



Ministério da  
**Ciência, Tecnologia  
e Inovação**



sid.inpe.br/mtc-m19/2012/08.07.02.07-TDI

## O MÉTODO DO Q-GRADIENTE PARA OTIMIZAÇÃO GLOBAL

Aline Cristina Soterroni

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Fernando Manoel Ramos, e Roberto Luiz Galski aprovada em 28 de agosto de 2012.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/3CDHUH8>>

INPE  
São José dos Campos  
2012

## **PUBLICADO POR:**

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

## **CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):**

### **Presidente:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

### **Membros:**

Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr<sup>a</sup> Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr<sup>a</sup> Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

### **BIBLIOTECA DIGITAL:**

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

### **REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

### **EDITORAÇÃO ELETRÔNICA:**

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)



Ministério da  
**Ciência, Tecnologia  
e Inovação**



sid.inpe.br/mtc-m19/2012/08.07.02.07-TDI

## O MÉTODO DO Q-GRADIENTE PARA OTIMIZAÇÃO GLOBAL

Aline Cristina Soterroni

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Fernando Manoel Ramos, e Roberto Luiz Galski aprovada em 28 de agosto de 2012.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/3CDHUH8>>

INPE  
São José dos Campos  
2012

Dados Internacionais de Catalogação na Publicação (CIP)

---

Soterroni, Aline Cristina.

So76m

O método do q-gradiente para otimização global / Aline Cristina Soterroni. – São José dos Campos : INPE, 2012.

xxiv + 125 p. ; (sid.inpe.br/mtc-m19/2012/08.07.02.07-TDI)

Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2012.

Orientadores : Drs. Fernando Manoel Ramos, e Roberto Luiz Galski.

1. q-derivada. 2. q-gradiente. 3. método do q-gradiente.  
I.Título.

CDU 004.832

---

Copyright © 2012 do MCT/INPE. Nenhuma parte desta publicação pode ser reproduzida, armazenada em um sistema de recuperação, ou transmitida sob qualquer forma ou por qualquer meio, eletrônico, mecânico, fotográfico, reprográfico, de microfilmagem ou outros, sem a permissão escrita do INPE, com exceção de qualquer material fornecido especificamente com o propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

Copyright © 2012 by MCT/INPE. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, microfilming, or otherwise, without written permission from INPE, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the reader of the work.

Aprovado (a) pela Banca Examinadora  
em cumprimento ao requisito exigido para  
obtenção do Título de **Doutor(a)** em  
**Computação Aplicada**

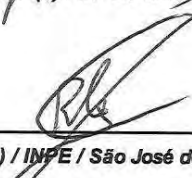
Dr. José Carlos Becceneri

  
\_\_\_\_\_  
Presidente / INPE / SJC Campos - SP

Dr. Fernando Manuel Ramos

  
\_\_\_\_\_  
Orientador(a) / INPE / SJC Campos - SP

Dr. Roberto Luiz Galski

  
\_\_\_\_\_  
Orientador(a) / INPE / São José dos Campos - SP

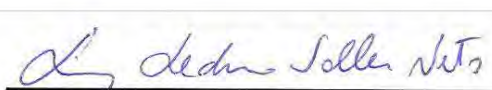
Dr. Stephan Stephany

  
\_\_\_\_\_  
Membro da Banca / INPE / SJC Campos - SP

Dr. Fernando José Von Zuben

  
\_\_\_\_\_  
Convidado(a) / UNICAMP / Campinas - SP

Dr. Luiz Leduino de Salles Neto

  
\_\_\_\_\_  
Convidado(a) / UNIFESP / São Paulo - SP

**Este trabalho foi aprovado por:**

( ) maioria simples

(X) unanimidade

Aluno (a): **Aline Cristina Soterroni**

São José dos Campos, 28 de Agosto de 2012



*The White Rabbit put on his spectacles.  
‘Where shall I begin, please your Majesty?’ he asked.  
‘Begin at the beginning,’ the King said gravely,  
‘and go on till you come to the end: then stop.’*

LEWIS CARROLL  
em “*Alice’s Adventures in Wonderland*”, 1865





*A minha família*



## AGRADECIMENTOS

Agradeço primeiramente a Deus.

Agradeço ao Professor Dr. Fernando Manuel Ramos, meu orientador desde o mestrado, que me acolheu desde o meu primeiro dia no INPE e desde então vem demonstrando confiança e entusiasmo com o meu trabalho. Foi o Fernando quem me introduziu na área de otimização com suas ideias de aplicação do  $q$ -cálculo em métodos de otimização, tema central desta tese. Agradeço ao Professor Dr. Roberto Luiz Galski que coorientou o trabalho com seriedade, motivação e competência. Agradeço aos dois orientadores por toda atenção, dedicação, compreensão, discussões e amizade que vou levar para sempre.

Agradeço ao Instituto Nacional de Pesquisas Espaciais pela oportunidade de prosseguir meus estudos e por toda infraestrutura concedida. Agradeço aos professores do programa de pós-graduação em Computação Aplicada do INPE por todo o conhecimento compartilhado e aos colegas de disciplina pelo companheirismo.

Agradeço à banca examinadora por todas as sugestões e comentários visando o aprimoramento deste trabalho.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro (processo 140720/2007-3).

Agradeço a todos os amigos que fiz nesta caminhada, principalmente, a segunda mãe e as irmãs que ganhei.

Agradeço imensamente à minha família, pais e irmãos, pelo apoio e amor incondicionais. E, em especial, ao meu marido Leonardo por compreender a minha ausência e por me fazer tão feliz.

Agradeço a todos que contribuíram direta e indiretamente para a conclusão deste trabalho.



## RESUMO

O reverendo inglês Frank Hilton Jackson foi o primeiro a desenvolver o  $q$ -cálculo de forma sistemática e, no início do século XX, reintroduziu a  $q$ -derivada, que ficou amplamente conhecida como derivada de Jackson. O  $q$ -cálculo, por sua vez, surgiu da generalização de expressões matemáticas por meio de um parâmetro  $q$ , dando origem a  $q$ -versões de funções, séries, operadores e números especiais que, no limite  $q \rightarrow 1$ , retomam suas respectivas versões clássicas. Este trabalho introduz o conceito de vetor  $q$ -gradiente na área de otimização por meio do método do  $q$ -gradiente, uma generalização do método da máxima descida que utiliza a direção contrária à direção do vetor  $q$ -gradiente como direção de busca. O uso dessa direção de busca, juntamente com estratégias apropriadas para a obtenção do parâmetro  $q$  e do tamanho do passo, mostrou que o método do  $q$ -gradiente realiza, ao longo do procedimento de otimização, uma transição suave entre busca global e busca local, além de possuir mecanismos para escapar de mínimos locais. O método do  $q$ -gradiente foi comparado com algoritmos determinísticos e extensivamente comparado com os Algoritmos Evolutivos (AEs), que participaram da competição do *IEEE Congress on Evolutionary Computation* (CEC) em 2005, sobre um conjunto de funções teste da literatura. Os resultados mostraram que o método do  $q$ -gradiente é competitivo em relação aos AEs, sobretudo nos problemas multimodais. O método do  $q$ -gradiente também foi aplicado na resolução de um problema da engenharia aeroespacial e os resultados apontaram para a viabilidade do seu uso em aplicações práticas.



# THE $q$ -GRADIENT METHOD FOR GLOBAL OPTIMIZATION

## ABSTRACT

The English reverend Frank Hilton Jackson was the first to develop the  $q$ -calculus in a systematic way, and in the beginning of the twentieth century he reintroduced the  $q$ -derivative, widely known as Jackson's derivative. The  $q$ -calculus, by its turn, came from generalizations of mathematical expressions called  $q$ -versions of functions, series, operators and special numbers that take into account a parameter  $q$ . In the limiting case of  $q \rightarrow 1$ , the  $q$ -versions reduce to its classical versions. In this work the concept of  $q$ -gradient is introduced in the optimization area by the  $q$ -gradient method, a generalization of the steepest descent method that uses the negative of the  $q$ -gradient as the search direction. The optimization procedure, with this direction and properly defined strategies for the parameter  $q$  and the step length, has shown that the search process gradually shifts from global in the beginning to local in the end with an effective mechanism for escaping from local minima. The  $q$ -gradient method was compared with some deterministic methods and extensively compared with Evolutionary Algorithms (EAs) of the 2005 IEEE Congress on Evolutionary Computation (CEC) over benchmark test functions. The results presented here have shown the competitiveness of the  $q$ -gradient over the EAs specially for the multimodal problems. The  $q$ -gradient method was also applied to an optimization problem from aerospace engineering and the results indicated the viability of the method for solving practical problems.





## LISTA DE FIGURAS

	<u>Pág.</u>
2.1 Mínimos: $x^1$ é o mínimo global, $x^2$ é um mínimo local estrito, $[x^3; x^4]$ é um conjunto de mínimos locais não estritos. . . . .	6
2.2 Conjunto convexo (esquerda) e conjunto não-convexo (direita). . . . .	8
2.3 Classificação dos problemas de otimização. . . . .	12
2.4 Convergência lenta e comportamento em zigzague do método da máxima descida. . . . .	17
3.1 Interpretação geométrica do $q$ -gradiente para uma função de uma única variável e diferentes valores do parâmetro $q$ . . . . .	33
3.2 Direções do anti-gradiente (vermelho) e do anti- $q$ -gradiente (preto) para diferentes valores do parâmetro $\mathbf{q}$ . . . . .	36
3.3 Função densidade de probabilidade uniforme para diferentes valores de $[q_{min}; q_{max}]$ . . . . .	38
3.4 Direções do anti- $q$ -gradiente para diferentes valores do parâmetro $\mathbf{q}$ , gerados segundo uma distribuição uniforme no intervalo $I$ . . . . .	39
3.5 Função densidade de probabilidade log-normal para $\mu = 1$ e diferentes valores para o desvio-padrão $\sigma$ . . . . .	40
3.6 Direções do anti- $q$ -gradiente para $\mathbf{q}$ 's obtidos segundo uma distribuição log-normal, com diferentes valores de $\sigma^0$ e sem decrescimento. . . . .	41
3.7 Evolução de $\sigma^k$ ao longo de 500 iterações, com $\sigma^0 = 1$ e diferentes fatores de redução $\beta$ . . . . .	42
3.8 Direções do anti- $q$ -gradiente para $\mathbf{q}$ 's obtidos segundo uma distribuição log-normal, com $\sigma^0 = 1$ e diferentes fatores de redução $\beta$ . . . . .	43
3.9 Direções do anti- $q$ -gradiente para $\mathbf{q}$ 's obtidos segundo uma distribuição log-normal, com $\sigma^0 = 1$ e fator de redução $\beta = 0,99$ . . . . .	44
3.10 Função densidade de probabilidade gaussiana para $\mu = x_i = 2$ e diferentes valores para o desvio-padrão $\sigma$ . . . . .	46
3.11 Direções do anti- $q$ -gradiente para $\mathbf{q}$ 's obtidos segundo uma distribuição gaussiana, com diferentes valores de $\sigma^0$ e sem decrescimento. . . . .	46
3.12 Evolução de $\sigma^k$ ao longo de 500 iterações, com $\sigma^0 = 10$ e diferentes fatores de redução $\beta$ . . . . .	47
3.13 Direções do anti- $q$ -gradiente para $\mathbf{q}$ 's obtidos segundo uma distribuição gaussiana, com $\sigma^0 = 10$ e diferentes fatores de redução $\beta$ . . . . .	48

3.14	Direções do anti- $q$ -gradiente para $\mathbf{q}$ 's obtidos segundo uma distribuição gaussiana, com $\sigma^0 = 10$ e $\beta = 0,99$ . . . . .	49
3.15	Densidade não-uniforme de pontos para (a) primeiro quadrante e (b) todos os quadrantes. Regiões próximas da origem são mais visitadas que as demais regiões do espaço de busca. . . . .	52
3.16	(a) Função bidimensional dada pela Equação 3.22 e (b) suas curvas de contorno para $\mathbf{o} = (0; 0)$ . . . . .	55
3.17	(a) Função bidimensional dada pela Equação 3.22 e (b) suas curvas de contorno para $\mathbf{o} = (10; 10)$ . . . . .	55
3.18	Sequência de pontos $\{\mathbf{x}^k\}$ obtida para a estratégia de geração dos parâmetros $\mathbf{q}$ 's segundo uma distribuição uniforme no intervalo $[q_{min}; q_{max}]$ e tamanho do passo adaptativo. Considerou-se o domínio da $f$ deslocado com $\mathbf{o} = (10; 10)$ . . . . .	56
3.19	Sequência de pontos $\{\mathbf{x}^k\}$ obtida para a estratégia de geração dos parâmetros $\mathbf{q}$ 's segundo uma distribuição log-normal com $\mu = 1$ , $\sigma^0 = 3$ e $\beta = 0,99$ , e tamanho do passo adaptativo para (a) $\mathbf{o} = (10; 10)$ e (b) $\mathbf{o} = (0; 0)$ . . . . .	57
3.20	Sequência de pontos $\{\mathbf{x}^k\}$ obtida para a estratégia de geração dos parâmetros $\mathbf{q}$ 's segundo uma distribuição log-normal com $\mu = 1$ , $\sigma^0 = 3$ e $\beta = 0,99$ , e controle do passo por redução com $\alpha^0 = 3,75$ para (a) $\mathbf{o} = (10; 10)$ e (b) $\mathbf{o} = (0; 0)$ . . . . .	57
3.21	Sequência de pontos $\{\mathbf{x}^k\}$ obtida para a estratégia de geração dos parâmetros $\mathbf{q}$ 's segundo uma distribuição gaussiana com $\mu = \mathbf{x}^k$ , $\sigma^0 = 3$ e $\beta = 0,99$ , e tamanho do passo adaptativo para (a) $\mathbf{o} = (10; 10)$ e (b) $\mathbf{o} = (0; 0)$ . . . . .	58
3.22	Sequência de pontos $\{\mathbf{x}^k\}$ obtida para a estratégia de geração dos parâmetros $\mathbf{q}$ 's segundo uma distribuição gaussiana com $\mu = \mathbf{x}^k$ , $\sigma^0 = 3$ e $\beta = 0,99$ , e controle do passo por redução com $\alpha^0 = 3,75$ para (a) $\mathbf{o} = (10; 10)$ e (b) $\mathbf{o} = (0; 0)$ . . . . .	58
4.1	Função De Jong . . . . .	64
4.2	Função <i>Ellipsoidal</i> . . . . .	64
4.3	Função Rosenbrock . . . . .	65
4.4	Função Rastrigin . . . . .	65
4.5	Evolução de $F(\mathbf{x}_{melhor})$ médio em função do NAFO para a função teste De Jong e os métodos do $q$ -gradiente e da máxima descida. . . . .	68
4.6	Evolução de $F(\mathbf{x}_{melhor})$ médio em função do NAFO para a função teste <i>Ellipsoidal</i> e os métodos do $q$ -gradiente e da máxima descida. . . . .	68

4.7	Evolução de $F(\mathbf{x}_{melhor})$ médio em função do NAFO para a função teste Rosenbrock e os métodos do $q$ -gradiente e da máxima descida. . . . .	69
4.8	Evolução de $F(\mathbf{x}_{melhor})$ médio em função do NAFO para a função teste Rastrigin e os métodos do $q$ -gradiente e da máxima descida. . . . .	69
4.9	Evolução de $F(\mathbf{x}_{melhor})$ médio em função de $4 \cdot 10^5$ avaliações da função objetivo para a função teste Rastrigin e os métodos do $q$ -gradiente e da máxima descida. . . . .	70
4.10	Função Schwefel . . . . .	71
4.11	Função Ackley . . . . .	72
4.12	Função Rastrigin Rotacionada . . . . .	73
5.1	Constelação de satélites GPS. . . . .	92
5.2	Tipos de geometria que influenciam na precisão de uma posição GPS. . . . .	93
5.3	Localização dos 5 receptores sobre o território brasileiro. . . . .	95
5.4	Localização dos 6 receptores sobre o território sul-americano. . . . .	96
5.5	GDOP para o RPS brasileiro com 4 satélites e 5 receptores. . . . .	100
5.6	Primeiro pico das curvas GDOP para o RPS brasileiro às 2h18. . . . .	101
5.7	Segundo pico das curvas GDOP para o RPS brasileiro às 14h18. . . . .	102
5.8	GDOP para o RPS sul-americano com 4 satélites e 6 receptores. . . . .	103
5.9	Primeiro pico das curvas GDOP para o RPS sul-americano às 11h30. . . . .	103
5.10	Segundo pico das curvas GDOP para o RPS sul-americano às 23h30. . . . .	104
5.11	Zoom da Figura 5.10. Note que os satélites S1 e S2 possuem órbitas próximas. . . . .	104



## LISTA DE TABELAS

	<u>Pág.</u>
4.1 Parâmetros usados pelo método do $q$ -gradiente. . . . .	66
4.2 Número de avaliações da função objetivo necessário para satisfazer a condição $ F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)  \leq 0,001$ . . . . .	67
4.3 Média dos melhores valores da função objetivo ( $F(\mathbf{x}_{melhor})$ Médio) obtidos em 50 execuções independentes dos algoritmos com NAFO máximo igual a $10^5$ . . . . .	67
4.4 Parâmetros usados pelo método do $q$ -gradiente para o primeiro conjunto de funções teste. . . . .	74
4.5 Desempenho comparativo, em termos no número de avaliações da função objetivo (NAFO), entre o algoritmo para o método do $q$ -gradiente, os AGs G3-PCX, SPC-vSBX e SPC-PNX, e a meta-heurística GEOvar+EE modificado para as funções unimodais <i>Ellipsoidal</i> e Schwefel, e a função Rosenbrock. . . . .	75
4.6 Desempenho comparativo, em termos no número de avaliações da função objetivo (NAFO), entre o algoritmo para o método do $q$ -gradiente, os AGs G3-PCX, SPC-vSBX e SPC-PNX, e a meta-heurística GEOvar+EE modificado para as funções multimodais Ackley, Rastrigin e Rastrigin Rotacionada. . . . .	76
4.7 $F(\mathbf{x}_{melhor})$ encontrado pelo <i>solver</i> GAMS/LINDOGlobal e pelo método do $q$ -gradiente sobre o primeiro conjunto de funções teste. . . . .	79
4.8 Principais características das funções teste do CEC-2005. . . . .	80
4.9 Parâmetros de ajuste utilizados pelo método do $q$ -gradiente para as funções F1 a F15 com as dimensões 10, 30 e 50. . . . .	82
4.10 Valores do NAFO para as funções F1 a F15 e $n = 10$ . . . . .	83
4.11 Valores do NAFO para as funções F1 a F15 e $n = 30$ . . . . .	84
4.12 Valores do NAFO para as funções F1 a F15 e $n = 50$ . . . . .	85
4.13 Taxas de desempenho normalizadas para as funções teste unimodais com $n = 10$ . . . . .	87
4.14 Taxas de desempenho normalizadas para as funções teste unimodais com $n = 30$ . . . . .	88
4.15 Taxas de desempenho normalizadas para as funções multimodais com $n = 10$ . . . . .	88
4.16 Taxas de desempenho normalizadas para as funções multimodais com $n = 30$ . . . . .	89

5.1	Valores de GDOP. . . . .	94
5.2	Parâmetros de ajuste utilizados pelo RPS brasileiro e sul-americano. . . . .	99
5.3	Parâmetros orbitais (em graus) e melhor GDOP médio (adimensional) para o RPS brasileiro e RPS sul-americano obtidos pelo método do $q$ -gradiente. . . . .	100
B.1	Precisões estabelecidas para as funções F1 a F15 do CEC 2005. . . . .	117
B.2	Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F1 a F8 com $n = 10$ .118	
B.3	Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F9 a F15 com $n = 10$ .119	
B.4	Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F1 a F8 com $n = 30$ .120	
B.5	Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F9 a F15 com $n = 30$ .121	
B.6	Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F1 a F8 com $n = 50$ .122	
B.7	Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F9 a F15 com $n = 50$ .123	

## LISTA DE ABREVIATURAS E SIGLAS

AE	–	Algoritmo Evolutivo
GAMS	–	General Algebraic Modeling System
CEC	–	IEEE Congress on Evolutionary Computation
DFP	–	Davidon-Fletcher-Powell
BFGS	–	Broyden-Fletcher-Goldfarb-Shanno
RS	–	Recozimento Simulado
AG	–	Algoritmo Genético
NAFO	–	Número de Avaliações da Função Objetivo
EE	–	Estratégia Evolutiva
GEO	–	Generalized Extremal Optimization
GEO+var	–	Variante do GEO
GEOvar+EE	–	Hibridização do GEOvar com EE
GNSS	–	Global Navigation Satellite System
GPS	–	Global Positioning System
RPS	–	Regional Positioning System
GDOP	–	Geometric Dilution of Precision





# SUMÁRIO

	<u>Pág.</u>
<b>1 INTRODUÇÃO</b> . . . . .	<b>1</b>
<b>2 CONCEITOS DE OTIMIZAÇÃO</b> . . . . .	<b>5</b>
2.1 Definição do problema de otimização mono-objetivo . . . . .	5
2.2 Condições de otimalidade . . . . .	6
2.3 Noções de convergência . . . . .	9
2.4 Critérios de parada . . . . .	10
2.5 Classificação dos problemas de otimização . . . . .	11
2.6 Métodos para otimização irrestrita . . . . .	13
2.6.1 Busca linear . . . . .	14
2.6.2 Método da máxima descida . . . . .	16
2.6.3 Método de Newton . . . . .	17
2.6.4 Método dos gradientes conjugados . . . . .	19
2.6.5 Métodos quase-Newton . . . . .	21
2.7 Métodos para otimização global . . . . .	22
2.7.1 Recozimento Simulado . . . . .	24
2.7.2 Algoritmos Genéticos . . . . .	26
<b>3 MÉTODOS BASEADOS EM <math>q</math>-GRADIENTE</b> . . . . .	<b>29</b>
3.1 $q$ -derivada . . . . .	29
3.2 $q$ -gradiente . . . . .	31
3.3 Algoritmo para métodos baseados em $q$ -gradiente . . . . .	34
3.4 Estratégias para cálculo do parâmetro $q$ . . . . .	35
3.4.1 Estratégia 1: distribuição uniforme . . . . .	37
3.4.2 Estratégia 2: distribuição log-normal . . . . .	39
3.4.3 Estratégia 3: distribuição gaussiana . . . . .	45
3.5 Estratégias para escolha do tamanho do passo . . . . .	50
3.5.1 Estratégia 1: busca linear . . . . .	50
3.5.2 Estratégia 2: passo adaptativo . . . . .	50
3.5.3 Estratégia 3: controle do passo por redução . . . . .	52
3.6 Método do $q$ -gradiente . . . . .	53
<b>4 RESULTADOS NUMÉRICOS</b> . . . . .	<b>61</b>

4.1	Parâmetros de ajuste . . . . .	61
4.2	Validações em funções teste . . . . .	63
4.3	Primeiro conjunto de funções teste . . . . .	71
4.3.1	Comparação com algoritmos genéticos . . . . .	73
4.3.2	Comparação com otimização global determinística . . . . .	77
4.4	Segundo conjunto de funções teste . . . . .	79
<b>5</b>	<b>APLICAÇÃO À ENGENHARIA ESPACIAL . . . . .</b>	<b>91</b>
5.1	Introdução . . . . .	91
5.2	Problema de otimização . . . . .	95
5.3	Metodologia . . . . .	98
5.4	Análises e resultados . . . . .	99
<b>6</b>	<b>CONCLUSÕES . . . . .</b>	<b>105</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS . . . . .</b>	<b>109</b>
	<b>APÊNDICE A - TABELAS COMPLEMENTARES PARA AS</b>	
	<b>FUNÇÕES DO CEC 2005 . . . . .</b>	<b>117</b>

# 1 INTRODUÇÃO

Otimizar é buscar a melhor resposta possível para um determinado objetivo. Esse objetivo vai além da construção de um equipamento que desempenhe máxima eficiência, da elaboração de uma rota que percorra um caminho no menor tempo possível, ou do aumento dos lucros ao mesmo tempo em que os custos de produção são reduzidos. “O conceito de otimização está intrinsecamente ligado ao desejo do homem em ser o melhor no que faz”, (VANDERPLAATS, 1984).

O conceito de otimização está arraigado como um princípio determinante na análise de muitos problemas de decisão complexos. Além de oferecer uma certa elegância filosófica ao tratamento dos problemas, o conceito de otimização frequentemente oferece um indispensável grau de simplicidade operacional. É possível, por exemplo, aproximar um problema de decisão complexo por meio de um único objetivo capaz de quantificar o desempenho e medir a qualidade de uma decisão (LUENBERGER, 1984).

Matematicamente, um objetivo pode ser expresso por meio de uma função objetivo sujeita a variáveis de decisão, e são as variáveis de decisão que devem ser modificadas para que o objetivo seja otimizado. Uma solução será boa ou ruim dependendo do valor que as variáveis de decisão assumem ao serem avaliadas pela função objetivo. Todos os valores possíveis das variáveis de decisão determinam o espaço de busca viável. E o conjunto de todos os valores dessas variáveis para os quais nenhuma restrição é violada representa o espaço de busca viável.

De acordo com as características da função objetivo, variáveis de decisão e restrições, os problemas de otimização podem ser classificados em contínuos ou discretos, restritos ou irrestritos, lineares ou não-lineares, convexos ou não-convexos, unimodais ou multimodais, locais ou globais, mono-objetivo ou multiobjetivo, dentre outros.

Dentre essas classes de problemas, destacam-se os problemas de otimização global. Problemas desse tipo, em geral, podem assumir outras características além da não-convexidade e existência de vários extremos locais. Eles podem ser não-lineares, não-diferenciáveis, descontínuos, ou ainda apresentarem funções objetivo e restrições do tipo caixa preta (*black box*), em que poucas informações sobre o comportamento delas são conhecidas e a avaliação é feita apenas por meio de simulação numérica de um modelo. Os métodos numéricos para problemas de otimização global não fornecem garantias de que o extremo encontrado seja o extremo global. Embora esses problemas sejam difíceis de se resolver, eles constituem a maioria dos problemas reais

com aplicações, por exemplo, em engenharia mecânica, civil e química, otimização estrutural, biologia molecular, projetos de chips VLSI, processamento de imagens e problemas de otimização combinatorial (ZABINSKY, 2003).

Os algoritmos para problemas de otimização global podem ser classificados em determinísticos e estocásticos (ZABINSKY, 2003). Os métodos estocásticos se referem aos algoritmos que utilizam algum tipo de aleatoriedade durante o procedimento de busca pelo ótimo como, por exemplo, um gerador de números pseudo-aleatórios, e podem produzir resultados diferentes a cada execução do algoritmo, mesmo que as condições iniciais sejam repetidas. Já os métodos determinísticos produzem sempre a mesma solução para uma mesma condição inicial (BECCENERI, 2008).

Os métodos estocásticos para otimização global têm mostrado grande potencial na resolução de problemas de otimização de grande escala em um tempo aceitável, incluindo problemas cujas funções são do tipo caixa preta. Ao contrário dos métodos determinísticos, que são mais apropriados para problemas com estruturas especiais, como problemas convexos em que o extremo local é também o extremo global, e cujo custo computacional aumenta com a dimensionalidade do problema tornando o seu uso proibitivo em muitas aplicações. Embora os métodos estocásticos não possuam uma base teórica sólida, com provas de convergência ou provas sobre a qualidade das soluções obtidas, e sejam ajustados para cada problema em questão por meio de parâmetros de ajuste, que muitas vezes são definidos por tentativa e erro, eles são de fácil implementação, dependem basicamente de avaliações da função objetivo e são utilizados na prática para resolver problemas reais (ZABINSKY, 2003).

“A natureza otimiza”, (NOCEDAL; WRIGHT, 2006). Não é por acaso que o homem busca em fenômenos naturais analogias para criar algoritmos de otimização. Exemplos de métodos estocásticos inspirados em processos naturais são as meta-heurísticas Recozimento Simulado, Algoritmos Genéticos, Colônia de Formigas e Enxame de Partículas. Esses métodos inteligentes e adaptáveis a diversos tipos de problemas tentam explorar eficientemente o espaço das soluções viáveis por meio de mecanismos que, ao longo do procedimento iterativo, permitem a escolha de soluções piores como forma de escapar de extremos locais e, dessa forma, fornecem soluções cada vez melhores e confiáveis na direção do extremo global, mesmo sem nenhuma garantia de otimalidade global.

São as características da formulação do problema de otimização que auxiliam a escolha do método ou algoritmo a ser utilizado, e existem inúmeras técnicas destinadas à tarefa de resolver esses problemas. Segundo o teorema da inexistência do almoço

grátis (*No-Free-Lunch Theorem*), se um algoritmo  $A_1$  é melhor que um algoritmo  $A_2$  para um conjunto de problemas, então existirá um outro conjunto de problemas para os quais o algoritmo  $A_2$  apresenta um desempenho superior ao algoritmo  $A_1$  (WOLPERT; MACREADY, 1997). De fato, não existe um método que seja o melhor ou o pior dentre todos, o que existe é um método mais ou menos apropriado para o problema em questão, e mesmo assim não é possível antecipar a escolha de tal método. A pesquisa em otimização global é uma área em desenvolvimento e junto com o surgimento de novos problemas tem-se também a necessidade do desenvolvimento de novos algoritmos.

O  $q$ -cálculo, do inglês *q-calculus*, surgiu da generalização de expressões matemáticas por meio de um parâmetro  $q$ . Essas generalizações deram origem a versões análogas, também chamadas de  $q$ -versões, de funções, séries, operadores, e números especiais que, no limite  $q \rightarrow 1$ , retomam suas respectivas versões clássicas. Essas generalizações remontam aos trabalhos de Fermat, Euler, Heine e Gauss, mas foi no início do século XX que o  $q$ -cálculo recebeu importantes contribuições do reverendo inglês Frank Hilton Jackson. Ele foi o primeiro a desenvolver o  $q$ -cálculo de forma sistemática, através de inúmeros artigos contendo generalizações de funções, séries e números especiais no contexto do  $q$ -cálculo (CHAUNDY, 1962; ERNST, 2000; KAC; CHEUNG, 2002). Em particular, F.H. Jackson reintroduziu o conceito de  $q$ -derivada, amplamente conhecida como derivada de Jackson, e criou o conceito de  $q$ -integral definida (JACKSON, 1904; JACKSON, 1908; JACKSON, 1910a; JACKSON, 1910b).

O principal objetivo do presente trabalho é estender a aplicabilidade do  $q$ -cálculo na área de otimização. Baseado no conceito de  $q$ -derivada, o  $q$ -gradiente – vetor das  $q$ -derivadas parciais de primeira ordem da função objetivo – foi introduzido e utilizado no método da máxima descida por meio da substituição do gradiente clássico pelo  $q$ -gradiente. Dessa forma, foi criada uma  $q$ -versão do método da máxima descida, denominado método do  $q$ -gradiente, que retoma a sua versão clássica à medida que o parâmetro  $q$  tende a 1. Com estratégias simples para a obtenção do parâmetro  $q$  (necessário para o cálculo do  $q$ -gradiente) e para o tamanho do passo na direção de busca, foi criado um algoritmo estocástico para o método do  $q$ -gradiente destinado aos problemas de otimização contínuos, irrestritos, não-lineares, não-convexos e multimodais, isto é, problemas de otimização global. Embora o algoritmo seja originalmente desenvolvido para problemas sem restrições, ele pode ser facilmente adaptado para funções com restrições através do uso de métodos de penalidade (IZMAILOV; SOLODOV, 2007).

De fácil implementação, o método do  $q$ -gradiente conta com, no máximo, três parâmetros de ajuste e introduz características de busca global em um método de busca local. Além da comparação com o método da máxima descida e o *solver* GAMS/LINDOGlobal para otimização global determinística disponível no servidor NEOS (CZYZYK et al., 1998), o método do  $q$ -gradiente foi extensivamente comparado com Algoritmos Evolutivos (AEs), como os desenvolvidos por Deb et al. (2002) e Ballester e Carter (2004), e os 11 AEs que participaram da competição do *IEEE Congress on Evolutionary Computation* em 2005 (CEC-2005). Os resultados mostraram que o método do  $q$ -gradiente é competitivo em relação aos AEs sobre os problemas multimodais, comprovando a sua capacidade de escapar de extremos locais e caminhar na direção do extremo global. O método também foi aplicado na resolução de um problema de engenharia aeroespacial.

Este trabalho está organizado da seguinte forma. No Capítulo 2, são apresentados conceitos gerais de problemas e métodos de otimização. No Capítulo 3, são introduzidos o  $q$ -gradiente e o método do  $q$ -gradiente, juntamente com as estratégias de cálculo do parâmetro  $q$  e do tamanho do passo dado nesta direção. No Capítulo 4, são apresentados resultados de validação e comparação do método do  $q$ -gradiente em funções teste da literatura. No Capítulo 5 são apresentados o problema de engenharia aeroespacial e os resultados obtidos da sua resolução pelo método do  $q$ -gradiente. E, finalmente, no Capítulo 6, são apresentadas as conclusões deste trabalho e sugestões para trabalhos futuros.

## 2 CONCEITOS DE OTIMIZAÇÃO

Este capítulo apresenta conceitos clássicos sobre problemas de otimização mono-objetivo, ou seja, problemas descritos por uma única função objetivo. São apresentados também os principais métodos determinísticos baseados em gradiente e métodos para otimização global estocástica.

Ao longo deste texto, o problema de otimização será tratado como um problema de minimização, por uma questão de notação, uma vez que sempre é possível transformar um problema de maximização em um problema de minimização.

### 2.1 Definição do problema de otimização mono-objetivo

O problema de otimização consiste na minimização de um objetivo que pode ser expresso, matematicamente, por meio de uma função objetivo  $f(\mathbf{x})$  sujeita as variáveis de decisão  $\mathbf{x}$ . São essas variáveis que devem ser modificadas para que a função objetivo seja otimizada. Todos os valores possíveis das variáveis de decisão determinam o espaço de busca viável.

Matematicamente, dada uma função objetivo  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  e as variáveis de decisão  $\mathbf{x} \in \mathcal{F}$  em que  $\mathcal{F} \subseteq \mathbb{R}^n$  é o conjunto viável, então o problema de otimização é dito sem restrições (ou irrestrito) se  $\mathcal{F} = \mathbb{R}^n$  e pode ser definido como (PILLO; PALAGI, 2002a)

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}). \quad (2.1)$$

Quando  $\mathcal{F}$  é um subconjunto do  $\mathbb{R}^n$  descrito por restrições de desigualdade, restrições de igualdade e/ou restrições laterais sobre as variáveis de decisão, ou seja,

$$\mathcal{F} = \{\mathbf{x} : g_i(\mathbf{x}) \leq 0, i = 1, \dots, p; h_j(\mathbf{x}) = 0, j = 1, \dots, m; x_k^I \leq x_k \leq x_k^S, k = 1, \dots, n\},$$

então o problema é dito com restrições e sua formulação é dada por

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{F}} \quad & f(\mathbf{x}) \\ & g(\mathbf{x}) \leq 0 \\ & h(\mathbf{x}) = 0 \\ & \mathbf{x}^I \leq \mathbf{x} \leq \mathbf{x}^S, \end{aligned} \quad (2.2)$$

em que  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\mathbf{x}^I$  é o limite inferior e  $\mathbf{x}^S$  é o limite superior.

Um ponto  $\mathbf{x}^* \in \mathcal{F}$  é uma solução global, ou mínimo global, do problema 2.1 se  $f(\mathbf{x}^*) \leq f(\mathbf{x})$ ,  $\forall \mathbf{x} \in \mathcal{F}$ ,  $\mathbf{x} \neq \mathbf{x}^*$ . Se  $\mathcal{F}$  é um conjunto compacto não vazio, então pelo Teorema de Weierstrass, o problema 2.1 possui uma solução global. Um ponto  $\mathbf{x}^* \in \mathcal{F}$  é uma solução local, ou mínimo local, do problema 2.1 se  $f(\mathbf{x}^*) \leq f(\mathbf{x})$ ,  $\forall \mathbf{x} \in \mathcal{F} \cap \mathcal{B}$  em que  $\mathcal{B}$  é uma vizinhança de  $\mathbf{x}^*$ . Quando a desigualdade é estrita para todo  $\mathbf{x} \neq \mathbf{x}^*$ , então  $\mathbf{x}^*$  é minimizador estrito (global ou local).

A Figura 2.1 ilustra mínimos locais e o mínimo global para uma função  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Note que todo mínimo global é também um mínimo local, mas a recíproca não é verdadeira.

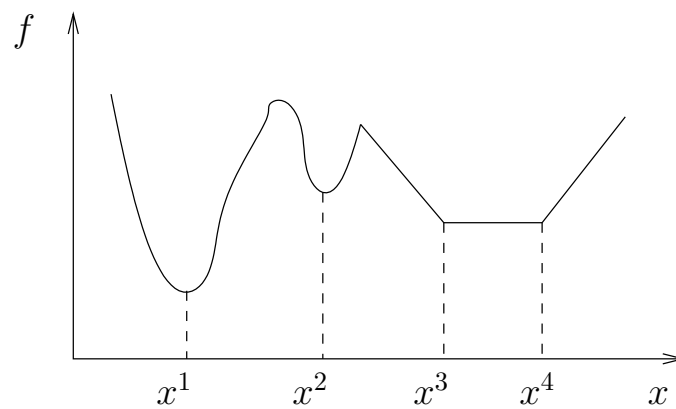


Figura 2.1 - Mínimos:  $x^1$  é o mínimo global,  $x^2$  é um mínimo local estrito,  $[x^3; x^4]$  é um conjunto de mínimos locais não estritos.

Fonte: Adaptado de [Izmailov e Solodov \(2005\)](#).

Para determinar se um ponto  $\mathbf{x}^*$  é uma solução local do problema de otimização, basta que ele satisfaça as condições de otimalidade. Essas condições utilizam informações do gradiente e da matriz Hessiana da função objetivo e, portanto, só podem ser aplicadas em problemas com funções objetivos contínuas e duas vezes diferenciáveis. A seguir, são apresentadas as condições de otimalidade para problemas de otimização da forma 2.1 e 2.2.

## 2.2 Condições de otimalidade

Seja  $\mathbf{x}^*$  um mínimo local do problema de otimização irrestrita 2.1 com  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  diferenciável, então o vetor gradiente da  $f$  em  $\mathbf{x}^*$  é nulo, ou seja,

$$\nabla f(\mathbf{x}^*) = \mathbf{0}, \quad (2.3)$$



em que

$$\nabla f(\mathbf{x}^*)^\top = \left[ \frac{\partial f(\mathbf{x}^*)}{\partial x_1} \quad \frac{\partial f(\mathbf{x}^*)}{\partial x_2} \quad \cdots \quad \frac{\partial f(\mathbf{x}^*)}{\partial x_n} \right].$$

Se  $f$  é duas vezes diferenciável, então além do vetor gradiente da função no ponto  $\mathbf{x}^*$  ser nulo, a matriz Hessiana de  $f$  em  $\mathbf{x}^*$  é semidefinida positiva, isto é,

$$\mathbf{y}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{y} \geq 0, \quad \forall \mathbf{y} \in \mathbb{R}^n, \quad (2.4)$$

em que

$$\nabla^2 f(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_n^2} \end{bmatrix}.$$

Se  $f$  é duas vezes diferenciável em  $\mathbf{x}^* \in \mathbb{R}^n$  e a matriz Hessiana de  $f$  em  $\mathbf{x}^*$  é definida positiva, ou seja, se existe  $\gamma > 0$  tal que

$$\mathbf{y}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{y} \geq \gamma \|\mathbf{y}\|^2, \quad \forall \mathbf{y} \in \mathbb{R}^n, \quad (2.5)$$

então  $\mathbf{x}^*$  é minimizador local estrito do problema de otimização 2.1.

A Equação 2.3 é a condição necessária de primeira ordem e os pontos que satisfazem esta condição são denominados pontos estacionários do problema 2.1. A Equação 2.4 é a condição necessária de segunda ordem, e a Equação 2.5 é a condição suficiente de segunda ordem para o problema 2.1 (IZMAILOV; SOLODOV, 2005).

Seja  $\mathbf{x}^*$  um mínimo local do problema de otimização com restrições 2.2 para  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$  e  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  diferenciáveis, então  $\mathbf{x}^*$  deverá satisfazer as condições de Kuhn-Tucker. Essas condições são definidas como os pontos estacionários do Lagrangiano (VANDERPLAATS, 1984)

$$\mathbf{L}(\mathbf{x}; \lambda) = f(\mathbf{x}) + \sum_{i=1}^p \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^m \lambda_{j+p} h_j(\mathbf{x}). \quad (2.6)$$

Logo, para  $\mathbf{x}^*$  ser um ponto de mínimo local, as seguintes condições de Kuhn-Tucker devem ser satisfeitas:

- i)  $\mathbf{x}^* \in \mathcal{F}$ , ou seja,  $\mathbf{x}^*$  não viola nenhuma restrição.  
 ii)  $\lambda_i g_i(\mathbf{x}^*) = 0$ ,  $\lambda_i \geq 0$ ,  $i = 1, \dots, p$ .

$$\text{iii) } \nabla f(\mathbf{x}^*) + \sum_{i=1}^p \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^m \lambda_{j+p} \nabla h_j(\mathbf{x}^*) = \mathbf{0}, \quad \lambda_i \leq 0, \quad \lambda_{j+p} \text{ com qualquer sinal.} \quad (2.7)$$

Para muitos problemas de otimização multimodais (i.e., com vários mínimos locais) existe a necessidade de se encontrar o melhor mínimo, ou seja, o mínimo global do problema. Ao contrário de soluções locais que possuem condições de otimalidade bem definidas, a solução global de um problema de otimização não é tão simples de ser determinada. As condições de otimalidade garantem que  $\mathbf{x}^*$  é uma solução global somente se o problema de otimização 2.1 possui alguma estrutura especial, por exemplo, se o problema é convexo e, nestes casos, o mínimo local é também o mínimo global do problema de otimização.

O conceito de convexidade é de grande importância na teoria de otimização. Um conjunto  $\mathcal{F} \subset \mathbb{R}^n$  é convexo se, para quaisquer  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$  e para qualquer número real  $a$ ,  $0 < a < 1$ , o ponto  $a\mathbf{x}_1 + (1-a)\mathbf{x}_2 \in \mathcal{F}$  (LUENBERGER, 1984). Geometricamente, um conjunto  $\mathcal{F}$  é convexo se o segmento de reta conectando dois pontos quaisquer deste conjunto está contido em  $\mathcal{F}$  (veja a Figura 2.2).

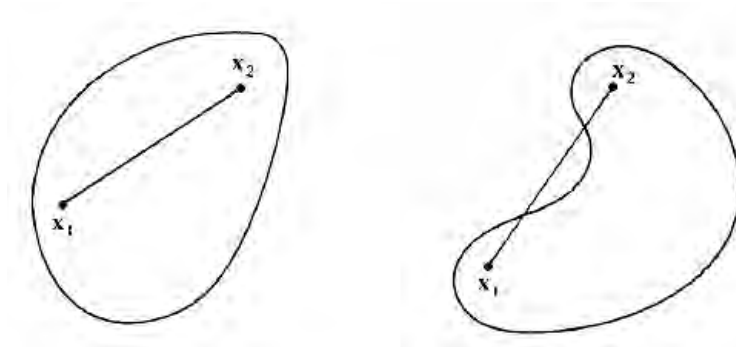


Figura 2.2 - Conjunto convexo (esquerda) e conjunto não-convexo (direita).  
 Fonte: (LUENBERGER, 1984).

Uma função  $f$  definida em um conjunto convexo  $\mathcal{F}$  é convexa se, para quaisquer  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$  e para qualquer número  $a$ ,  $0 \leq a \leq 1$ , tem-se

$$f(a\mathbf{x}_1 + (1-a)\mathbf{x}_2) \leq af(\mathbf{x}_1) + (1-a)f(\mathbf{x}_2).$$

Se, para qualquer número  $a$ ,  $0 < a < 1$ , e  $\mathbf{x}_1 \neq \mathbf{x}_2$ , tem-se

$$f(a\mathbf{x}_1 + (1 - a)\mathbf{x}_2) < af(\mathbf{x}_1) + (1 - a)f(\mathbf{x}_2),$$

então  $f$  é estritamente convexa (LUENBERGER, 1984). Logo, seja  $\mathcal{F}$  é um conjunto convexo com função objetivo  $f$  e restrições, se existirem, convexas e definidas em  $\mathcal{F}$ , então o mínimo local do problema de otimização é também o mínimo global.

Em um problema de otimização linear, também conhecido como programação linear (PL), a função objetivo é linear e as restrições podem ser expressas na forma de equações ou inequações lineares. Se o problema estiver bem definido, então o conjunto de soluções candidatas é composto pelos vértices do poliedro que representam o espaço de busca viável (IZMAILOV; SOLODOV, 2007).

### 2.3 Noções de convergência

Seja  $\{\mathbf{x}^k\}$ ,  $k = 0, 1, \dots$  uma sequência de pontos gerada por um algoritmo de otimização e seja  $\Omega \subset \mathcal{F}$  um subconjunto de pontos que satisfaz as condições de otimalidade de primeira ordem. Um algoritmo converge se um ponto limite  $\mathbf{x}^*$  da sequência  $\{\mathbf{x}^k\}$  é obtido tal que  $\mathbf{x}^* \in \Omega$ . Em outras palavras, o procedimento iterativo termina quando o algoritmo encontra um ponto estacionário. Se o problema é de otimização sem restrições, então  $\Omega = \{\mathbf{x} \in \mathbb{R}^n : \nabla f(\mathbf{x}) = 0\}$ . E se o problema é de otimização com restrições,  $\Omega$  pode ser definido pelo conjunto de pontos que satisfaz as condições de Kuhn-Tucker (PILLO; PALAGI, 2002a).

Se a partir de qualquer ponto inicial  $\mathbf{x}^0 \in \mathbb{R}^n$  o algoritmo gera uma sequência  $\{\mathbf{x}^k\}$  cujo ponto limite é  $\mathbf{x}^* \in \Omega$ , diz-se que o algoritmo possui convergência global. Quando o ponto limite  $\mathbf{x}^* \in \Omega$  é obtido apenas para  $\mathbf{x}^0$  em uma vizinhança de  $\Omega$ , diz-se que o algoritmo possui convergência local (PILLO; PALAGI, 2002a). Note que, tanto para convergência global quanto convergência local, a solução encontrada  $\mathbf{x}^*$  é apenas local. Um mínimo  $\mathbf{x}^* \in \Omega$  será global se o problema de otimização for convexo.

A noção de convergência será forte se qualquer subsequência de  $\{\mathbf{x}^k\}$  possuir um ponto limite que pertence a  $\Omega$ . No caso irrestrito, isto implica em (PILLO; PALAGI, 2002a)

$$\lim_{k \rightarrow \infty} \|\nabla f(\mathbf{x}^k)\| = 0.$$

Supondo que um algoritmo gera uma sequência  $\{\mathbf{x}^k\}$  que converge para o ponto

$\mathbf{x}^* \in \Omega$ , a ordem de convergência desta sequência é definida como o supremo dos números não negativos  $p$  que satisfaçam (BAZARAA; SHETTY, 1979)

$$\overline{\lim}_{k \rightarrow \infty} \frac{|\mathbf{x}^{k+1} - \mathbf{x}^*|}{|\mathbf{x}^k - \mathbf{x}^*|^p} = \beta < \infty.$$

Se  $p = 1$  e  $\beta < 1$ , a sequência converge linearmente para  $\mathbf{x}^*$  com taxa de convergência  $\beta$ . Se  $p > 1$  ou  $p = 1$ , e  $\beta = 0$ , a sequência possui convergência superlinear (BAZARAA; SHETTY, 1979).

Taxas de convergência podem ser úteis na comparação entre métodos. Porém, existem outras ferramentas de comparação relevantes. É importante, por exemplo, analisar o custo computacional de uma iteração, pois um algoritmo com uma taxa de convergência rápida pode ser computacionalmente lento (IZMAILOV; SOLODOV, 2007).

## 2.4 Critérios de parada

Na prática, nenhum procedimento iterativo pode ser infinito e os algoritmos devem levar isso em consideração, estabelecendo regras ou critérios de parada. Os critérios de parada podem se basear em estimativas de taxa de convergência, ou seja, se a taxa de convergência é conhecida, o algoritmo pode utilizá-la para definir o número de iterações necessárias (IZMAILOV; SOLODOV, 2007).

Critérios de parada podem se basear também em informações obtidas a partir da parte mais recente da sequência  $\{\mathbf{x}^k\}$  ou da sequência  $\{f(\mathbf{x}^k)\}$ , ou ainda em alguma medida de violação de condições de otimalidade (IZMAILOV; SOLODOV, 2007).

Se o mínimo global  $\mathbf{x}^*$  de  $f$  é conhecido, então o procedimento iterativo termina se a precisão  $\varepsilon$  for atingida e a condição a seguir for satisfeita para  $\varepsilon > 0$  e pequeno

$$|f(\mathbf{x}^k) - f(\mathbf{x}^*)| \leq \varepsilon. \quad (2.8)$$

Esse caso é comum para funções teste em que o objetivo é testar e comparar a eficiência de métodos de otimização.

Se o tamanho do passo é muito pequeno, de forma que não é mais possível melhorar significativamente a solução, então o critério de parada pode ser dado por

$$\|\mathbf{x}^{k+1} - \mathbf{x}^k\| \leq \varepsilon. \quad (2.9)$$

Analogamente, o mesmo raciocínio pode ser aplicado para o valor da função objetivo e o procedimento iterativo termina se

$$|f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k)| \leq \varepsilon. \quad (2.10)$$

Vale ressaltar que os critérios de parada dados pelas Equações 2.9 e 2.10 não garantem convergência da sequência  $\{\mathbf{x}^k\}$  para nenhum mínimo (local ou global) da função  $f$ . Apesar de não serem confiáveis do ponto de vista teórico, são de fácil implementação e muito utilizados na prática (IZMAILOV; SOLODOV, 2007).

Um critério de parada que pode ser aplicado em qualquer procedimento iterativo é o número máximo de iterações ou de avaliações da função objetivo. Este critério é indicado para problemas de difícil resolução, em que o procedimento iterativo é lento (VANDERPLAATS, 1984). Este critério é inclusive usado como medida de desempenho entre algoritmos. Em geral, a comparação é feita em termos da melhor solução encontrada por cada algoritmo quando o número máximo de avaliações da função objetivo é atingido.

## 2.5 Classificação dos problemas de otimização

Os algoritmos de otimização disponíveis são mais ou menos indicados para certas classes de problemas e reconhecer as características do problema a ser otimizado pode auxiliar na escolha do método mais apropriado e, conseqüentemente, na obtenção de uma solução satisfatória para o problema.

A classificação dos problemas de otimização pode ser feita de acordo com as características da função objetivo, variáveis de decisão e restrições. Zabinsky (2003) utiliza a classificação ilustrada na Figura 2.3. Essa classificação considera apenas os problemas de otimização mono-objetivo. A maior parte da pesquisa e aplicação da área de otimização é voltada para esses problemas (DEB, 2001). Quando o problema de otimização envolve dois ou mais objetivos, diz-se que o problema é de otimização multiobjetivo. Os problemas de otimização multiobjetivo não pertencem ao escopo deste trabalho.

Se as variáveis de decisão assumem valores reais, ou contínuos, tem-se os problemas de otimização contínua, mas se as variáveis assumem valores inteiros, ou discretos, os problemas são de otimização combinatória ou discreta. Existe ainda a classe dos problemas de otimização mista, cujas variáveis assumem valores reais e inteiros, e que não estão na Figura 2.3 (BECCENERI, 2008).

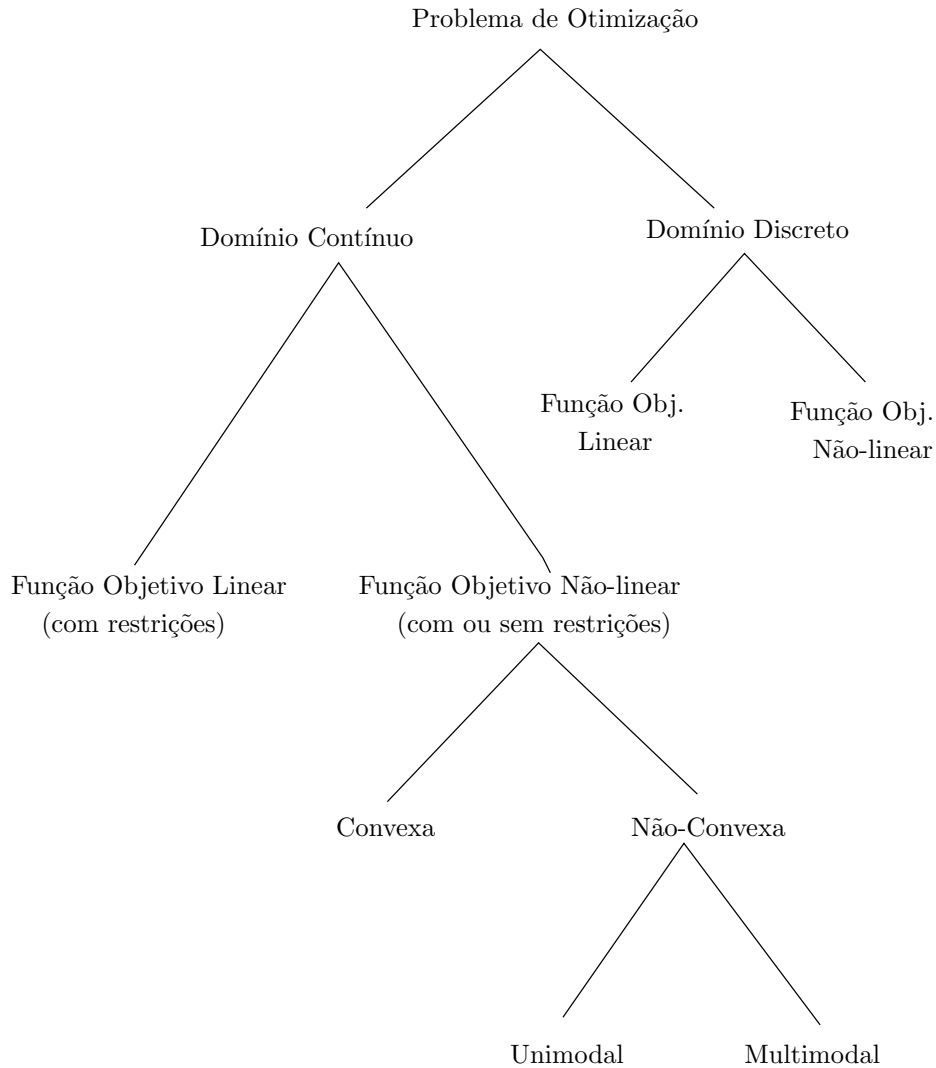


Figura 2.3 - Classificação dos problemas de otimização.  
 Fonte: Adaptado de Zabinsky (2003).

Se as variáveis de decisão estão sujeitas a restrições de igualdade, desigualdade ou laterais, então o problema é de otimização com restrições. Problemas com restrições são comuns em muitos problemas da física e da engenharia e surgem naturalmente de suas modelagens. Se as variáveis não estão sujeitas a nenhuma restrição, então o problema é de otimização irrestrita. Problemas de otimização com restrições podem ser escritos como problemas de otimização irrestrita substituindo as restrições por penalidades na função objetivo (IZMAILOV; SOLODOV, 2007).

Se a função objetivo é linear e as restrições podem ser escritas sob a forma de equações ou inequações lineares, então o problema é de programação linear, mas basta que a função ou uma das restrições seja não-linear nas variáveis de decisão para que

o problema seja definido como problema de otimização não-linear ou programação não-linear (PILLO; PALAGI, 2002a).

Se a função objetivo e as restrições, caso existam, forem convexas e estiverem definidas em um conjunto convexo, então o problema é de otimização convexa e, neste caso, o mínimo local é também o mínimo global. Toda função que não é convexa é dita não-convexa e, neste caso, tem-se os problemas de otimização não-convexa que possuem vários mínimos locais (IZMAILOV; SOLODOV, 2007).

Os problemas de **otimização global** se encaixam na Figura 2.3 nos problemas discretos com função objetivo não-linear, e nos problemas contínuos com função objetivo não-linear e não-convexa (ZABINSKY, 2003). Esses problemas são caracterizados por serem multimodais e de natureza geral, em que a função objetivo e as restrições podem ser não-convexas, não-lineares, não-diferenciáveis e até mesmo descontínuas. A dificuldade em garantir que a solução encontrada é uma solução global é um desafio para os métodos destinados à resolução dessa classe de problemas (ZABINSKY, 2003).

Ao mesmo tempo em que os problemas de otimização global podem possuir características comuns aos problemas clássicos, muitos deles podem se enquadrar na categoria de problemas do tipo caixa preta, em que a função objetivo não é avaliada explicitamente por uma função matemática, mas por meio de simulação numérica de um modelo (ZABINSKY, 2003). Dessa forma, a área de otimização global precisa de novos paradigmas de modelagem e métodos de resolução combinados com conceitos e técnicas clássicas de otimização (PINTER, 2002).

A seguir, alguns métodos determinísticos clássicos baseados em gradiente são descritos brevemente.

## 2.6 Métodos para otimização irrestrita

Nesta seção são descritos os principais métodos de otimização para problemas irrestritos e não-lineares baseados em gradiente. Por simplicidade, considera-se funções objetivo  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  duas vezes diferenciáveis no  $\mathbb{R}^n$ , mesmo que para alguns métodos apenas derivadas contínuas de primeira ordem sejam necessárias.

Uma estratégia geral para a resolução de problemas de otimização irrestrita da forma 2.1 é dada pelo procedimento iterativo

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k, \quad (2.11)$$

em que  $\mathbf{d}^k \in \mathbb{R}^n$  é a direção de busca,  $\alpha^k \in \mathbb{R}$  é o tamanho do passo dado ao longo da direção  $\mathbf{d}^k$ , e  $\mathbf{x}^k \in \mathbb{R}^n$  representa o vetor de variáveis de decisão do problema na iteração  $k$ .

Os métodos de otimização diferem na forma em que a direção e o tamanho do passo são escolhidos. Existem métodos que utilizam informações do gradiente da função objetivo para orientar a busca a cada iteração, como o método da máxima descida e o método dos gradientes conjugados. Outros métodos utilizam informações da matriz Hessiana, como o método de Newton. Já os métodos quase-Newton, como o método Davidon-Fletcher-Powell (DFP), utilizam informações do gradiente nas sucessivas iterações para construir uma aproximação da inversa da matriz Hessiana. E, existem ainda, os métodos que não fazem uso de informações de derivadas, como o método das direções conjugadas e o método simplex, entre outros (VANDERPLAATS, 1984).

A maioria dos métodos que utilizam a estratégia 2.11 determinam o tamanho do passo  $\alpha^k$  na direção de busca  $\mathbf{d}^k$ , ou seja, esses métodos realizam uma busca pelo mínimo da função objetivo localizado na linha determinada por esta direção. O processo de determinação do ponto de mínimo em uma dada linha é chamado busca em linha ou busca linear (*line search*) (LUENBERGER, 1984).

### 2.6.1 Busca linear

As técnicas de busca linear utilizadas em parceria com o procedimento iterativo 2.11 requerem que  $\mathbf{d}^k$  seja uma direção de descida, ou seja,

$$\nabla f(\mathbf{x}^k)\mathbf{d}^k < 0.$$

Logo, se  $\mathbf{d}^k$  é uma direção de descida, então existe  $\alpha^k > 0$  tal que

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

As técnicas de busca linear podem ser classificadas em exatas e inexatas. Uma técnica de busca linear inexata bem difundida é a de redução de um intervalo de incerteza. O método da seção áurea é um exemplo desta técnica, em que a partir de um intervalo  $[a, b]$  que cerca o mínimo de uma função unimodal, pontos interiores são obtidos de acordo com a proporção áurea e são realizadas substituições de forma a reduzir progressivamente o comprimento do intervalo de incerteza original até uma tolerância desejada (LUENBERGER, 1984).



O método da seção áurea pode ser estendido para o caso em que a função está definida na semi-reta  $\mathbf{x}^k + \alpha^k \mathbf{d}^k$  em vez de um intervalo limitado, e também para funções multimodais. Neste último caso, o método retorna apenas uma solução local do problema de otimização. Embora garantida, a convergência (local) do método da seção áurea é, em geral, lenta (IZMAILOV; SOLODOV, 2007).

Na busca linear exata, o tamanho do passo  $\alpha^k$  é obtido da minimização da função objetivo na semi-reta  $\mathbf{x}^k + \alpha^* \mathbf{d}^k$ , com  $\alpha^* \geq 0$ , ou seja, a obtenção de  $\alpha^k$  pode ser formulada como um problema de minimização unidimensional dado por

$$\alpha^k = \arg \min_{\alpha^* \geq 0} f(\mathbf{x}^k + \alpha^* \mathbf{d}^k). \quad (2.12)$$

Quando a função objetivo é quadrática dada por

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{b}^\top \mathbf{x}, \quad (2.13)$$

em que  $\mathbf{b} \in \mathbb{R}^n$  e  $\mathbf{A} \in \mathbb{R}^{n \times n}$  é uma matriz simétrica definida positiva, então é possível determinar o tamanho do passo  $\alpha^k$  que fornece exatamente o mínimo de  $f$  ao longo da direção  $\mathbf{d}^k$  explicitamente pela expressão (IZMAILOV; SOLODOV, 2007)

$$\alpha^k = -\frac{\mathbf{d}^{k \top} \mathbf{d}^k}{\mathbf{d}^{k \top} \mathbf{A} \mathbf{d}^k}. \quad (2.14)$$

Com exceção deste último caso, a resolução de um problema de minimização unidimensional a cada iteração do procedimento iterativo 2.11 pode requerer muitas avaliações da função objetivo e se tornar uma estratégia computacionalmente cara. Uma estratégia menos custosa e mais utilizada na prática é a busca linear inexata, em que um tamanho de passo  $\alpha^k$  é escolhido de forma a garantir apenas um decréscimo suficiente da função objetivo.

Em geral, os métodos de busca linear inexata possuem critérios para o término do procedimento de busca, antes que a convergência seja alcançada. Um critério prático e popular é a regra de Armijo, cuja principal ideia é a de garantir que o tamanho do passo  $\alpha^k$  não seja tão grande e nem tão pequeno ao longo do procedimento de busca pelo mínimo em uma direção. Dessa forma, considere a função objetivo  $f$ , o tamanho do passo  $\alpha$ , e os parâmetros  $0 < \varepsilon < 1$  e  $\eta > 1$ . Por um lado, o tamanho

do passo  $\alpha$  não será tão grande se (LUENBERGER, 1984)

$$f(\mathbf{x}^k + \alpha \mathbf{d}^k) \leq f(\mathbf{x}^k) + \varepsilon \alpha \nabla f(\mathbf{x}^k) \mathbf{d}^k, \quad (2.15)$$

e, por outro lado, o tamanho do passo não será tão pequeno se

$$f(\mathbf{x}^k + \eta \alpha \mathbf{d}^k) > f(\mathbf{x}^k) + \varepsilon \eta \alpha \nabla f(\mathbf{x}^k) \mathbf{d}^k.$$

Esta última desigualdade significa que, se  $\alpha$  é aumentado por um fator  $\eta$ , então a condição 2.15 será avaliada.

A regra de Armijo também é usada para definir uma busca linear inexata simplificada. Neste caso, inicia-se a busca com um valor arbitrário para o tamanho do passo  $\alpha$ . Se  $\alpha$  satisfizer a condição 2.15, então  $\alpha$  é aumentado repetidamente por  $\eta$  ( $\eta = 2$  ou  $\eta = 10$ , e  $\varepsilon = 0,2$  são frequentemente utilizados) até que a condição 2.15 não seja mais satisfeita, e o último valor de  $\alpha$  que satisfizer 2.15 será o tamanho do passo selecionado. Mas se  $\alpha$  não satisfizer a condição 2.15, então  $\alpha$  será repetidamente dividido por  $\eta$  até que a desigualdade 2.15 seja satisfeita (LUENBERGER, 1984).

As técnicas de busca linear, que na verdade são procedimentos para a resolução de problemas de minimização unidimensionais, formam a espinha dorsal dos algoritmos de programação não-linear, pois os problemas de dimensões mais elevadas são resolvidos simplesmente pela execução de uma sequência de buscas lineares consecutivas ao longo de linhas convenientemente escolhidas (LUENBERGER, 1984).

### 2.6.2 Método da máxima descida

O método da máxima descida, também conhecido como *steepest descent method*, é considerado um método básico em otimização irrestrita (PILLO; PALAGI, 2002b). Neste método, a direção de busca  $\mathbf{d}^k$  é dada pela direção contrária à direção do gradiente da função objetivo no ponto  $\mathbf{x}^k$ , ou seja, a direção de busca é a direção do anti-gradiente da função expressa por

$$\mathbf{d}^k = -\nabla f(\mathbf{x}^k). \quad (2.16)$$

Dessa forma, o método da máxima descida utiliza o procedimento iterativo 2.11 com direção de busca 2.16 e tamanho do passo  $\alpha^k$  que pode ser determinado por alguma estratégia de busca linear. O Algoritmo 1, a seguir, ilustra um possível algoritmo para o método da máxima descida.

---

**Algoritmo 1:** Algoritmo para o método da máxima descida

---

**Dados**  $f(\mathbf{x})$  contínua e diferenciável com  $\mathbf{x} \in \mathbb{R}^n$  e um ponto inicial  $\mathbf{x}^0$

- 1: **Faça**  $k = 0$
  - 2: **Enquanto**  $\nabla f(\mathbf{x}^k) \neq 0$ , **faça**
  - 3:      $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$
  - 4:     Obtenha  $\alpha^k > 0$  tal que  $f(\mathbf{x}^k + \alpha^k \mathbf{d}^k) < f(\mathbf{x}^k)$
  - 5:      $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
  - 6:      $k = k + 1$
  - 7: **Retorne**  $\mathbf{x}^k$
- 

Para funções objetivo mal escalonadas, com curvas de nível elipsoidais alongadas, o método da máxima descida apresenta (no caso bidimensional) um comportamento em ziguezague ao caminhar na direção do mínimo, como ilustra a Figura 2.4.

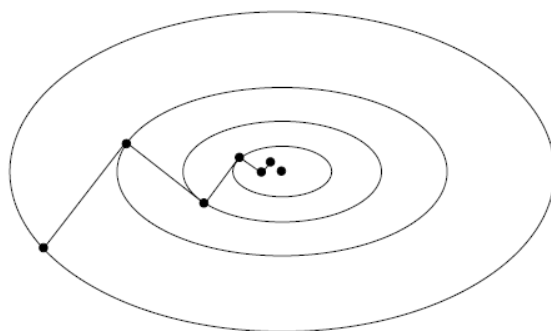


Figura 2.4 - Convergência lenta e comportamento em ziguezague do método da máxima descida.

Fonte: Extraído de Nocedal e Wright (2006)

Embora a convergência do método da máxima descida seja muito lenta, ela é garantida para estratégias adequadas do tamanho do passo. Outras vantagens incluem o uso de apenas informações das derivadas primeiras, e custo computacional e de armazenamento reduzidos, por iteração (PILLO; PALAGI, 2002b). Vale ressaltar que o método da máxima descida é frequentemente incorporado em outras técnicas de otimização, servindo como ponto de partida de métodos mais elaborados.

### 2.6.3 Método de Newton

Seja o problema de otimização irrestrita 2.1, o princípio do método de Newton é minimizar a função  $f$  através de uma aproximação local por uma função quadrática.

Na vizinhança de  $\mathbf{x}^k$  a função  $f$  pode ser aproximada pela sua expansão em série de Taylor até 2ª ordem, dada por

$$f(\mathbf{x}) = f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^k)^\top \nabla^2 f(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k),$$

em que  $\nabla f(\mathbf{x}^k)$  é o vetor gradiente e  $\nabla^2 f(\mathbf{x}^k)$  é a matriz Hessiana de  $f$  no ponto  $\mathbf{x}^k$ . Assumindo que a inversa da matriz Hessiana existe, o lado direito da expressão anterior é minimizado por

$$\mathbf{x}^{k+1} = \mathbf{x}^k - [\nabla^2 f(\mathbf{x}^k)]^{-1} \nabla f(\mathbf{x}^k), \quad (2.17)$$

e este esquema iterativo é conhecido como método de Newton puro (LUENBERGER, 1984; BAZARAA; SHETTY, 1979).

Seja  $\mathbf{x}^0$  um ponto inicial próximo do mínimo local  $\mathbf{x}^*$  da função  $f$ , e seja  $\nabla^2 f(\mathbf{x}^*)$  a matriz Hessiana em  $\mathbf{x}^*$  definida positiva, então os pontos gerados pelo método de Newton convergem para  $\mathbf{x}^*$  com ordem de convergência no mínimo dois (LUENBERGER, 1984). Se a função  $f$  for quadrática da forma 2.13 com matriz Hessiana definida positiva, então o método de Newton encontrará o mínimo local (que é também o mínimo global) em uma única iteração, com passo de tamanho unitário. No entanto, quando  $\mathbf{x}^0$  está longe de  $\mathbf{x}^*$ , o método de Newton pode não convergir para o mínimo local e, portanto, requer modificações (LUENBERGER, 1984).

Uma modificação possível é a introdução de um tamanho de passo  $\alpha^k$  tal que

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k [\nabla^2 f(\mathbf{x}^k)]^{-1} \nabla f(\mathbf{x}^k), \quad (2.18)$$

em que  $\alpha^k$  pode ser obtido por alguma técnica de busca linear (LUENBERGER, 1984).

Esta última modificação, porém, não apresenta qualquer garantia de convergência. Uma modificação que garanta convergência, independentemente do ponto inicial, deve testar a positividade da matriz Hessiana  $\nabla^2 f(\mathbf{x}^k)$  a cada iteração. Inicialmente, considere o procedimento iterativo

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k \mathbf{M}_k \nabla f(\mathbf{x}^k), \quad (2.19)$$

onde  $\mathbf{M}_k$  é uma matriz de ordem  $n \times n$  e  $\alpha^k$  é um tamanho de passo. Logo, o decréscimo da função  $f$ , para  $\alpha^k$  positivo e suficientemente pequeno, acontecerá apenas se  $\nabla f(\mathbf{x}^k) \mathbf{M}_k \nabla f(\mathbf{x}^k)^\top > 0$ , ou seja, se  $\mathbf{M}_k$  for definida positiva. Uma escolha comum é  $\mathbf{M}_k = [\varepsilon_k \mathbf{I} + \nabla^2 f(\mathbf{x}^k)]^{-1}$ , para  $\varepsilon_k > 0$  arbitrário. Esta escolha

pode ser considerada um compromisso entre o método da máxima descida ( $\varepsilon_k$  muito grande) e o método de Newton ( $\varepsilon_k = 0$ ). Além disso, é sempre possível obter um  $\varepsilon_k$  que transforma  $\mathbf{M}_k$  em uma matriz definida positiva (LUENBERGER, 1984).

Embora o método de Newton possua uma taxa de convergência de ordem dois, ele deve ser modificado para garantir convergência, além de ser custoso do ponto de vista computacional, pois requer a avaliação, o armazenamento e a inversão da matriz Hessiana de  $f$  em cada iteração.

#### 2.6.4 Método dos gradientes conjugados

O método dos gradientes conjugados foi desenvolvido com o objetivo de acelerar a taxa de convergência do método da máxima descida e, ao mesmo tempo, evitar o alto custo computacional do método de Newton. Dessa forma, ele pode ser considerado um procedimento intermediário entre estes dois últimos métodos (LUENBERGER, 1984).

Tal como o método de Newton, o método dos gradientes conjugados foi desenvolvido para obter soluções exatas em um número finito de iterações quando a função objetivo é quadrática da forma 2.13, com matriz  $\mathbf{A}$  simétrica definida positiva. Para funções mais gerais, as mesmas técnicas são estendidas e espera-se que a convergência seja similar, pois próximo do ponto de mínimo toda função é aproximadamente quadrática (LUENBERGER, 1984).

Inicialmente, considere o método das direções conjugadas para a minimização de uma função quadrática da forma 2.13, com matriz  $\mathbf{A}$  simétrica definida positiva. Este método constrói um conjunto de  $n$  direções  $\{\mathbf{d}^0, \dots, \mathbf{d}^{n-1}\}$   $\mathbf{A}$ -ortogonais, ou seja, direções conjugadas com respeito à matriz  $\mathbf{A}$  em que

$$(\mathbf{d}^i)^\top \mathbf{A} \mathbf{d}^j = 0, \quad \forall i \neq j.$$

Para qualquer matriz  $\mathbf{A}$  de ordem  $n \times n$  simétrica positiva definida, qualquer conjunto de direções  $\mathbf{A}$ -ortogonais é linearmente independente. A importância desta propriedade de conjugação está na garantia de que todo método de direções conjugadas encontra a solução do problema de minimização irrestrita de uma função quadrática da forma 2.13 com matriz  $\mathbf{A}$  definida positiva em, no máximo,  $n$  iterações (IZMAILOV; SOLODOV, 2007).

O método dos gradientes conjugados, por sua vez, é o método das direções conjugadas em que as direções de busca correspondem a uma versão conjugada dos

sucessivos gradientes obtidos ao longo do procedimento iterativo. Logo, considere o procedimento iterativo 2.11 e um ponto inicial  $\mathbf{x}^0 \in \mathbb{R}^n$ . As direções de busca do método dos gradientes conjugados são obtidas por

$$\mathbf{d}^0 = -\nabla f(\mathbf{x}^0), \quad (2.20)$$

e

$$\mathbf{d}^k = -\nabla f(\mathbf{x}^k) + \delta^{k-1} \mathbf{d}^{k-1}, \quad k > 0, \quad (2.21)$$

com

$$\delta^{k-1} = \frac{\mathbf{g}^k \top \mathbf{A} \mathbf{d}^{k-1}}{\mathbf{d}^{k-1 \top} \mathbf{A} \mathbf{d}^{k-1}}, \quad (2.22)$$

onde  $\mathbf{g}^k = \nabla f(\mathbf{x}^k)$ .

Note que, a primeira direção de busca é a direção de máxima descida. Em todas as demais iterações, a direção de busca será a soma da direção contrária à direção do vetor gradiente no ponto atual com a combinação linear das direções de busca anteriores.

Dentre outras fórmulas para o cálculo de  $\delta^{k-1}$ , uma expressão muito utilizada é a definida por Fletcher e Reeves (1964) e dada por

$$\delta^{k-1} = \frac{\|\nabla f(\mathbf{x}^k)\|^2}{\|\nabla f(\mathbf{x}^{k-1})\|^2}. \quad (2.23)$$

O Algoritmo 2 do método dos gradientes conjugados para funções não-lineares, adaptado de Fletcher e Reeves (1964), é apresentado a seguir. Considere, para a obtenção do tamanho do passo  $\alpha^k$ , qualquer técnica de busca linear inexata.

---

**Algoritmo 2:** Algoritmo do método dos gradientes conjugados para funções não-lineares

---

**Dados**  $f(\mathbf{x})$  contínua e diferenciável com  $\mathbf{x} \in \mathbb{R}^n$  e um ponto inicial  $\mathbf{x}^0$

- 1: **Faça**  $k = 0$
  - 2: **Enquanto**  $\nabla f(\mathbf{x}^k) \neq 0$ , **faça**
  - 3:   **Se**  $k = 0$  **então**  $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$  e vá para o passo 6
  - 4:   Calcule  $\delta^{k-1}$  pela Equação 2.23
  - 5:    $\mathbf{d}^k = -\nabla f(\mathbf{x}^k) + \delta^{k-1} \mathbf{d}^{k-1}$
  - 6:   Obtenha  $\alpha^k$  tal que  $f(\mathbf{x}^k + \alpha^k \mathbf{d}^k) < f(\mathbf{x}^k)$
  - 7:    $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
  - 8:    $k = k + 1$
  - 9: **Retorne**  $\mathbf{x}^k$
-

No método dos gradientes conjugados, a direção  $\mathbf{d}^k$  é obtida apenas com base na direção anterior  $\mathbf{d}^{k-1}$ . Com isso, não é preciso armazenar, na iteração  $k$ , todas as direções anteriores  $\mathbf{d}^0, \mathbf{d}^1, \dots, \mathbf{d}^{k-2}$  e, dessa forma, o custo de armazenamento do método se torna baixo. Outra vantagem reside na formulação simples usada na obtenção das direções de busca. Esta simplicidade faz com que o método dos gradientes conjugados seja apenas um pouco mais complexo que o método da máxima descida. Por fim, o fato das direções serem baseadas em gradientes, o processo exibe a cada passo um progresso uniforme em direção à solução. Apesar de um progresso uniforme não ser de grande importância para um problema puramente quadrático, é importante para generalizações de problemas não-quadráticos (LUENBERGER, 1984).

Devido à presença de termos não-quadráticos em funções objetivos mais gerais, o método dos gradientes conjugados pode apresentar perda de conjugação ao longo das iterações e gerar direções ineficientes. Uma estratégia usada para contornar este problema e assegurar as propriedades de convergência do método é o uso da direção de máxima descida como direção de busca ( $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$ ) sempre que

$$|\nabla f(\mathbf{x}^k)^\top \nabla f(\mathbf{x}^{k-1})| > \omega \|\nabla f(\mathbf{x}^{k-1})\|^2,$$

com  $0 < \omega < 1$  (PILLO; PALAGI, 2002b).

### 2.6.5 Métodos quase-Newton

O custo computacional elevado do método de Newton se deve à avaliação e uso da matriz Hessiana de  $f$ . A ideia por trás dos métodos quase-Newton é realizar uma aproximação iterativa da inversa da matriz Hessiana.

Os métodos quase-Newton utilizam o procedimento iterativo 2.11 com direção de busca dada por

$$\mathbf{d}^k = -\mathbf{B}_k \nabla f(\mathbf{x}^k), \tag{2.24}$$

em que  $\mathbf{B}_k$  é uma matriz simétrica definida positiva. Note que, se  $\mathbf{B}_k = I_n$  ( $I_n$  é a matriz identidade de ordem  $n$  por  $n$ ), a direção de busca  $\mathbf{d}^k$  do método quase-Newton se transforma na direção do método da máxima descida e, se  $\mathbf{B}_k = [\nabla^2 f(\mathbf{x}^k)]^{-1}$ , a direção de busca se transforma na direção do método de Newton.

Os métodos quase-Newton diferem entre si na forma como as atualizações da matriz  $\mathbf{B}_k$  são realizadas. Essa matriz deve se aproximar da inversa da matriz Hessiana de

$f$  a cada iteração  $k$ . Logo, dados

$$\mathbf{r}^k = \mathbf{x}^{k+1} - \mathbf{x}^k,$$

e

$$\mathbf{s}^k = \nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k),$$

a matriz  $\mathbf{B}_k$  pode ser atualizada pela expressão

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{\mathbf{r}^k(\mathbf{r}^k)^\top}{(\mathbf{r}^k)^\top \mathbf{s}^k} - \frac{\mathbf{B}_k \mathbf{s}^k (\mathbf{B}_k \mathbf{s}^k)^\top}{(\mathbf{s}^k)^\top \mathbf{B}_k \mathbf{s}^k} + c(\mathbf{s}^k)^\top \mathbf{B}_k \mathbf{s}^k \mathbf{v}^k (\mathbf{v}^k)^\top, \quad (2.25)$$

com

$$\mathbf{v}^k = \frac{\mathbf{r}^k}{(\mathbf{r}^k)^\top \mathbf{s}^k} - \frac{\mathbf{B}_k \mathbf{s}^k}{(\mathbf{s}^k)^\top \mathbf{B}_k \mathbf{s}^k}.$$

Para  $c = 0$  na expressão 2.25, tem-se a fórmula de Davidon-Fletcher-Powell (DFP). Já para  $c = 1$ , tem-se a fórmula de Broyden-Fletcher-Goldfarb-Shanno (BFGS) (PILLO; PALAGI, 2002b).

Dessa forma, o procedimento iterativo com a direção de busca 2.24 requer, além do ponto inicial  $\mathbf{x}^0$ , uma matriz  $\mathbf{B}_0$  simétrica definida positiva. Em geral, toma-se  $\mathbf{B}_0 = I_n$ . Para completar o esquema iterativo, o tamanho do passo  $\alpha^k$  deve ser obtido por uma técnica de busca linear exata ou inexata. Neste último caso, o comprimento  $\alpha^k$  deve ser positivo e garantir decréscimo suficiente da função (IZMAILOV; SOLODOV, 2007).

Quando a função é quadrática da forma 2.13, com a matriz  $\mathbf{A}$  definida positiva e o tamanho do passo  $\alpha^k$  é dado por busca linear exata, os métodos quase-Newton DFP e BFGS possuem convergência finita. No caso não-quadrático, a convergência finita não é mais assegurada. Porém, quando a convergência ocorre, a sua taxa é frequentemente superlinear (IZMAILOV; SOLODOV, 2007). Embora o custo computacional de métodos quase-Newton seja relativamente baixo, para problemas de grande escala o armazenamento da matriz  $\mathbf{B}_k$  na memória do computador pode inviabilizar o uso desses métodos (PILLO; PALAGI, 2002b).

## 2.7 Métodos para otimização global

Os métodos para problemas de otimização global são frequentemente classificados em métodos determinísticos e métodos estocásticos. Como exemplos de métodos determinísticos, tem-se os métodos *branch-and-bound*, método de planos-de-corte,



métodos de decomposição primal-dual e métodos de análise intervalar (FLOUDAS, 2002). Entre os métodos estocásticos, destacam-se busca aleatória, recozimento simulado, algoritmos genéticos, colônia de formigas e enxame de partículas.

Métodos determinísticos são, em geral, aplicados em problemas de otimização global com estruturas especiais. Estratégias do tipo *branch-and-bound* são bastante empregadas e consistem em partições sucessivas do problema original em subproblemas. Cada subproblema pode ser formulado na forma de um problema convexo relaxado, em que é possível encontrar o mínimo global para uma precisão desejada. Estes mínimos globais representam limites inferiores para o mínimo global do problema não-convexo original (ADJIMAN; FLOUDAS, 2009; GOUNARIS; FLOUDAS, 2009; ZHU, 2009). Teoricamente, os métodos determinísticos garantem convergência de acordo com uma precisão desejada. No entanto, envolvem operações complexas e o custo computacional aumenta com a dimensão do problema, tornando o seu uso proibitivo em muitos problemas reais.

Métodos estocásticos podem ser aplicados em qualquer tipo de função, desde as funções com estruturas especiais até funções com estruturas desconhecidas e que são calculadas apenas por meio de simulação numérica. Esses métodos são caracterizados por possuírem algum tipo de aleatoriedade durante o procedimento de otimização, como o uso de um gerador de números pseudo-aleatórios. Métodos estocásticos têm mostrado eficiência na resolução de problemas de grande escala, em um tempo aceitável, embora não garantam otimalidade da solução. São de fácil implementação, no entanto dependem de parâmetros de ajuste que muitas vezes são determinados por tentativa e erro (ZABINSKY, 2003).

O nome heurística é derivado da palavra grega *heuriskein* e significa descobrir. Hoje o termo heurística é usado para descrever um método “que, baseado na experiência ou julgamento, parece conduzir a uma boa solução de um problema, mas que não garante produzir uma solução ótima”. Já o termo meta-heurística deriva da composição das palavras gregas heurística, explicada anteriormente, e meta, que significa “após”, indicando um nível superior de descoberta. Muitas definições sobre o termo meta-heurísticas podem ser encontradas e, dentre elas, tem-se que “Uma metaheurística é um conjunto de conceitos que podem ser usados para definir métodos heurísticos que podem ser aplicados a um conjunto amplo de diferentes problemas. Em outras palavras, uma metaheurística pode ser vista como uma ferramenta algorítmica geral que pode ser aplicada a diferentes problemas de otimização com modificações relativamente pequenas, para torna-la adaptável a um problema específico” (BECCENERI,

2008).

A seguir, as meta-heurísticas Recozimento Simulado e Algoritmos Genéticos são apresentadas brevemente.

### 2.7.1 Recozimento Simulado

O Recozimento Simulado (RS), do inglês *Simulated Annealing*, foi introduzido por Kirkpatrick et al. (1983) com base no trabalho de Metropolis et al. (1953). Usando uma analogia com o processo físico de têmpera (*annealing*) presente no resfriamento lento de metais submetidos a altas temperaturas, e motivados para resolver o problema de otimização combinatorial do caixeiro viajante, Kirkpatrick et al. (1983) criaram um novo algoritmo de otimização em que movimentos de subida (ou de piora no valor da função objetivo) são aceitos probabilisticamente (GOLDEN; WASIL, 2002).

A principal vantagem do RS é justamente a presença de um mecanismo que permite que o método escape de mínimos locais. Soluções geradas na vizinhança do ponto atual da busca e que melhoram o valor da função objetivo (movimentos de descida) são sempre aceitas, e as soluções piores (movimentos de subida) são aceitas com uma certa probabilidade controlada pelo parâmetro  $T$ , que representa a temperatura. A temperatura decresce a cada iteração  $k$ , em geral, seguindo a regra  $T^{k+1} = \tau \cdot T^k$  em que  $\tau$  é o fator de decréscimo ou resfriamento.

No início do procedimento iterativo, a temperatura é relativamente alta e a probabilidade de aceitar soluções piores é maior. Ao longo das iterações, a temperatura decresce e a probabilidade de aceitar soluções piores diminui. É essa característica de aceitar movimentos de subida de forma controlada que permite que o algoritmo escape de mínimos locais e forneça soluções cada vez melhores.

A seguir, o Algoritmo 3 para o método de Recozimento Simulado, adaptado de Golden e Wasil (2002), é apresentado. Considere  $\mathcal{F}$  o espaço de busca viável,  $\mathbf{x}'$  uma solução candidata em  $\mathcal{F}$ , e os parâmetros de ajuste  $T^0$  (temperatura inicial) e  $\tau$  (fator de resfriamento). Usualmente, o critério de parada é dado por um número máximo de avaliações da função objetivo. O algoritmo retorna em  $\mathbf{x}_{melhor}$  a solução com o menor valor da função objetivo encontrada ao longo da busca.

A temperatura inicial  $T^0$  e o cronograma de resfriamento são fundamentais no desempenho do algoritmo RS. Para valores altos de  $T^0$ , a probabilidade de aceitar soluções piores também é alta e, neste caso, o algoritmo explora o espaço de busca.

---

**Algoritmo 3:** Algoritmo do Recozimento Simulado

---

Dados  $T^0$ ,  $\tau$  e  $\mathbf{x}^0 \in \mathcal{F}$

- 1: **Faça**  $k = 0$
  - 2: **Faça**  $\mathbf{x}_{melhor} = \mathbf{x}^0$
  - 3: **Enquanto** algum critério de parada não for atingido, **faça**
  - 4:   Gere uma solução  $\mathbf{x}'$  na vizinhança de  $\mathbf{x}^k$
  - 5:   Gere  $r \in [0; 1]$  com distribuição uniforme
  - 6:   Calcule  $\Delta = f(\mathbf{x}') - f(\mathbf{x}^k)$
  - 7:   **Se**  $\Delta < 0$  ou  $\exp(-\Delta/T^k) > r$  **então**
  - 8:      $\mathbf{x}^{k+1} = \mathbf{x}'$
  - 9:   **Caso contrário**
  - 10:     $\mathbf{x}^{k+1} = \mathbf{x}^k$
  - 11:    **Se**  $f(\mathbf{x}^{k+1}) < f(\mathbf{x}_{melhor})$  **então**  $\mathbf{x}_{melhor} = \mathbf{x}^{k+1}$
  - 12:     $T^{k+1} = \tau \cdot T^k$
  - 13:     $k = k + 1$
  - 14: **Retorne**  $\mathbf{x}_{melhor}$
- 

Porém, para valores baixos de  $T^0$ , a probabilidade de aceitar soluções piores é menor e o algoritmo desempenha uma busca intensificada ou localizada.

O cronograma de resfriamento determina a forma como a temperatura decai e, conseqüentemente, como ocorre a transição entre busca global e busca local. Por um lado, se o fator de redução  $\tau$  for muito pequeno, a temperatura cairá rapidamente e a busca se tornará localizada em poucas iterações, com uma probabilidade de aceitar movimentos de piora muito pequena. Por outro lado, se  $\tau$  for muito grande, a temperatura cairá lentamente e a busca permanecerá global por várias iterações, pois a probabilidade de aceitar movimentos de piora será grande. No primeiro caso, o algoritmo pode ficar preso em um mínimo local e, no último caso, o algoritmo pode não convergir.

Propriedades de convergência do Recozimento Simulado têm sido analisadas e mostram que o método converge assintoticamente para o mínimo global com probabilidade 1 para um número de iterações muito grande (ZABINSKY, 2003). O método do Recozimento Simulado pode ser aplicado em problemas de variáveis discretas, contínuas ou mistas. Exemplos de aplicações incluem o problema de particionamento e coloração de grafos, projetos de chips VLSI, problema quadrático de alocação, processamento de imagens, entre outros (PARDALOS; MAVRIDOU, 2009).

### 2.7.2 Algoritmos Genéticos

Os algoritmos genéticos (AGs) pertencem a uma classe mais geral de métodos inspirados no processo de evolução natural das espécies, denominada Algoritmos Evolutivos (AEs).

Os AGs foram desenvolvidos por J. H. Holland e seus colaboradores, e são algoritmos baseados em populações compostas por indivíduos que representam variáveis de decisão no espaço de busca viável (HOLLAND, 1992; GOLDBERG, 1989). A principal ideia é permitir a “evolução” dos indivíduos de uma população por meio da aplicação de operadores genéticos como reprodução, mutação e seleção. Espera-se que, após várias gerações, a população seja composta por indivíduos bem adaptados, representativos de boas soluções para o problema de otimização.

O Algoritmo 4, a seguir, é um exemplo de algoritmo genético básico. O critério de parada é, em geral, dado por um número máximo de gerações, mas outros critérios, como um número máximo de avaliações da função objetivo, podem ser aplicados. O algoritmo retorna como solução o indivíduo com melhor aptidão.

---

**Algoritmo 4:** Algoritmo genético básico

---

- 1: **Faça**  $k = 0$
  - 2: Construa uma população inicial de indivíduos  $\mathbf{S}^k$
  - 3: **Enquanto** algum critério de parada não for atingido, **faça**
  - 4:   Avalie os indivíduos de  $\mathbf{S}^k$  e obtenha a aptidão de cada um
  - 5:   Armazene o indivíduo com melhor aptidão
  - 6:   Aplique o operador de cruzamento e crie novos indivíduos para  $\mathbf{S}^k$
  - 7:   Aplique o operador de mutação sobre os novos indivíduos de  $\mathbf{S}^k$
  - 8:   Selecione uma nova população  $\mathbf{S}^{k+1}$  a partir de  $\mathbf{S}^k$
  - 9:    $k = k + 1$
  - 10: **Retorne** o indivíduo com melhor aptidão
- 

A aptidão de um indivíduo é dada de acordo com o valor da função objetivo. Quanto menor é o valor da função objetivo, melhor é a aptidão do indivíduo. Em geral, a cada geração os indivíduos mais adaptados denominados “pais” são selecionados para gerar novos indivíduos denominados “filhos”, ou seja, os pais são submetidos ao operador genético de cruzamento (*crossover*) dando origem aos filhos que, por sua vez, são submetidos ao operador de mutação (*mutation*). A mutação tem o papel de manter a diversidade na população e prevenir que a busca convirja rapidamente para uma solução local, o que é chamado de convergência prematura no jargão da

literatura de AGs.

Os indivíduos ou variáveis de decisão podem ser definidos com codificação binária, inteira ou real. Detalhes sobre os operadores de cruzamento e mutação, estratégias de avaliação e seleção, tamanho da população, taxa de cruzamento, taxa de mutação e intervalo de geração devem ser especificados para definir as particularidades do algoritmo genético.

Os AGs trabalham com uma população em vez de um único ponto e, portanto, permitem a exploração simultânea do espaço de busca. São de fácil implementação, trivialmente paralelizáveis e adaptáveis a problemas com variáveis contínuas, discretas ou mistas, com restrições arbitrárias, com funções objetivo não-lineares, não-convexas, descontínuas, com ou sem ruído, entre outros. Como desvantagens, tem-se a existência de vários parâmetros de ajuste, embora exista uma lista de parâmetros de referência que podem ser empregados em problemas gerais (LACERDA; CARVALHO, 1999).



### 3 MÉTODOS BASEADOS EM $q$ -GRADIENTE

Neste capítulo, são apresentados os conceitos de  $q$ -derivada,  $q$ -gradiente e um algoritmo geral para as  $q$ -versões de métodos baseados em  $q$ -gradiente. Estratégias para obtenção do parâmetro  $q$  e tamanho do passo, dado na direção contrária à direção do  $q$ -gradiente, também são apresentadas.

#### 3.1 $q$ -derivada

O  $q$ -cálculo, do inglês *q-calculus*, surgiu da generalização de funções por meio de um parâmetro  $q$ . Neste contexto, surgiram os  $q$ -números, funções  $q$ -análogas elementares,  $q$ -derivadas,  $q$ -integrals, séries  $q$ -hipergeométricas, polinômios  $q$ -ortogonais, entre outros. Essas generalizações remontam aos trabalhos de Fermat, Euler, Heine e Gauss, mas foi no início do século XX que o  $q$ -cálculo recebeu importantes contribuições. Em particular, contribuições do reverendo inglês Frank Hilton Jackson, que foi o primeiro a desenvolver o  $q$ -cálculo de forma sistemática (CHAUNDY, 1962; ERNST, 2000; KAC; CHEUNG, 2002). Dentre uma série de artigos publicados, F. H. Jackson estudou a generalização de funções, séries e números especiais no contexto do  $q$ -cálculo, além de reintroduzir o operador  $q$ -derivada, amplamente conhecido como derivada de Jackson, e de criar o conceito de  $q$ -integral definida (JACKSON, 1904; JACKSON, 1908; JACKSON, 1910a; JACKSON, 1910b).

Nos últimos anos, o  $q$ -cálculo tem encontrado na física um espaço para diversas aplicações, sobretudo nos campos da teoria quântica e da mecânica estatística, como por exemplo a  $q$ -entropia (ERNST, 2003; BORGES, 2004). Abe (1997) sugeriu o uso da derivada de Jackson como forma de relacionar o  $q$ -cálculo com a entropia de Tsallis por meio da substituição da derivada tradicional pela derivada de Jackson em uma equação utilizada para gerar a entropia de Boltzmann-Gibbs. Com isso, ele mostrou que uma conexão entre a entropia generalizada e a teoria de grupos quânticos apresentadas por Tsallis (1994) pode ser naturalmente entendida no contexto do  $q$ -cálculo. Johal (1998) estabeleceu uma relação entre a derivada de Jackson e a entropia para uma distribuição de probabilidades multifractal. E Lavagnoa e Swamy (2000) mostraram que conceitos da termodinâmica estatística de grupos quânticos  $q$ -deformados podem ser construídos sobre o formalismo do  $q$ -cálculo.

A derivada clássica de uma função diferenciável  $f$  de uma única variável no ponto  $x$  é dada por

$$\frac{df(x)}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}, \quad (3.1)$$

em que  $h$  é um infinitésimo. Substituindo  $x + h$  por  $qx$  ( $q$  é um número real diferente de 1) na última equação e desconsiderando o limite, tem-se a definição do operador  $q$ -derivada<sup>1</sup> dado por

$$D_q f(x) = \frac{f(qx) - f(x)}{qx - x}, \quad (3.2)$$

para  $x \neq 0$  e  $q \neq 1$ . E no limite,  $q \rightarrow 1$ , a  $q$ -derivada retorna à derivada clássica, ou seja,

$$\frac{df(x)}{dx} = \lim_{q \rightarrow 1} \frac{f(qx) - f(x)}{qx - x}. \quad (3.3)$$

A derivada clássica avalia o quanto uma dada função  $f(x)$  é sensível a pequenas *translações* em sua variável independente. A generalização do conceito de derivada desenvolvida por Jackson é baseada em *deformações* na variável independente, ou seja, em vez da variável independente  $x$  ser transladada por uma quantidade  $h$ , ela é dilatada ou contraída por uma quantidade  $q$ .

Seja, por exemplo,  $f(x) = x^n$ , tem-se que a derivada de  $f$  é

$$\frac{df(x)}{dx} = nx^{n-1}.$$

A  $q$ -derivada de  $f$ , por sua vez, é dada por

$$D_q f(x) = \frac{(qx)^n - x^n}{qx - x} = \frac{q^n x^n - x^n}{qx - x} = \frac{(q^n - 1)x^n}{(q - 1)x} = \frac{q^n - 1}{q - 1} x^{n-1}.$$

No entanto, a fração  $(q^n - 1)/(q - 1)$  é a  $q$ -versão de  $n$ , ou seja,

$$[n] = \frac{q^n - 1}{q - 1},$$

para qualquer número  $n$  inteiro e positivo, pois quando  $q \rightarrow 1$  tem-se que  $[n] = q^{n-1} + \dots + 1 \rightarrow 1 + 1 + \dots + 1 = n$ . Portanto,

$$D_q x^n = [n]x^{n-1}.$$

Esta definição é usada para calcular o  $q$ -binômio e estabelecer uma  $q$ -versão da fórmula de Taylor, utilizada em muitos resultados, como a identidade de Euler em funções  $q$ -exponenciais, o  $q$ -binômio de Gauss, a fórmula de Heine para funções  $q$ -hipergeométricas, entre outros resultados matemáticos (KAC; CHEUNG, 2002).

---

<sup>1</sup>O operador  $q$ -derivada também é conhecido como operador  $q$ -diferença, derivada de Jackson, ou simplesmente  $q$ -derivada.



Considerando funções arbitrárias  $f(x)$  e  $g(x)$ , o operador  $q$ -derivada satisfaz as seguintes propriedades (KAC; CHEUNG, 2002):

1) A  $q$ -derivada é um operador linear para quaisquer constantes  $a$  e  $b$

$$D_q(af(x) + bg(x)) = aD_qf(x) + bD_qg(x).$$

2) A  $q$ -derivada do produto de  $f(x)$  por  $g(x)$  é

$$D_q(f(x)g(x)) = f(qx)D_qg(x) + g(x)D_qf(x)$$

e, por simetria, é equivalente a

$$D_q(f(x)g(x)) = f(x)D_qg(x) + g(qx)D_qf(x).$$

3) A  $q$ -derivada do quociente de  $f(x)$  por  $g(x)$  é calculada por

$$D_q\left(\frac{f(x)}{g(x)}\right) = \frac{g(x)D_qf(x) - f(x)D_qg(x)}{g(x)g(qx)}$$

ou, de forma equivalente,

$$D_q\left(\frac{f(x)}{g(x)}\right) = \frac{g(qx)D_qf(x) - f(qx)D_qg(x)}{g(x)g(qx)}.$$

A regra da cadeia para  $q$ -derivadas não existe, exceto para funções da forma  $f(u(x))$ , em que  $u(x) = \alpha x^\beta$ , com  $\alpha, \beta$  constantes. Mais propriedades da  $q$ -derivada podem ser encontradas em Kac e Cheung (2002).

### 3.2 $q$ -gradiente

Para funções diferenciáveis de  $n$  variáveis,  $f(\mathbf{x})$ , se a derivada parcial clássica de primeira ordem com respeito a uma variável  $x_i$  é dada por

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = \lim_{h \rightarrow 0} \frac{f(x_1; \dots; x_i + h; \dots; x_n) - f(x_1; \dots; x_i; \dots; x_n)}{h}, \quad (3.4)$$

então a  $q$ -derivada parcial de primeira ordem com respeito a  $x_i$  é

$$D_{q_i, x_i} f(\mathbf{x}) = \frac{f(x_1; \dots; q_i x_i; \dots; x_n) - f(x_1; \dots; x_i; \dots; x_n)}{q_i x_i - x_i}. \quad (3.5)$$

Quando  $x_i = 0$  ou  $q_i = 1$ ,  $\forall i = 1, 2, \dots, n$ , a  $q$ -derivada parcial de primeira ordem é dada pela derivada parcial clássica de primeira ordem

$$D_{q_i, x_i} f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_i}. \quad (3.6)$$

Logo, a  $q$ -derivada parcial de primeira ordem de  $f$  com respeito à variável  $x_i$  pode ser naturalmente definida como

$$D_{q_i, x_i} f(\mathbf{x}) = \begin{cases} \frac{f(x_1; \dots; q_i x_i; \dots; x_n) - f(x_1; \dots; x_i; \dots; x_n)}{q_i x_i - x_i}, & x_i \neq 0 \text{ e } q_i \neq 1 \\ \frac{\partial f(\mathbf{x})}{\partial x_i}, & x_i = 0 \text{ ou } q_i = 1 \end{cases}. \quad (3.7)$$

Finalmente, se o gradiente clássico de uma função objetivo de  $n$  variáveis  $f(\mathbf{x})$  é o vetor das  $n$  derivadas parciais de primeira ordem de  $f$

$$\nabla f(\mathbf{x})^\top = \left[ \frac{\partial f(\mathbf{x})}{\partial x_1} \quad \dots \quad \frac{\partial f(\mathbf{x})}{\partial x_i} \quad \dots \quad \frac{\partial f(\mathbf{x})}{\partial x_n} \right], \quad (3.8)$$

então o  $q$ -gradiente é o vetor nas  $n$   $q$ -derivadas parciais de primeira ordem de  $f$  dado por

$$\nabla_{\mathbf{q}} f(\mathbf{x})^\top = [D_{q_1, x_1} f(\mathbf{x}) \quad \dots \quad D_{q_i, x_i} f(\mathbf{x}) \quad \dots \quad D_{q_n, x_n} f(\mathbf{x})], \quad (3.9)$$

em que o parâmetro  $q$  é um vetor de  $n$  posições, ou seja,  $\mathbf{q} = (q_1; \dots; q_i; \dots; q_n)$ . A partir de agora denotamos o parâmetro  $q$  por  $\mathbf{q}$  quando a função objetivo se tratar de uma função de  $n$  variáveis.

Quando  $q_i \rightarrow 1 \forall i$ , o vetor  $q$ -gradiente retorna ao vetor gradiente clássico, isto é,

$$\nabla_{\mathbf{q}} f(\mathbf{x})|_{\mathbf{q}=\mathbf{1}} = \nabla f(\mathbf{x}). \quad (3.10)$$

A Figura 3.1 ilustra a interpretação geométrica do vetor gradiente clássico e do vetor  $q$ -gradiente para a função  $f : \mathbb{R} \rightarrow \mathbb{R}$

$$f(x) = 2 - (e^{-x^2} + 2e^{-(x-3)^2}). \quad (3.11)$$

Para funções de uma única variável, a interpretação geométrica do vetor gradiente em um ponto  $x$  é simplesmente a interpretação da derivada em  $x$ . Logo, o módulo

do vetor gradiente em  $x$  é dado pelo coeficiente angular da reta *tangente* à curva de  $f$  nesse ponto e a direção de máximo crescimento será à direita de  $x$  se o coeficiente angular for positivo ou à esquerda se for negativo. Analogamente, o vetor  $q$ -gradiente é dado pela reta *secante* à curva de  $f$  passando pelos pontos  $(x; f(x))$  e  $(qx; f(qx))$ . Se o coeficiente angular da reta secante for positivo (negativo), então a direção de máximo crescimento dada pelo vetor  $q$ -gradiente será à direita (esquerda).

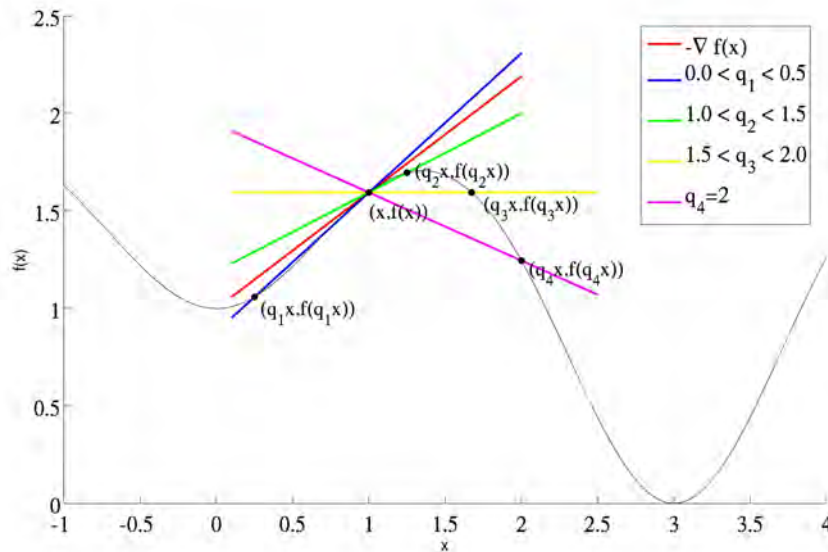


Figura 3.1 - Interpretação geométrica do  $q$ -gradiente para uma função de uma única variável e diferentes valores do parâmetro  $q$ .

O vetor gradiente clássico, ou melhor, a derivada de  $f$  em  $x = 1$  é dada pela reta em vermelho na Figura 3.1, e como o coeficiente angular dessa reta é positivo, então a direção de máximo crescimento da função nesse ponto é à direita. Se o problema é de minimização e a solução é obtida pelo método da máxima descida, então a direção de busca será a direção contrária à direção dada pelo gradiente clássico, também conhecida como direção do anti-gradiente. Dessa forma, o ponto  $x$  irá se mover para a esquerda e ficará necessariamente preso no mínimo local em  $x = 0$ . O vetor  $q$ -gradiente, por sua vez, é dado pelas demais retas passando por  $(x; f(x))$  e  $(q_i x; f(q_i x))$  ( $i = 1, 2, 3, 4$ ) na Figura 3.1, e pode ter coeficiente angular positivo ou negativo, dependendo do parâmetro  $q_i$  utilizado. Por exemplo, para  $q_4 = 2$  a  $q$ -derivada é negativa em  $x = 1$  (reta em magenta), permitindo que a estratégia de minimização baseada no  $q$ -gradiente tome a direita como direção de busca de  $x = 1$  e, portanto, se mova na direção do mínimo global dessa função. Note que existe um

valor  $1,5 < q_3 < 2$  em que o ponto  $x = 1$  é estacionário, isto é, o  $q$ -gradiente é nulo, embora esse ponto não seja nem um ponto de máximo e nem de mínimo da função. Por fim, para  $0 < q_1 < 0,5$  (reta em azul) ou  $1 < q_2 < 1,5$  (reta em verde) o coeficiente angular da reta secante é positivo e o método de otimização baseado em  $q$ -gradiente se move para a esquerda, se comportando como o método da máxima descida.

Este exemplo simples mostra que o uso do vetor  $q$ -gradiente como direção de busca em métodos de otimização oferece mecanismos para escapar de mínimos locais. Além disso, a transição entre busca global e busca local pode ser controlada pelo parâmetro  $\mathbf{q}$ , desde que estratégias adequadas para a sua obtenção e a determinação do tamanho do passo sejam incorporadas no algoritmo de otimização.

### 3.3 Algoritmo para métodos baseados em $q$ -gradiente

O procedimento iterativo a seguir já foi apresentado na Seção 2.6 e é repetido aqui. Trata-se de uma estratégia geral para métodos de otimização, que gera uma sequência de pontos  $\{\mathbf{x}^k\}$  a partir de um ponto inicial  $\mathbf{x}^0$ :

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k, \quad (3.12)$$

em que  $\alpha^k$  é o tamanho do passo ou a distância percorrida ao longo da direção de busca  $\mathbf{d}^k$  na iteração  $k$ .

É intuitivo pensar na direção de busca em um método de otimização como a direção do anti-gradiente da função objetivo no ponto em que se está na busca, pois o vetor gradiente fornece a direção de máximo crescimento da função e, conseqüentemente, sua direção contrária fornece a direção de máximo decréscimo. Logo, para os métodos de otimização baseados no conceito de  $q$ -gradiente, de modo geral, deve-se substituir o vetor gradiente da função objetivo pelo vetor  $q$ -gradiente e utilizar a sua direção contrária, ou direção do anti- $q$ -gradiente, na determinação da direção de busca.

Um algoritmo geral para métodos baseados em  $q$ -gradiente para funções diferenciáveis  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  é apresentado a seguir (Algoritmo 5). O ponto inicial  $\mathbf{x}^0 \in \mathbb{R}^n$  pode ser obtido por meio de sorteio uniforme no espaço de busca. O algoritmo retorna  $\mathbf{x}_{melhor} \in \mathbb{R}^n$  como o menor valor de  $f$  ao longo do procedimento iterativo, isto é,  $f(\mathbf{x}_{melhor}) < f(\mathbf{x}^k), \forall k$ .

Regras para satisfazer o passo 3 do Algoritmo 5 podem se basear nos critérios defi-

---

**Algoritmo 5:** Algoritmo (geral) para métodos baseados em  $q$ -gradiente

---

**Dados**  $f(\mathbf{x})$  contínua e diferenciável com  $\mathbf{x} \in \mathbb{R}^n$  e um ponto inicial  $\mathbf{x}^0$ 

- 1: **Faça**  $k = 0$
  - 2: **Faça**  $\mathbf{x}_{melhor} = \mathbf{x}^0$
  - 3: **Enquanto** não atingir um critério de parada, **faça**
  - 4:   Obtenha o parâmetro  $\mathbf{q}^k$
  - 5:   Obtenha a direção de busca  $\mathbf{d}^k$  com base no vetor  $q$ -gradiente ( $-\nabla_{\mathbf{q}}f(\mathbf{x}^k)$ )
  - 6:   Obtenha o tamanho do passo  $\alpha^k$
  - 7:    $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
  - 8:   **Se**  $f(\mathbf{x}^{k+1}) < f(\mathbf{x}_{melhor})$  **então**  $\mathbf{x}_{melhor} = \mathbf{x}^{k+1}$
  - 9:    $k = k + 1$
  - 10: **Retorne**  $\mathbf{x}_{melhor}$
- 

nidos na Seção 2.4, ou ainda pode-se utilizar como critério de parada a expressão

$$\|\nabla_{\mathbf{q}}f(\mathbf{x}^k)\| \leq \varepsilon, \quad (3.13)$$

para  $\varepsilon > 0$  pequeno, desde que  $q \rightarrow 1$  e o vetor  $q$ -gradiente tenha convergido para o vetor gradiente clássico, pois como visto na Figura 3.1 existem valores para o parâmetro  $\mathbf{q}^k$  e o ponto  $\mathbf{x}^k$  em que o  $q$ -gradiente é nulo, embora  $\mathbf{x}^k$  não represente necessariamente um ponto estacionário da função objetivo. Logo, quando  $q \rightarrow 1$ , métodos baseados em  $q$ -gradiente retomam suas versões clássicas e o critério de parada dado pela Equação 3.13 nada mais é que a condição necessária de otimalidade de primeira ordem.

O mais adequado na escolha do critério de parada é a utilização de várias regras de parada combinadas de forma hierárquica. A definição desta hierarquia e de vários parâmetros envolvidos em métodos de otimização é uma questão de arte e de experiência empírica (IZMAILOV; SOLODOV, 2007).

O uso do  $q$ -gradiente na determinação da direção de busca requer o cálculo dos parâmetros  $q_i$  ( $i = 1, 2, \dots, n$ ). Além disso, uma vez definida a direção de busca, é preciso saber o quanto caminhar nessa direção. A seguir, são descritas algumas estratégias para a obtenção do parâmetro  $\mathbf{q}$  e do tamanho do passo  $\alpha$  que, combinadas, completam o algoritmo para métodos de otimização baseados em  $q$ -gradiente.

### 3.4 Estratégias para cálculo do parâmetro $\mathbf{q}$

De modo geral, uma direção qualquer é considerada de descida se o ângulo formado entre ela e  $-\nabla f$  for menor que  $\pi/2$  radianos (NOCEDAL; WRIGHT, 2006). Por um

lado, quanto mais próximo de  $\mathbf{1}$  o parâmetro  $\mathbf{q}$  for definido, menor será o ângulo formado entre  $-\nabla f$  e  $-\nabla_{\mathbf{q}} f$ . Consequentemente, a direção do anti- $q$ -gradiente será próxima da direção de máxima descida. Por outro lado, para valores de  $\mathbf{q}$  diferentes e não tão próximos de  $\mathbf{1}$ , esse ângulo pode ser qualquer e a direção do anti- $q$ -gradiente pode apontar para qualquer direção do espaço de busca, dependendo das características da função objetivo e do ponto  $\mathbf{x}$  em que se está na busca.

A Figura 3.2 ilustra a direção do anti-gradiente (reta em vermelho) e as direções do anti- $q$ -gradiente (retas em preto) com diferentes valores do parâmetro  $\mathbf{q}$  para a função bidimensional ( $n = 2$ )  $f : \mathbb{R}^n \rightarrow \mathbb{R}$

$$f(\mathbf{x}) = \sum_{i=1}^n (x_i - o_i)^2, \quad (3.14)$$

com  $\mathbf{o} = (10; 10)$  e  $\mathbf{x} = (12; 12)$ . Essa função nada mais é que a função quadrática  $f(\mathbf{x}) = \sum_{i=1}^n x_i^2$  deslocada e com mínimo global em  $\mathbf{o}$ , ou seja,  $f(\mathbf{o}) = \mathbf{0}$ . A função 3.9, representada por meio de curvas de nível em todas as figuras desta seção, ilustra um caso particular do comportamento das direções de anti- $q$ -gradiente.

Para traçar as retas numeradas de 1 a 12 na Figura 3.2, foram tomados  $\mathbf{q}$ 's simétricos em relação ao  $\mathbf{1}$  e, respectivamente, iguais a  $(0, 5; 1, 5)$ ,  $(0, 6; 1, 4)$ ,  $(0, 7; 1, 3)$ ,  $(0, 8; 1, 2)$ ,  $(0, 9; 1, 1)$ ,  $(0, 95; 1, 05)$ ,  $(0, 96; 1, 04)$ ,  $(0, 97; 1, 03)$ ,  $(0, 98; 1, 02)$ ,  $(0, 99; 1, 01)$  e  $(0, 999; 1, 001)$ . Já para a reta 12 tem-se  $\mathbf{q} = (-10; -3)$ .

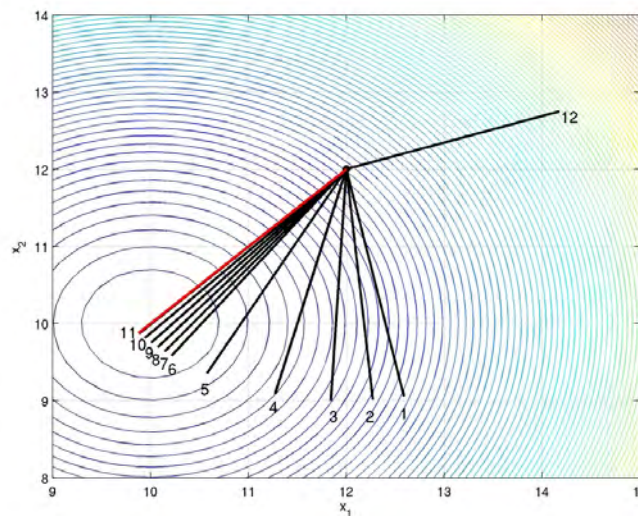


Figura 3.2 - Direções do anti-gradiente (vermelho) e do anti- $q$ -gradiente (preto) para diferentes valores do parâmetro  $\mathbf{q}$ .

À medida que os valores de  $\mathbf{q} = (q_1; q_2)$  se aproximam de 1, à esquerda para a coordenada  $q_1$  e à direita para a coordenada  $q_2$ , a direção do anti- $q$ -gradiente se aproxima da direção do anti-gradiente, como mostram as retas de 1 a 11. Note que a reta 11 e a reta determinada pelo anti-gradiente (reta vermelha) são praticamente retas coincidentes. Finalmente, para a reta 12, a direção do anti- $q$ -gradiente apresenta uma direção que não é de descida.

Para funções unimodais cujas curvas de nível formam elipsóides alongados, o uso de uma direção de busca que é de descida, mas não exatamente igual à direção negativa do gradiente clássico, pode reduzir o movimento de ziguezague durante a busca do mínimo global e acelerar a convergência do método. Para funções multimodais, direções de busca que não apontam necessariamente para a direção de máxima descida podem permitir que a busca escape de mínimos locais, aumentando as chances do algoritmo caminhar na direção do mínimo global da função.

É fácil observar que o parâmetro  $\mathbf{q}$  é um parâmetro chave para métodos baseados em  $q$ -gradiente. Quando  $\mathbf{q}$  é diferente e não tão próximo de  $\mathbf{1}$ , direções de busca dadas pelo anti- $q$ -gradiente podem ser tanto de descida quanto de subida, e os métodos baseados em  $q$ -gradiente exploram o espaço de busca escapando de mínimos locais, uma vez que é possível caminhar em direções de subida. Já quando  $\mathbf{q}$  tende a  $\mathbf{1}$ , as direções de busca dadas pelo anti- $q$ -gradiente tendem à direção do anti-gradiente clássico, e os métodos baseados em  $q$ -gradiente retomam suas respectivas versões clássicas, realizando buscas locais. Um método baseado em  $q$ -gradiente pode ser mais estocástico ou mais determinístico, dependendo dos valores do parâmetro  $\mathbf{q}$  utilizados ao longo do procedimento iterativo.

A estratégia utilizada para a escolha do parâmetro  $\mathbf{q}$  consiste na geração de números aleatórios segundo uma distribuição de probabilidade. As distribuições de probabilidade utilizadas foram:

- i) Uniforme em um intervalo  $[q_{min}; q_{max}]$ ,
- ii) Log-normal com média  $\mu = 1$  e desvio-padrão reduzindo para zero ( $\sigma \rightarrow 0$ ),
- iii) gaussiana com média  $\mu = x_i^k$  e desvio-padrão reduzindo para zero ( $\sigma \rightarrow 0$ ).

### 3.4.1 Estratégia 1: distribuição uniforme

Na estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição uniforme, determina-se um intervalo simétrico em relação ao 1, denominado  $[q_{min}; q_{max}]$  e cada

coordenada  $q_i$  do parâmetro  $\mathbf{q}$  é gerada neste intervalo de acordo com a função densidade de probabilidade uniforme

$$f(q_i; q_{min}, q_{max}) = \begin{cases} \frac{1}{q_{max} - q_{min}} & , q_{min} \leq q_i \leq q_{max}, \\ 0 & , \text{ caso contrário.} \end{cases} \quad (3.15)$$

A Figura 3.3 ilustra a função densidade de probabilidade de uma distribuição uniforme com diferentes valores de  $[q_{min}; q_{max}]$ . Note que a probabilidade de ocorrência é a mesma para qualquer valor no intervalo  $[q_{min}; q_{max}]$ . Logo, quanto mais próximo de 1 este intervalo for definido, mais próximo de 1 será o parâmetro  $q_i$  gerado segundo esta distribuição.

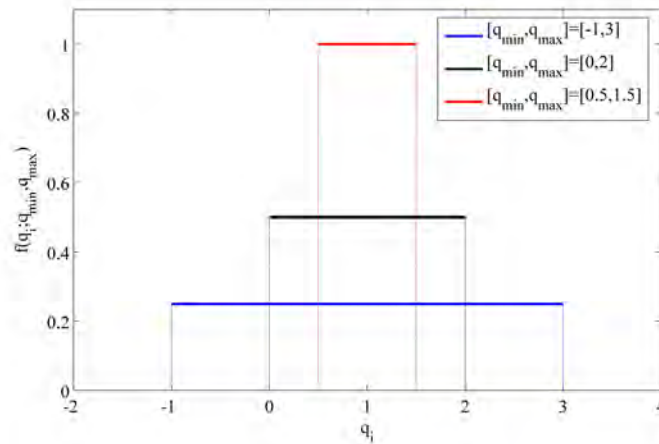


Figura 3.3 - Função densidade de probabilidade uniforme para diferentes valores de  $[q_{min}; q_{max}]$ .

Para ilustrar as possíveis direções de busca obtidas a partir dessa estratégia, foram gerados parâmetros  $\mathbf{q}$ 's uniformemente distribuídos nos intervalos  $[0; 2]$ ,  $[0, 5; 1, 5]$ ,  $[0, 9; 1, 1]$  e  $[0, 99; 1, 01]$ . Para cada um dos 200 parâmetros gerados em cada intervalo, foram traçadas as respectivas 200 direções do anti- $q$ -gradiente para a função 3.14, com  $\mathbf{o} = (10; 10)$  e no ponto  $\mathbf{x} = (11; 11)$ , como mostra a Figura 3.4. A semi-reta em vermelho representa a direção de anti-gradiente.

Note que, para o intervalo  $[0; 2]$ , as direções do anti- $q$ -gradiente apontam para qualquer direção do espaço de busca, mas à medida que esse intervalo vai se aproximando de 1, as direções do anti- $q$ -gradiente vão se tornando predominantemente de descida.

Métodos baseados em  $q$ -gradiente, com essa estratégia de sorteio do parâmetro  $\mathbf{q}$ ,



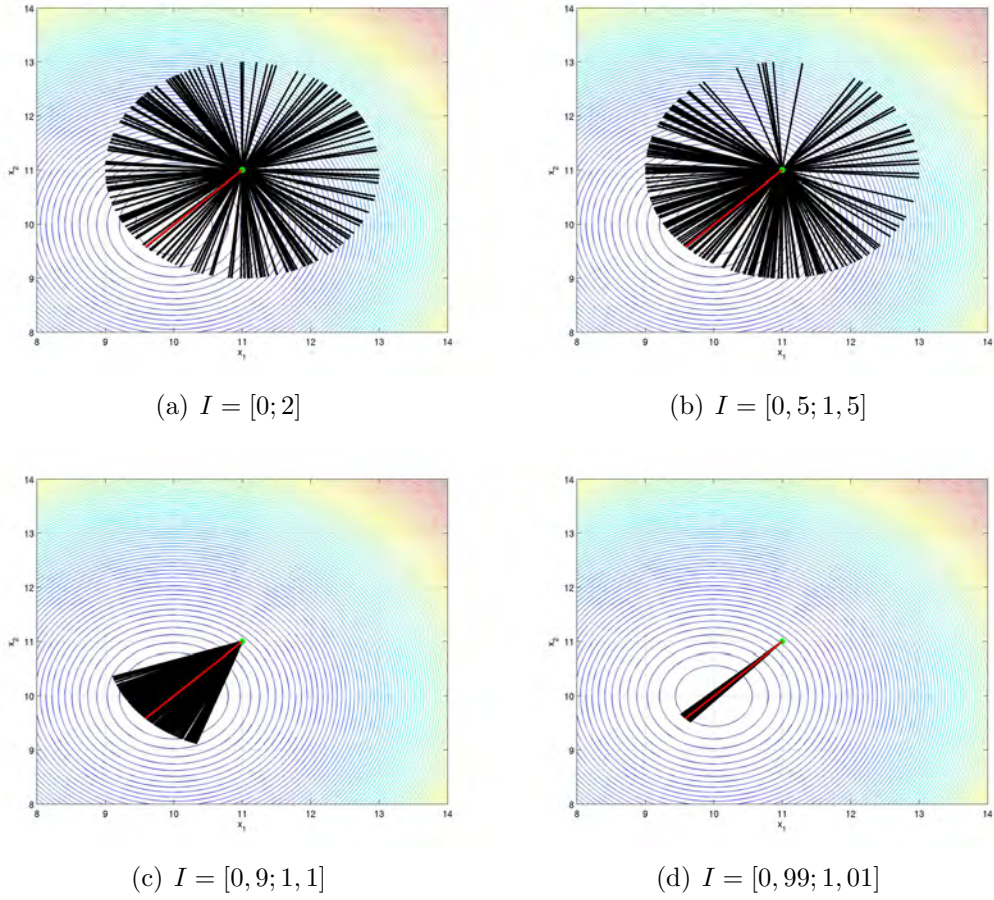


Figura 3.4 - Direções do anti- $q$ -gradiente para diferentes valores do parâmetro  $\mathbf{q}$ , gerados segundo uma distribuição uniforme no intervalo  $I$ .

podem desempenhar busca global (*exploration*) ou busca local (*exploitation*), dependendo do intervalo considerado. No entanto, essa estratégia não realiza uma transição entre exploração e intensificação. Para que o algoritmo apresente essa característica (busca global no início, com  $\mathbf{q}$ 's diferentes de  $\mathbf{1}$ , e busca local no final, com  $\mathbf{q}$ 's próximos ou numericamente iguais a  $\mathbf{1}$ ), o intervalo de geração dos  $\mathbf{q}$ 's deve ser modificado automaticamente ao longo das iterações.

### 3.4.2 Estratégia 2: distribuição log-normal

Com o objetivo de realizar uma transição entre busca global e busca local, propõe-se uma estratégia para a geração dos parâmetros  $\mathbf{q}$ 's em um intervalo de amostragem inicialmente grande e que, ao longo do procedimento iterativo, é reduzido para um intervalo pequeno em torno de  $\mathbf{1}$ . Uma distribuição de probabilidade que atende essa expectativa é a distribuição log-normal, com média  $\mu = 1$  e desvio-padrão  $\sigma^k$

reduzindo, ou decrescendo, ao longo das iterações pela expressão

$$\sigma^{k+1} = \beta \cdot \sigma^k, \quad (3.16)$$

em que  $0 < \beta < 1$  é o fator de redução. Em muitas aplicações, quando a operação relevante é a de multiplicação, a distribuição de probabilidade mais adequada é a distribuição log-normal (LIMPERT et al., 2001). Essa característica influenciou a escolha da distribuição log-normal, pois o  $q$ -gradiente é baseado em multiplicações das variáveis de decisão por um parâmetro  $\mathbf{q}$ . Logo, cada componente  $q_i$  do parâmetro  $\mathbf{q}$  é gerada segundo uma distribuição log-normal com função densidade de probabilidade dada por

$$f(q_i; \mu, \sigma) = \frac{1}{q_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln q_i - \mu)^2}{2\sigma^2}\right), \quad (3.17)$$

com  $\sigma = \sigma^k$  dado pela Equação 3.16 e  $\mu = 1$ .

A Figura 3.5 ilustra a função densidade de probabilidade da distribuição log-normal com média  $\mu = 1$  e diferentes valores de desvio-padrão  $\sigma$ . Note que, quanto menor é o desvio-padrão  $\sigma$ , menor é o intervalo de ocorrência dos números gerados em torno da média.

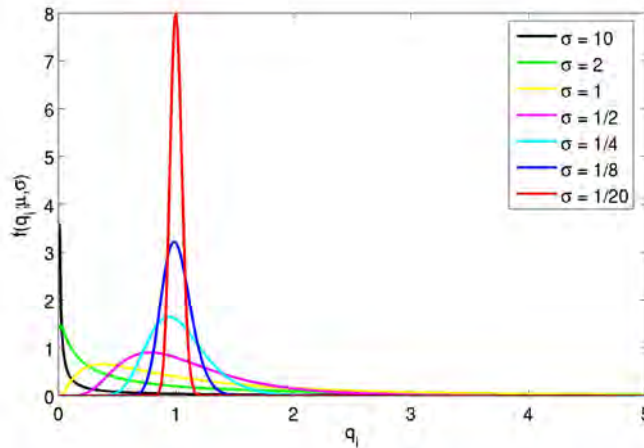


Figura 3.5 - Função densidade de probabilidade log-normal para  $\mu = 1$  e diferentes valores para o desvio-padrão  $\sigma$ .

Nesta estratégia, o desvio-padrão inicial  $\sigma^0 \neq 0$  é responsável pela geração de diferentes valores de  $q_i$  com igual probabilidade de ocorrência nos intervalos  $(0; 1]$  e

$[1; \infty)$ . À medida que  $\sigma^k$  é reduzido ao longo das iterações, a distribuição de probabilidade log-normal tende a uma distribuição Delta de Dirac centrada na média  $\mu = 1$  e os valores de  $q_i$  serão muito próximos ou numericamente iguais a 1. É a redução de  $\sigma$  que garante a transição entre busca de caráter global e busca de caráter local ao longo do procedimento iterativo.

As Figuras 3.6, 3.8 e 3.9, a seguir, ilustram as 500 possíveis direções de busca da função 3.14, com  $\mathbf{o} = (10; 10)$  e no ponto  $\mathbf{x} = (11; 11)$ , para a estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição log-normal. Na Figura 3.6, são considerados diferentes valores do desvio-padrão inicial  $\sigma^0$  sem decrescimento. A semi-reta em vermelho representa a direção de anti-gradiente.

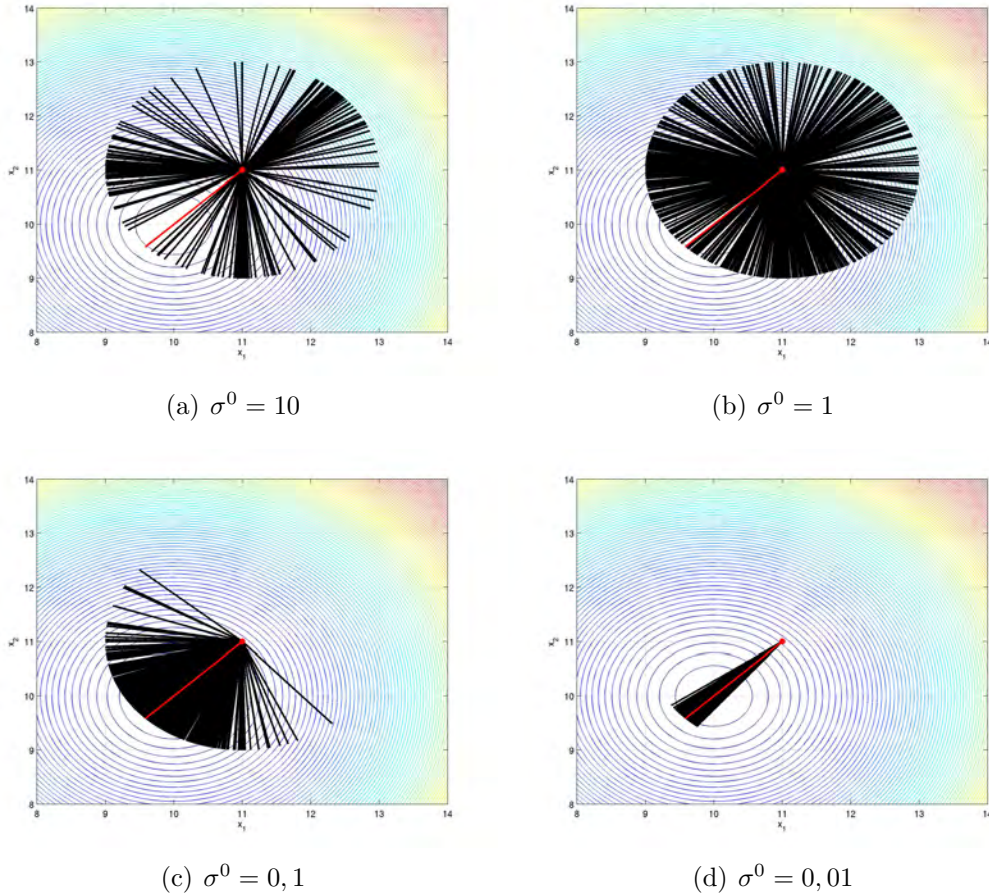


Figura 3.6 - Direções do anti- $q$ -gradiente para  $\mathbf{q}$ 's obtidos segundo uma distribuição log-normal, com diferentes valores de  $\sigma^0$  e sem decrescimento.

O valor do desvio-padrão inicial tem um impacto direto nas direções do anti- $q$ -gradiente. Quanto maior é esse valor, maior a probabilidade da direção do anti- $q$ -

gradiente apontar para qualquer direção do espaço de busca. Note que, para  $\sigma^0 = 0,01$ , as direções são predominantemente de descida e muito próximas da direção de máxima descida.

Para as Figuras 3.8 e 3.9, o desvio-padrão inicial é igual a 1 ( $\sigma^0 = 1$ ) e decresce de acordo com a Equação 3.16 ( $k = 0, \dots, 499$ ) para um determinado fator de redução  $\beta$ , como ilustra a Figura 3.7. Note que, quanto menor é o valor de  $\beta$ , mais rápido é o decaimento do desvio-padrão.

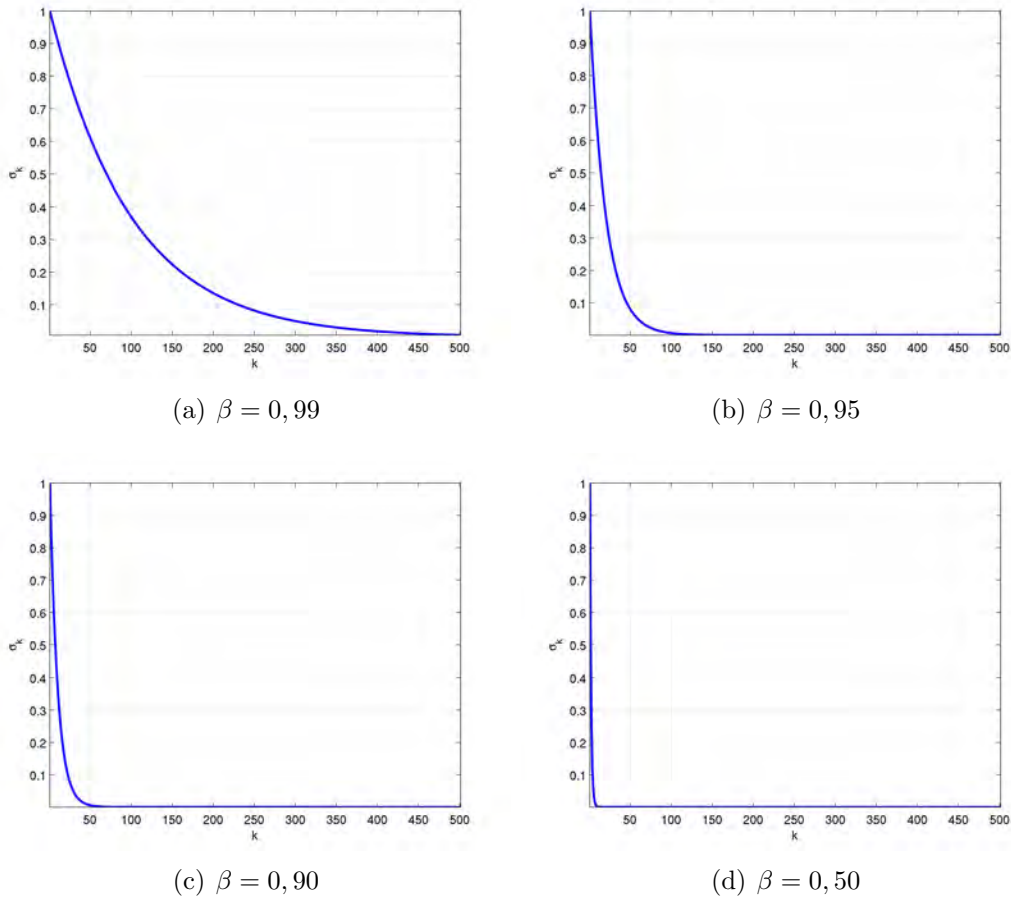


Figura 3.7 - Evolução de  $\sigma^k$  ao longo de 500 iterações, com  $\sigma^0 = 1$  e diferentes fatores de redução  $\beta$ .

Logo, as Figuras 3.8 e 3.9 ilustram, para o mesmo ponto  $\mathbf{x} = (11; 11)$ , as 500 possíveis direções de anti- $q$ -gradiente obtidas de acordo com a evolução de  $\sigma^k$  ( $k = 0, \dots, 499$ ) para um determinado  $\beta$ . Note que, essas figuras não ilustram a sequência de pontos  $\{\mathbf{x}^k\}$  obtida do procedimento iterativo 3.12 ( $k = 0, \dots, 499$ ), mas elas ilustram as possíveis direções de busca obtidas da estratégia de geração dos parâmetros  $\mathbf{q}$ 's



segundo uma distribuição log-normal para um mesmo ponto. A Figura 3.8 considera, para  $\sigma^0 = 1$ , os fatores de redução  $\beta = 0,99$ ,  $\beta = 0,95$ ,  $\beta = 0,90$  e  $\beta = 0,50$ . E a Figura 3.9 ilustra as possíveis direções do anti- $q$ -gradiente em diferentes etapas das 500 iterações, para  $\sigma^0 = 1$  e fator de redução  $\beta = 0,99$ .

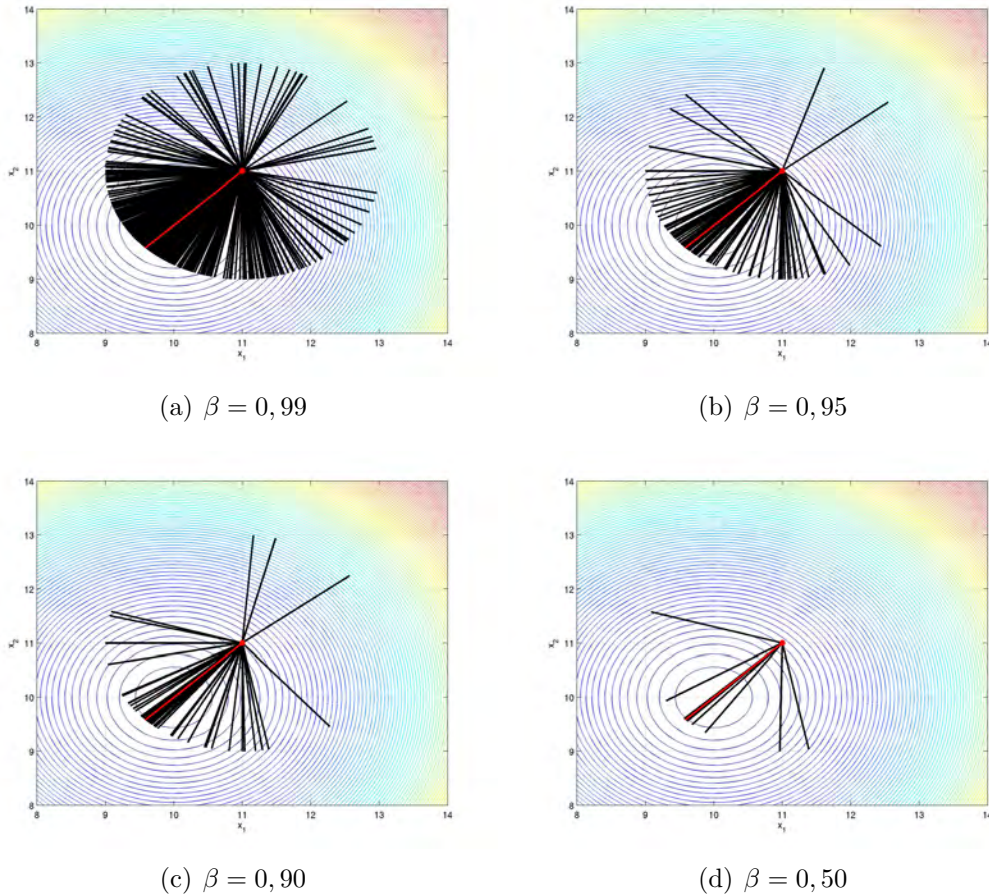


Figura 3.8 - Direções do anti- $q$ -gradiente para  $\mathbf{q}$ 's obtidos segundo uma distribuição log-normal, com  $\sigma^0 = 1$  e diferentes fatores de redução  $\beta$ .

Na Figura 3.8, observa-se a influência do fator de decrescimento  $\beta$ . Para  $\beta$ 's maiores, os  $\mathbf{q}$ 's sorteados tenderão a  $\mathbf{1}$  mais devagar e mais direções do anti- $q$ -gradiente apontarão para direções que não são de descida ao longo do procedimento iterativo. Mas se o fator de redução  $\beta$  for muito pequeno, o desvio-padrão decrescerá rapidamente para zero e os  $\mathbf{q}$ 's sorteados serão predominantemente próximos de  $\mathbf{1}$ .

Na Figura 3.9, as 250 primeiras possíveis direções do anti- $q$ -gradiente apontam para diferentes regiões do espaço de busca. As possíveis direções geradas nas iterações 250 a 500 são predominantemente de descida e o ângulo formado com o anti-gradiente

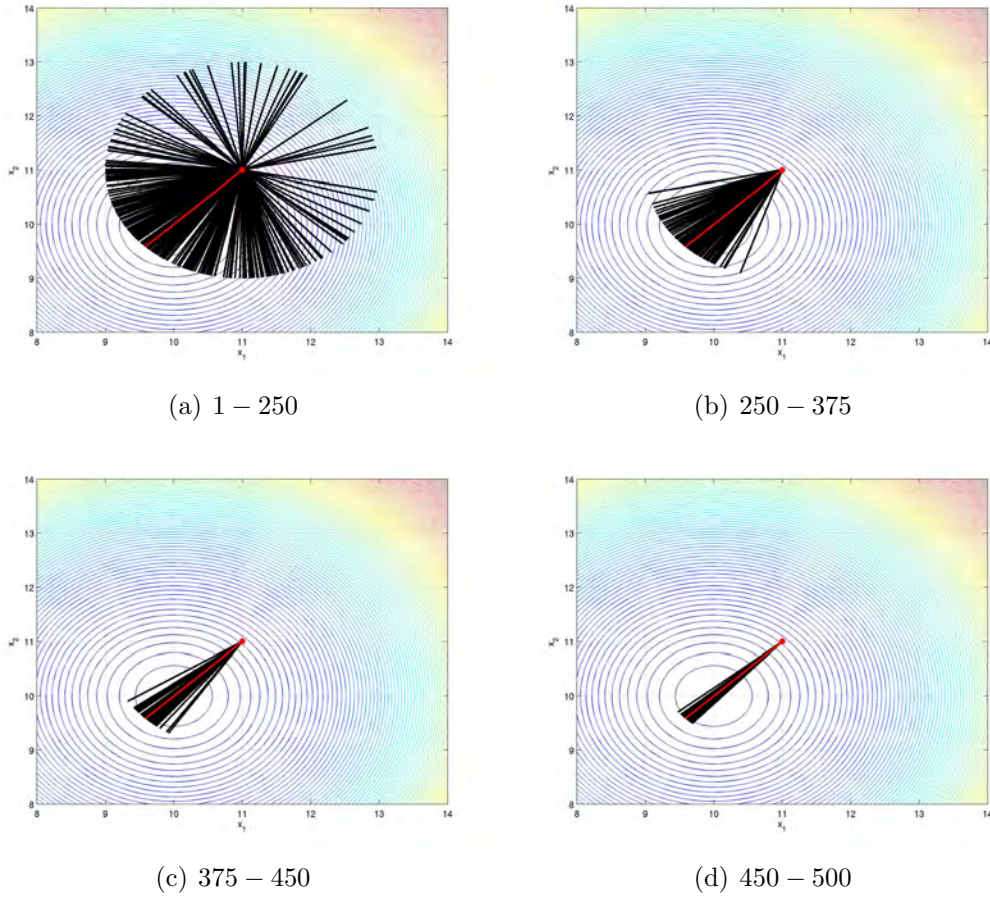


Figura 3.9 - Direções do anti- $q$ -gradiente para  $\mathbf{q}$ 's obtidos segundo uma distribuição log-normal, com  $\sigma^0 = 1$  e fator de redução  $\beta = 0,99$ .

$-\nabla f$  diminui ao longo das iterações, tendendo para a direção de máxima descida  $-\nabla f$  (semi-reta em vermelho).

A estratégia de geração do parâmetro  $\mathbf{q}$  segundo uma distribuição log-normal, com desvio-padrão reduzindo para zero, mostra que, para valores adequados de  $\sigma^0$  e  $\beta$ , as direções do anti- $q$ -gradiente podem apontar para qualquer direção no início do procedimento iterativo e se aproximar da direção do anti-gradiente clássico, ou direção de máxima descida, ao longo das iterações. Dessa forma, métodos de otimização baseados em  $q$ -gradiente com essa estratégia para obtenção dos parâmetros  $\mathbf{q}$ 's podem migrar suavemente de uma busca global para uma busca local.

Por definição, o parâmetro  $q$  necessário para o cálculo da  $q$ -derivada de uma função de uma única variável deve ser um número diferente de 1 (veja a Equação 3.2). A estratégia de obtenção dos parâmetros  $\mathbf{q}$ 's para funções de  $n$  variáveis por meio

de sorteio aleatório, segundo uma distribuição log-normal, exclui a faixa de valores  $(-\infty; 0]$ . Como esta distribuição é assimétrica em torno da média (veja a Figura 3.5), ela fornece a mesma equiprobabilidade de ocorrência de valores nos intervalos  $(0; 1]$  e  $[1; \infty)$ . Para valores altos de  $\sigma^0$ , verificou-se uma alta probabilidade de geração de valores muito distintos entre as coordenadas  $q_i$ 's e, conseqüentemente, entre os fatores  $q_i x_i$ 's que são utilizados no cálculo do vetor  $q$ -gradiente.

### 3.4.3 Estratégia 3: distribuição gaussiana

Para fornecer parâmetros  $\mathbf{q}$ 's com coordenadas  $q_i$ 's não tão distintas entre si, além da possibilidade de gerar  $q_i \in \mathbb{R}$  (e não apenas  $q_i$ 's positivos), foi elaborada a estratégia de geração do fator  $\mathbf{q}^k \mathbf{x}^k$  segundo uma distribuição gaussiana centrada no ponto atual da busca  $\mathbf{x}^k$  e com desvio-padrão  $\sigma^k$ . Nesta estratégia, são gerados os fatores  $q_i x_i$  necessários para calcular as  $q$ -derivadas parciais que compõem o vetor  $q$ -gradiente. Dessa forma, cada coordenada  $q_i$  é obtida a partir do fator  $q_i x_i$  gerado segundo uma distribuição gaussiana com função densidade de probabilidade dada por

$$f(q_i x_i; \sigma, \mu) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{(q_i x_i - \mu)^2}{2\sigma^2} \right], \quad (3.18)$$

com  $\sigma = \sigma^k$  obtido de acordo com a Equação 3.16 na iteração  $k$  e média  $\mu = x_i^k$ , em que  $x_i^k$  é a  $i$ -ésima coordenada do ponto atual da busca  $\mathbf{x}^k$  na iteração  $k$ .

A Figura 3.10 ilustra a função densidade de probabilidade gaussiana com média  $\mu = x_i = 2$  e diferentes valores de desvio-padrão  $\sigma$ . Note que a função é simétrica em relação à média, e quanto menor é o desvio-padrão  $\sigma$ , mais próximos de  $x_i$  o fator  $q_i x_i$  será gerado e, conseqüentemente, mais próximo de 1 a coordenada  $q_i$  será obtida.

As Figuras 3.11, 3.13 e 3.14, a seguir, ilustram as 500 possíveis direções de busca da função 3.14 com  $\mathbf{o} = (10; 10)$  e no ponto  $\mathbf{x} = (11; 11)$  para a estratégia de geração dos fatores  $\mathbf{q}\mathbf{x}$ 's e, conseqüentemente, dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição gaussiana. Na Figura 3.11, são considerados diferentes valores do desvio-padrão inicial  $\sigma^0$  sem decréscimo. A semi-reta em vermelho representa a direção de anti-gradiente.

O valor do desvio-padrão inicial, nesta estratégia, também influencia as direções do anti- $q$ -gradiente. Novamente, quanto maior é esse valor, maior a probabilidade da direção do anti- $q$ -gradiente apontar para qualquer direção do espaço de busca, e quanto menor, mais próxima a direção do anti- $q$ -gradiente será da direção de máxima



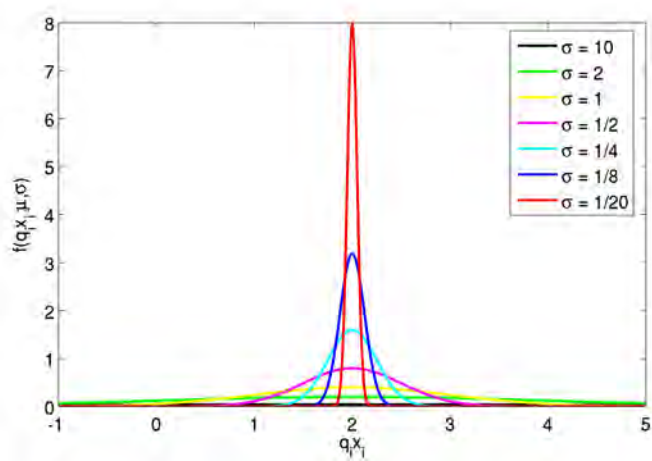
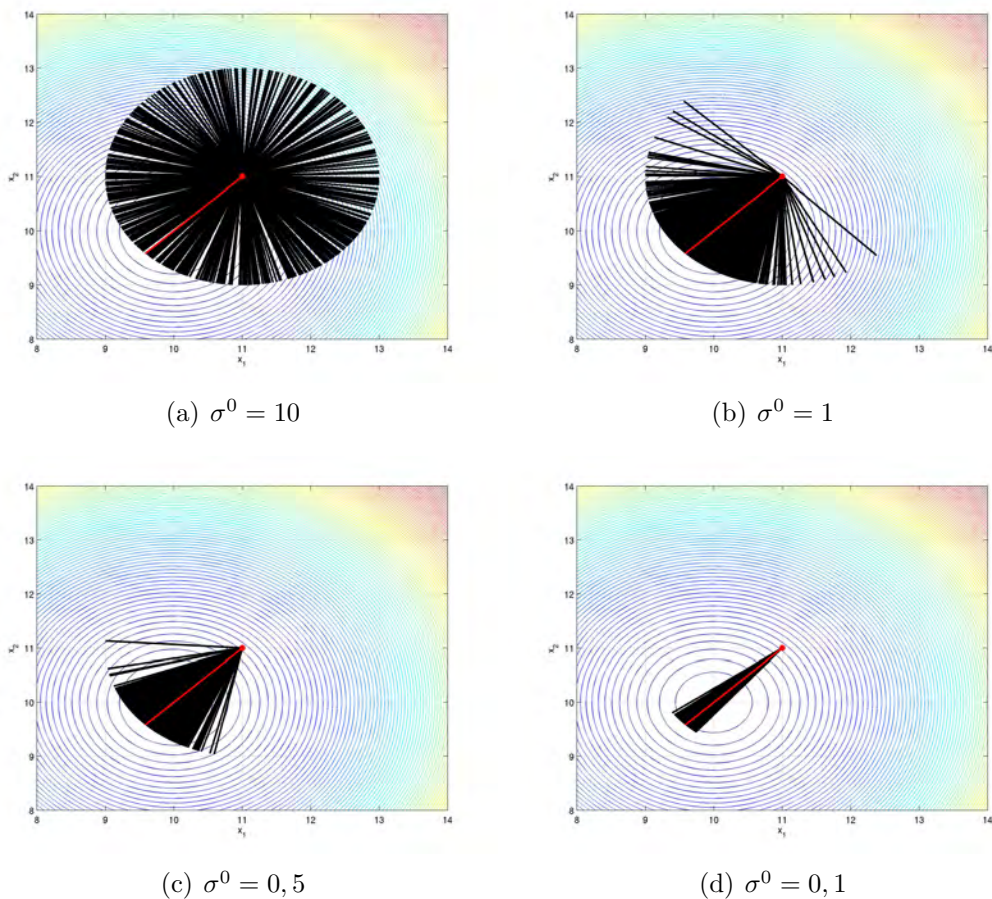


Figura 3.10 - Função densidade de probabilidade gaussiana para  $\mu = x_i = 2$  e diferentes valores para o desvio-padrão  $\sigma$ .



(a)  $\sigma^0 = 10$

(b)  $\sigma^0 = 1$

(c)  $\sigma^0 = 0,5$

(d)  $\sigma^0 = 0,1$

Figura 3.11 - Direções do anti- $q$ -gradiente para  $\mathbf{q}$ 's obtidos segundo uma distribuição gaussiana, com diferentes valores de  $\sigma^0$  e sem decrescimento.



descida.

Para as Figuras 3.13 e 3.14, tem-se  $\sigma^0 = 10$  que decresce conforme a Equação 3.16 ( $k = 0, \dots, 499$ ) para um determinado fator de redução  $\beta$ , como ilustra a Figura 3.12.

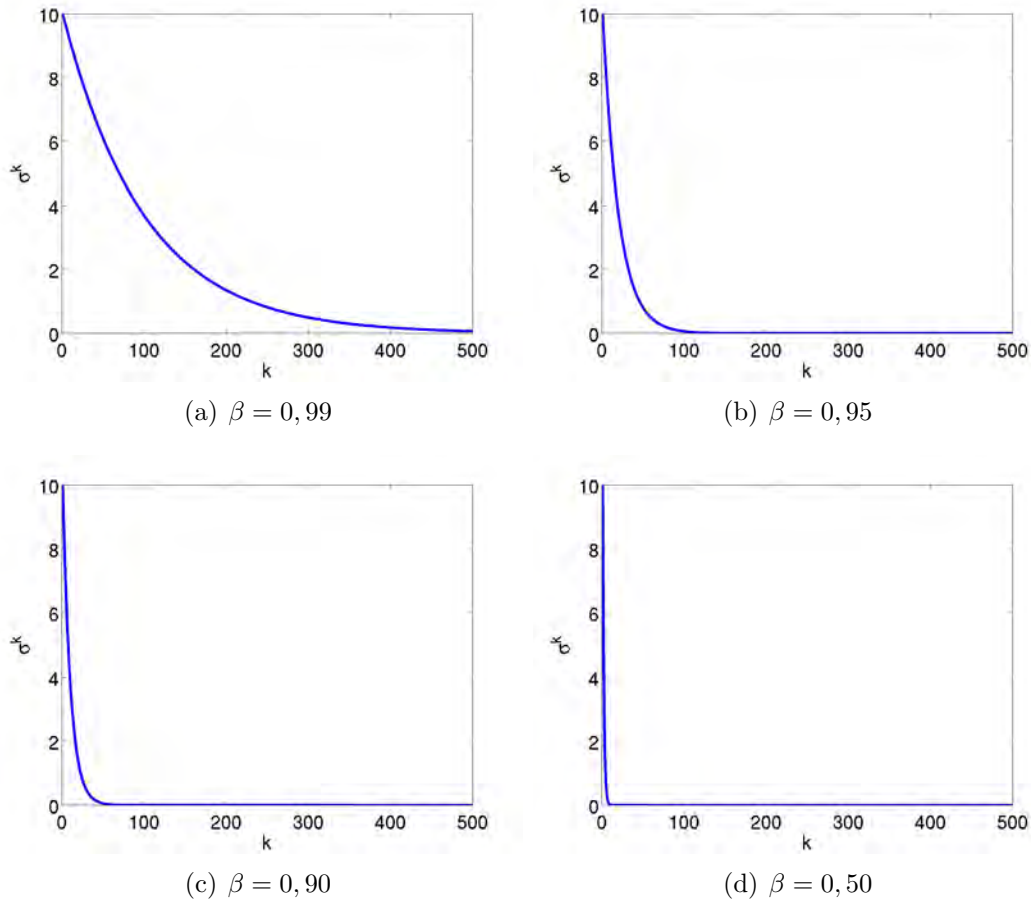


Figura 3.12 - Evolução de  $\sigma^k$  ao longo de 500 iterações, com  $\sigma^0 = 10$  e diferentes fatores de redução  $\beta$ .

Na Figura 3.13, são considerados diferentes fatores de redução  $\beta = 0,99$ ,  $\beta = 0,95$ ,  $\beta = 0,90$  e  $\beta = 0,50$ . E a Figura 3.14 ilustra as direções do anti- $q$ -gradiente em diferentes etapas do procedimento iterativo, para  $\sigma^0 = 10$  e fator de redução  $\beta = 0,99$ .

A Figura 3.13 ilustra a influência do fator de decrescimento  $\beta$  na geração das direções do anti- $q$ -gradiente e, conseqüentemente, nas direções de busca. Quanto menor é o valor de  $\beta$ , mais direções do anti- $q$ -gradiente serão de descida e/ou máxima descida

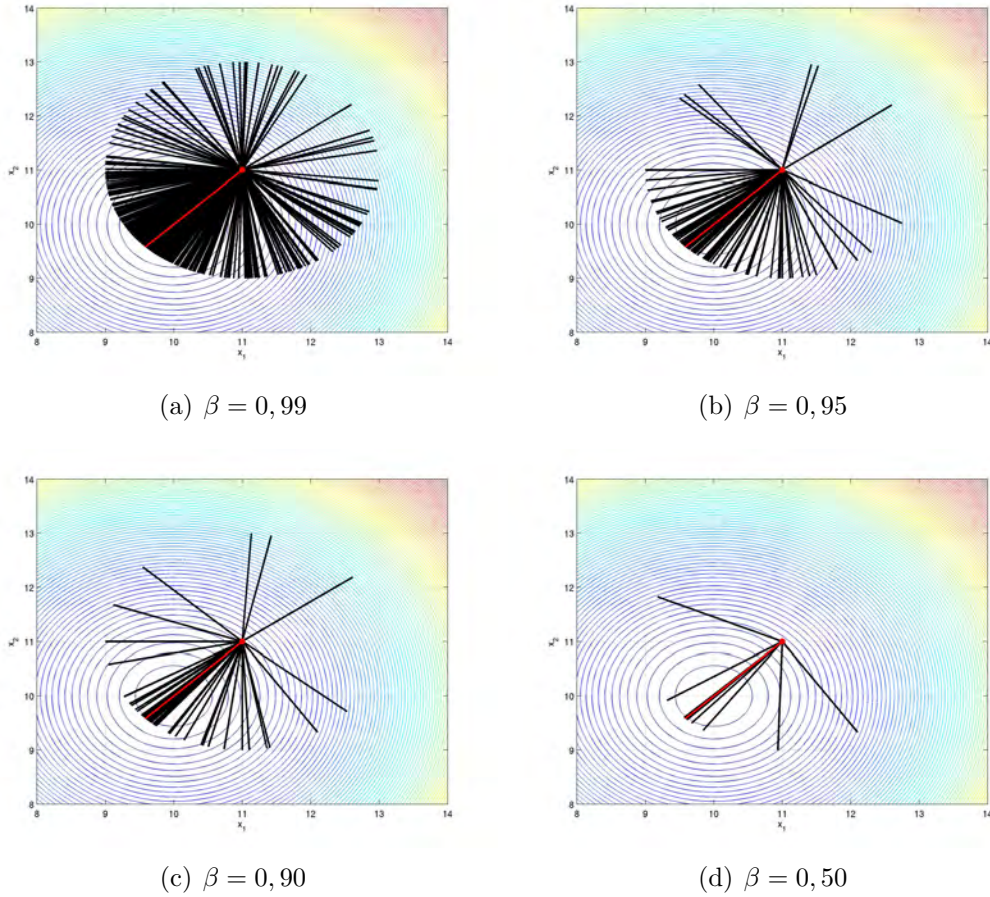


Figura 3.13 - Direções do anti- $q$ -gradiente para  $\mathbf{q}$ 's obtidos segundo uma distribuição gaussiana, com  $\sigma^0 = 10$  e diferentes fatores de redução  $\beta$ .

e a busca realizada por métodos baseados em  $q$ -gradiente será predominantemente local. Se a função objetivo for multimodal, então a busca ficará presa em um mínimo local em poucas iterações. Para  $\beta$ 's maiores, os  $\mathbf{q}$ 's sorteados tenderão a  $\mathbf{1}$  mais devagar e mais direções do anti- $q$ -gradiente apontarão para direções que não são de descida, ou seja, a busca realizada por métodos baseados em  $q$ -gradiente exibirá características de busca global por mais tempo. Na Figura 3.14, as 250 primeiras direções do anti- $q$ -gradiente apontam para diferentes regiões do espaço de busca. As direções geradas nas demais iterações são predominantemente de descida e tendem para a direção de máxima descida  $-\nabla f$  (semi-reta em vermelho).

As conclusões para as estratégias de gerações dos parâmetros  $\mathbf{q}$ 's segundo as distribuições log-normal e gaussiana são semelhantes no que tange a transição entre busca global e busca local ao longo do procedimento iterativo. Essa transição se deve à redução do desvio-padrão  $\sigma$  ao longo das iterações. O  $\sigma$  está relacionado à geração

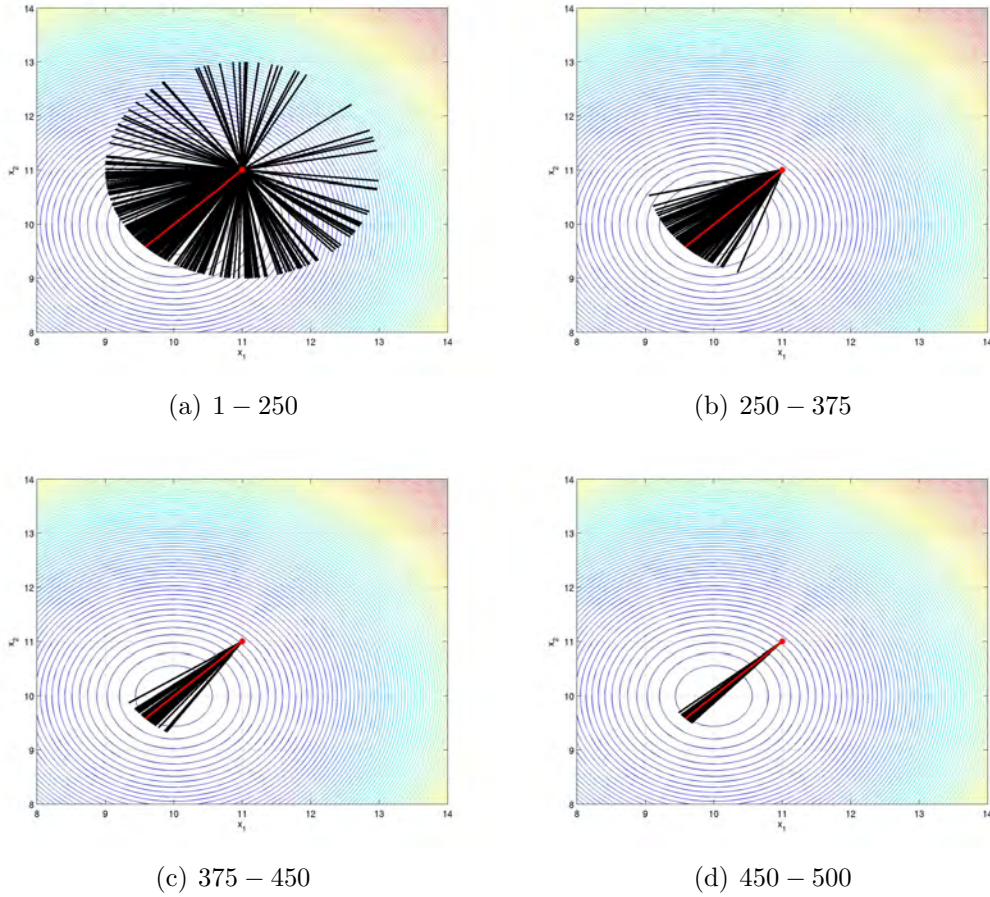


Figura 3.14 - Direções do anti- $q$ -gradiente para  $\mathbf{q}$ 's obtidos segundo uma distribuição gaussiana, com  $\sigma^0 = 10$  e  $\beta = 0,99$ .

dos parâmetros  $\mathbf{q}$ 's e, conseqüentemente, às direções de busca de métodos baseados em  $q$ -gradiente.

Dessa forma, a influência do desvio-padrão  $\sigma$  sobre o comportamento de métodos baseados em  $q$ -gradiente é similar à da temperatura no Recozimento Simulado, ou seja, valores altos de  $\sigma^k$  implicam em uma maior probabilidade de geração de direções de busca de subida, e valores baixos de  $\sigma^k$  implicam em maior probabilidade de geração de direções de busca de descida e, no caso dos métodos baseados em  $q$ -gradiente, tendendo à direção de máxima descida.

Os exemplos desta seção ilustram a relação entre o parâmetro  $\mathbf{q}$  e a direção de busca em métodos de otimização baseados em  $q$ -gradiente. Para que esses métodos sejam eficientes, é preciso elaborar uma estratégia adequada para o tamanho do passo dado na direção do anti- $q$ -gradiente.

### 3.5 Estratégias para escolha do tamanho do passo

Para métodos de otimização baseados em  $q$ -gradiente, as estratégias para a obtenção do tamanho do passo utilizadas foram:

- i) Busca linear via seção áurea,
- ii) Passo adaptativo via diferença  $\| \mathbf{q}\mathbf{x} - \mathbf{x} \|$ ,
- iii) Controle do passo por redução.

#### 3.5.1 Estratégia 1: busca linear

Os métodos baseados em  $q$ -gradiente são generalizações que retomam suas respectivas versões clássicas sempre que o parâmetro  $\mathbf{q}$  tende a  $\mathbf{1}$ . Como os métodos clássicos baseados em gradiente utilizam frequentemente técnicas de busca linear no cálculo do tamanho do passo, pensou-se na utilização de uma técnica de busca linear intervalar, como o método da seção áurea, para a obtenção do tamanho do passo na direção do anti- $q$ -gradiente.

O método da seção áurea não requer o cálculo de derivadas da função objetivo e pode ser aplicado em funções lineares ou não-lineares, com convergência garantida. Em contrapartida, o método requer um número relativamente alto de avaliações da função objetivo (VANDERPLAATS, 1984). Além disso, o método precisará encontrar três pontos que cerquem um mínimo antes de iniciar seu procedimento iterativo, tornando-o ainda mais custoso.

Como o principal objetivo dos métodos baseados em  $q$ -gradiente é introduzir características de busca global em métodos originalmente locais, é intuitivo observar que essa estratégia de obtenção do tamanho do passo pelo método da seção áurea é possível, mas não é a mais adequada, pois a direção do anti- $q$ -gradiente nem sempre será uma direção de descida e as funções objetivo nem sempre serão unimodais. Dessa forma, não há garantias de que o mínimo encontrado pelo método da seção áurea, quando a direção de busca não é de descida, seja realmente um ponto de mínimo.

#### 3.5.2 Estratégia 2: passo adaptativo

Outra estratégia utilizada foi a de um passo adaptativo que deve ser grande na fase de exploração e pequeno na fase de intensificação do algoritmo. Note que as estratégias de geração do parâmetro  $\mathbf{q}^k$ , segundo uma distribuição log-normal ou

gaussiana e descritas na Seção 3.4, geram valores de  $\mathbf{q}^k$  grandes (com maior probabilidade) no início do procedimento iterativo, enquanto o desvio-padrão  $\sigma^k$  também é grande, e geram valores de  $\mathbf{q}^k$  próximos de  $\mathbf{1}$  (com maior probabilidade) no final do procedimento iterativo, quando  $\sigma^k$  é próximo de zero.

Logo, a estratégia de obtenção do tamanho do passo adaptativo pode ser dada pela multiplicação do parâmetro  $\mathbf{q}^k$  sobre o ponto  $\mathbf{x}^k$  da busca, ou seja,

$$\alpha^k = \|\mathbf{q}^k \mathbf{x}^k - \mathbf{x}^k\|, \quad (3.19)$$

com  $\mathbf{q}^k$  gerado segundo uma distribuição de probabilidades log-normal ou gaussiana e desvio-padrão  $\sigma^k$  que vai sendo reduzido ao longo das iterações. É a redução de  $\sigma^k$  que garante a modificação do tamanho do passo  $\alpha^k$ .

Na fase de exploração do método, em que as direções de busca podem ser quaisquer, pois  $\sigma^k$  é grande e  $\mathbf{q}^k$  tem maior probabilidade de assumir valores grandes, o tamanho do passo  $\alpha^k$ , de acordo com a Equação 3.19, também pode assumir valores grandes, desde que o ponto  $\mathbf{x}^k$  seja diferente de zero. Na fase de intensificação, em que as direções de busca são de descida, pois  $\sigma^k \rightarrow 0$  e  $\mathbf{q}^k \rightarrow \mathbf{1}$ , o tamanho do passo  $\alpha^k$ , de acordo com a Equação 3.19, será pequeno e tendendo a zero.

Note que, para  $\mathbf{x}^k = \mathbf{0}$  na Equação 3.19, tem-se  $\alpha^k = 0$  independentemente do valor do parâmetro  $\mathbf{q}^k$ . Dessa forma, métodos baseados em  $q$ -gradiente com esta estratégia de passo adaptativo devem levar em consideração a influência da proximidade com a origem, que implica uma geração muito maior de novos pontos da busca na região em torno de  $\mathbf{x} = \mathbf{0}$ , quando comparado com as demais regiões do espaço de busca. Esse fenômeno foi denominado densidade não-uniforme de pontos e está ilustrado na Figura 3.15.

Os pontos da Figura 3.15 representam os pontos da busca calculados a partir do procedimento iterativo 3.12 para os eixos coordenados tomados como direções de busca e tamanho de passo obtido pela Equação 3.19. Considerou-se, inicialmente, a direção  $\mathbf{d} = (d_1; d_2) = (1; 0)$  e o ponto inicial  $(x_1^0; x_2^0) = (1; 1)$ . Com isso, tem-se  $x_2^k = 1$  e apenas  $x_1^k$  é atualizado. O cálculo do tamanho do passo e dos demais pontos da busca são realizados da seguinte forma, para  $q = 1,5$  a partir de  $x_1^0 = 1$ :

$$\alpha^0 = |qx_1^0 - x_1^0| = |1,5 \cdot 1 - 1| = 0,5,$$

$$x_1^1 = x_1^0 + \alpha^0 \cdot d_1 = 1 + 0,5 \cdot 1 = 1,5,$$



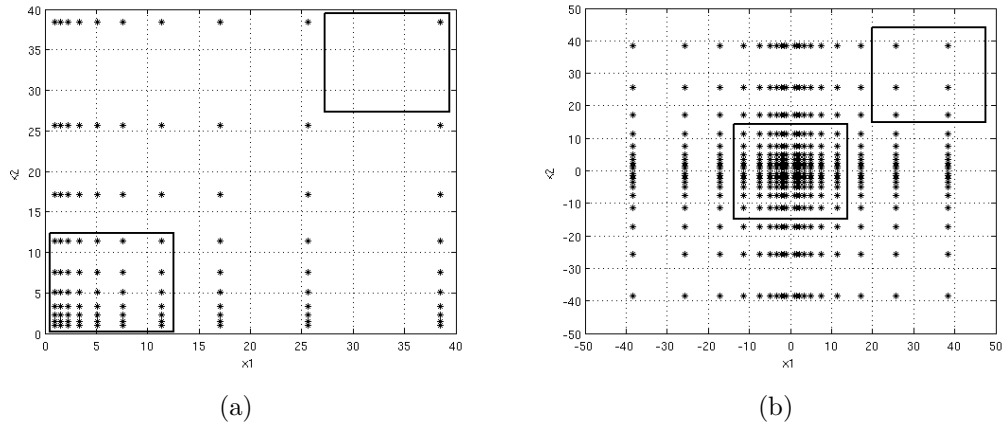


Figura 3.15 - Densidade não-uniforme de pontos para (a) primeiro quadrante e (b) todos os quadrantes. Regiões próximas da origem são mais visitadas que as demais regiões do espaço de busca.

e o ponto  $(x_1^1; x_2^1) = (1, 5; 1)$  é plotado. Em seguida, para  $q = 1,5$  em  $x_1^1 = 1,5$  tem-se:

$$\alpha^1 = |qx_1^1 - x_1^1| = |1,5 \cdot 1,5 - 1,5| = 0,75,$$

$$x_1^2 = x_1^1 + \alpha^1 \cdot d_1 = 1,5 + 0,75 \cdot 1 = 2,25,$$

e o ponto  $(x_1^2; x_2^2) = (2, 25; 1)$  é plotado. Esse procedimento se repete e o resultado é o conjunto de pontos plotados na horizontal para  $x_2 = 1$  e diferentes valores para  $x_1$ . De forma análoga, considera-se a direção  $\mathbf{d} = (d_1; d_2) = (0; 1)$ , o ponto inicial  $(x_1^0; x_2^0) = (1; 1)$  com  $q = 1,5$  e obtém-se o conjunto de pontos plotados na vertical para  $x_1 = 1$  e diferentes valores de  $x_2$ . O mesmo procedimento é realizado para outros pontos  $(x_1^0; x_2^0)$  sempre com  $q = 1,5$  e fixando um dos eixos coordenados como direção de busca.

### 3.5.3 Estratégia 3: controle do passo por redução

A última estratégia utilizada é a de controle do passo por redução, em que um valor inicial para o tamanho do passo ( $\alpha^0$ ) é reduzido a cada iteração por um fator de redução  $0 < \beta < 1$ , ou seja,

$$\alpha^{k+1} = \beta \cdot \alpha^k. \quad (3.20)$$

Por simplicidade, considerou-se o mesmo fator de redução  $\beta$  utilizado na redução do desvio-padrão  $\sigma$ , responsável pela geração do parâmetro  $\mathbf{q}$ .

Um método que utiliza apenas direções de descida com tamanho de passo  $\alpha$  cons-

tante em todas as iterações, provavelmente, não converge se  $\alpha$  for muito grande e apresenta convergência lenta se  $\alpha$  for muito pequeno (IZMAILOV; SOLODOV, 2007). Dessa forma, a redução do tamanho inicial do passo  $\alpha^0$  ao longo das iterações permite passos grandes no início do procedimento iterativo (quando a busca é global, o desvio-padrão inicial  $\sigma^0$  é diferente de zero e as direções de busca são quaisquer) e passos pequenos e tendendo a zero no final (quando a busca é local, com  $\sigma^k \rightarrow 0$  e as direções de busca são de descida ou de máxima descida). Ao contrário da Regra de Armijo (veja a Seção 2.6.1), aqui o tamanho do passo  $\alpha^k$  é reduzido em todas as iterações, e não apenas se um decréscimo suficiente da função objetivo for obtido.

### 3.6 Método do $q$ -gradiente

Nesta seção, a  $q$ -versão do método da máxima descida, denominado método do  $q$ -gradiente, é apresentado com base no Algoritmo 5 e nas estratégias para obtenção do parâmetro  $\mathbf{q}$  (Seção 3.4) e tamanho do passo (Seção 3.5).

Embora o método do  $q$ -gradiente seja uma generalização do método da máxima descida, não é apropriado denominá-lo método da  $q$ -máxima descida, pois dependendo do parâmetro  $\mathbf{q}$  utilizado para o cálculo do  $q$ -gradiente a direção de busca pode não ser uma direção de máxima descida, mas apenas uma direção de descida ou ainda uma direção de subida.

Considere uma função diferenciável  $f$  de  $n$  variáveis. O método do  $q$ -gradiente utiliza o procedimento iterativo 3.12, a partir de um ponto inicial  $\mathbf{x}^0 \in \mathbb{R}^n$ , com direção de busca  $\mathbf{d}^k \in \mathbb{R}^n$  na iteração  $k$  dada pela direção do anti- $q$ -gradiente normalizada, ou seja,

$$\mathbf{d}^k = -\nabla_{\mathbf{q}}f(\mathbf{x}^k) / \|\nabla_{\mathbf{q}}f(\mathbf{x}^k)\|, \quad (3.21)$$

em que  $\nabla_{\mathbf{q}}f(\mathbf{x}^k)$  é o vetor  $q$ -gradiente de  $f$  no ponto  $\mathbf{x}^k$ , dado pela Equação 3.9.

A seguir, o algoritmo do método do  $q$ -gradiente, desenvolvido a partir do algoritmo geral para métodos baseados em  $q$ -gradiente, é apresentado para as diferentes estratégias de obtenção do parâmetro  $\mathbf{q}$  e de tamanho do passo.

Para a estratégia de geração do parâmetro  $\mathbf{q}$  segundo uma distribuição log-normal ou gaussiana, é preciso considerar a instrução  $\sigma^{k+1} = \beta \cdot \sigma^k$  (Passo 9) para garantir a redução de  $\sigma^0$  ao longo do procedimento iterativo. Para a estratégia de controle do passo por redução, é preciso considerar a instrução  $\alpha^{k+1} = \beta \cdot \alpha^k$  (Passo 10) para que o passo seja grande no início e pequeno no final do procedimento iterativo.

---

**Algoritmo 6:** Algoritmo geral para o método do  $q$ -gradiente

---

**Dados**  $f(\mathbf{x})$  contínua e diferenciável com  $\mathbf{x} \in \mathbb{R}^n$  e um ponto inicial  $\mathbf{x}^0$ . Escolher uma das três estratégias para a geração do parâmetro  $\mathbf{q}$  e os parâmetros livres associados:  $[q_{min}; q_{max}]$  para distribuição uniforme;  $\mu = 1$ ,  $\sigma^0 > 0$  e  $0 < \beta < 1$  para a distribuição log-normal; e  $\mu = \mathbf{x}$ ,  $\sigma^0 > 0$  e  $0 < \beta < 1$  para a distribuição gaussiana. Escolher uma entre as três estratégias para o cálculo do tamanho do passo  $\alpha$ : busca linear exata (Seção Áurea), passo adaptativo (Equação 3.19) e, no caso do controle do passo por redução (Equação 3.20), escolher também o parâmetro livre  $\alpha^0 > 0$ . Escolha um dos critérios de parada ilustrados na Seção 2.4.

- 1: **Faça**  $k = 0$
  - 2: **Faça**  $\mathbf{x}_{melhor} = \mathbf{x}^0$
  - 3: **Enquanto** não atingir um critério de parada, **faça**
  - 4:   Obtenha o parâmetro  $\mathbf{q}^k$  por uma distribuição de probabilidade
  - 5:    $\mathbf{d}^k = -\nabla_{\mathbf{q}} f(\mathbf{x}^k) / \|\nabla_{\mathbf{q}} f(\mathbf{x}^k)\|$
  - 6:   Obtenha o tamanho do passo  $\alpha^k$
  - 7:    $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
  - 8:   **Se**  $f(\mathbf{x}^{k+1}) < f(\mathbf{x}_{melhor})$  **então**  $\mathbf{x}_{melhor} = \mathbf{x}^{k+1}$
  - 9:    $\sigma^{k+1} = \beta \cdot \sigma^k$  (se necessário)
  - 10:    $\alpha^{k+1} = \alpha \cdot \alpha^k$  (se necessário)
  - 11:    $k = k + 1$
  - 12: **Retorne**  $\mathbf{x}_{melhor}$
- 

Com o objetivo de ilustrar as estratégias de obtenção do parâmetro  $\mathbf{q}$ , combinadas com algumas estratégias de obtenção do tamanho do passo, o Algoritmo 6 foi aplicado na função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  no caso bidimensional ( $n = 2$ )

$$f(\mathbf{x}) = 2 - \left[ \exp \left( - \sum_{i=1}^n (x_i - o_i)^2 \right) + 2 \exp \left( - \sum_{i=1}^n ((x_i - o_i) - 3)^2 \right) \right] \quad (3.22)$$

em que  $\mathbf{o} \in \mathbb{R}^n$  é o deslocamento no espaço de busca. Quando  $\mathbf{o} = (0; 0)$ , o mínimo local da  $f$  é  $\mathbf{x} = (0; 0)$  e o mínimo global é  $\mathbf{x}^* = (3; 3)$  (veja a Figura 3.16). Já para  $\mathbf{o} = (10; 10)$  os mínimos são  $\mathbf{x} = (10; 10)$  (local) e  $\mathbf{x}^* = (13; 13)$  (global) como ilustra a Figura 3.17. Tanto para  $\mathbf{o} = (0; 0)$  quanto para  $\mathbf{o} = (10; 10)$ , as características da função  $f$  são as mesmas.

O comportamento do Algoritmo 6, para algumas combinações de estratégias de obtenção do parâmetro  $\mathbf{q}$  e tamanho do passo  $\alpha$ , está ilustrado nas Figuras 3.18 a 3.22. Como o objetivo é apenas ilustrar as estratégias, o critério de parada utilizado em todas as execuções foi um número máximo de avaliações da função objetivo igual a  $1000 \cdot n$ , isto é, 2000 avaliações. O ponto inicial é sempre  $\mathbf{x}^0 = (1; 1) + \mathbf{o}$ , ilustrado em vermelho nas figuras, e a sequência de números pseudo-aleatórios utilizada na



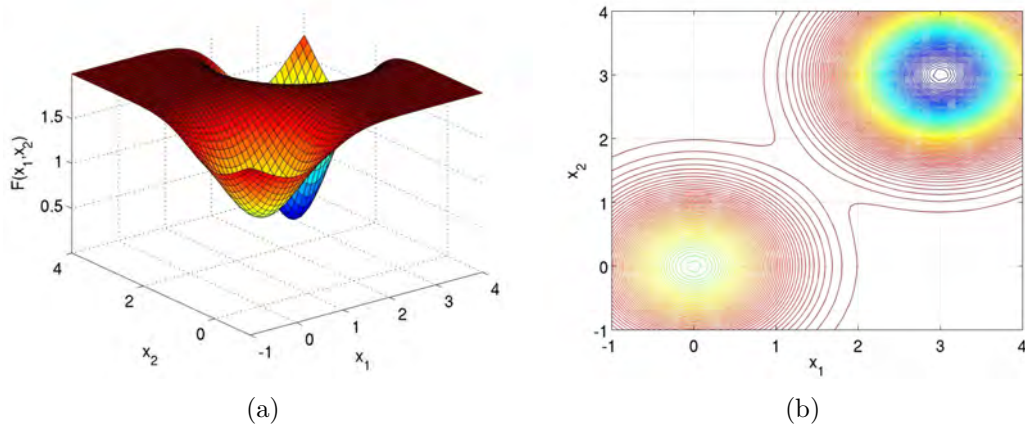


Figura 3.16 - (a) Função bidimensional dada pela Equação 3.22 e (b) suas curvas de contorno para  $\mathbf{o} = (0; 0)$ .

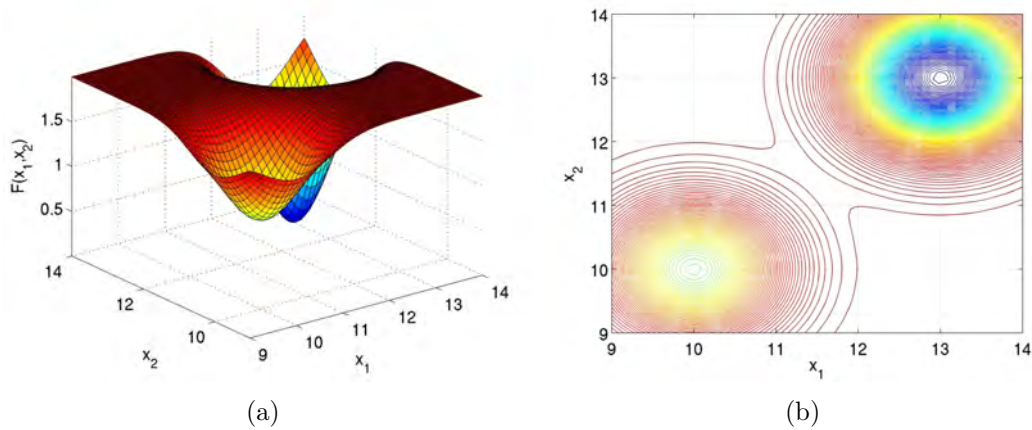


Figura 3.17 - (a) Função bidimensional dada pela Equação 3.22 e (b) suas curvas de contorno para  $\mathbf{o} = (10; 10)$ .

geração dos parâmetros  $\mathbf{q}$ 's é a mesma para todas as execuções. Em alguns casos, são considerados o domínio com e sem deslocamento para ilustrar a influência do fenômeno da densidade não-uniforme de pontos.

A Figura 3.18 ilustra a sequência de pontos  $\{\mathbf{x}^k\}$  gerada pelo Algoritmo 6 para a estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição uniforme no intervalo  $[q_{min}, q_{max}]$  e estratégia de obtenção do tamanho do passo adaptativo. Os únicos parâmetros livres são  $q_{min}$  e  $q_{max}$ . Como esta estratégia de geração dos parâmetros  $\mathbf{q}$ 's não fornece uma transição entre busca global e busca local, considerou-se um intervalo mais largo e outro mais estreito em torno de 1:  $[q_{min}; q_{max}] = [0, 5; 1, 5]$

e  $[q_{min}; q_{max}] = [0, 99; 1, 01]$ .

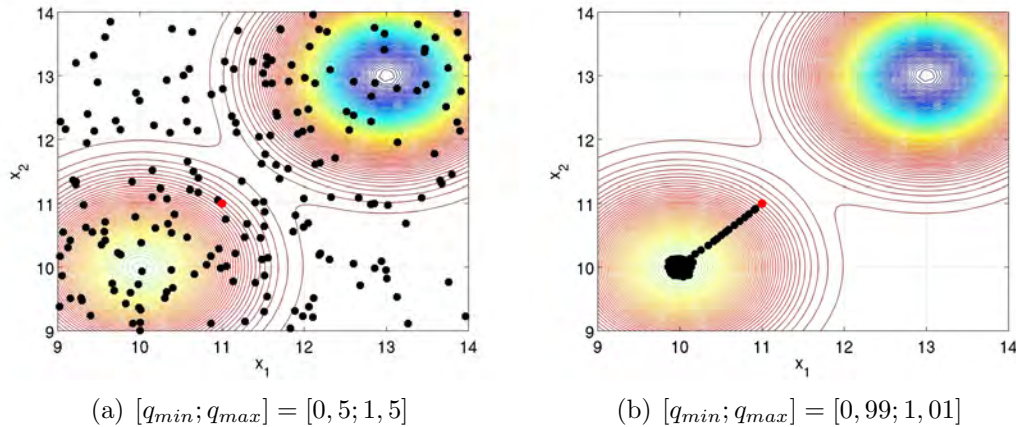


Figura 3.18 - Sequência de pontos  $\{\mathbf{x}^k\}$  obtida para a estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição uniforme no intervalo  $[q_{min}; q_{max}]$  e tamanho do passo adaptativo. Considerou-se o domínio da  $f$  deslocado com  $\mathbf{o} = (10; 10)$ .

Note que, para a mesma estratégia de cálculo do tamanho do passo, o algoritmo apresenta comportamentos distintos de acordo com o intervalo de geração dos parâmetros  $\mathbf{q}$ 's. Para  $\mathbf{q} \in [0, 5; 1, 5]$ , observa-se uma exploração maior do espaço de busca viável, ou seja, a busca tem caráter global. Já para  $\mathbf{q} \in [0, 99; 1, 01]$  a sequência de pontos gerada concentra-se na bacia de atração do mínimo local mais próximo do ponto inicial da busca, ou seja, a busca tem caráter local e de intensificação. Essas características de exploração e intensificação não ocorrem na mesma execução do algoritmo e, para que isso ocorra, conforme discutido na Seção 3.4, o intervalo de geração dos parâmetros  $\mathbf{q}$ 's deve ser modificado automaticamente. Logo, a geração dos  $\mathbf{q}$ 's segundo uma distribuição log-normal ou gaussiana com desvio-padrão decrescente ao longo das iterações é mais adequada.

As Figuras 3.19 e 3.20 ilustram a sequência de pontos  $\{\mathbf{x}^k\}$  gerada pelo Algoritmo 6 para a estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição log-normal com  $\mu = 1$ ,  $\sigma^0 = 3$  e fator de redução  $\beta = 0,99$ . Esta estratégia foi combinada com a estratégia de tamanho do passo adaptativo (ver Figura 3.19) e com a estratégia de controle do passo por redução, para  $\alpha^0 = 3,75$  (ver Figura 3.20). Considerou-se a função 3.22 com e sem deslocamento para ilustrar a influência do fenômeno da densidade não-uniforme de pontos no desempenho do método.

Quando a função está deslocada (Figuras 3.19a e 3.20a) é possível observar uma

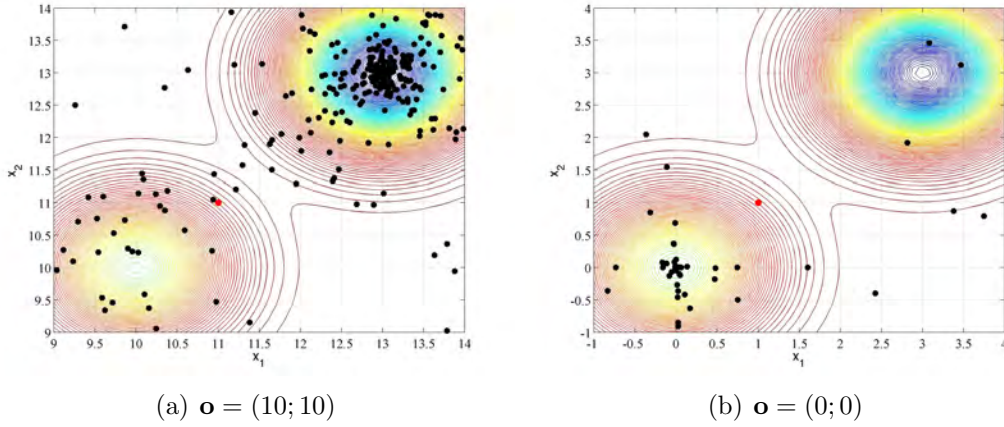


Figura 3.19 - Sequência de pontos  $\{\mathbf{x}^k\}$  obtida para a estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição log-normal com  $\mu = 1$ ,  $\sigma^0 = 3$  e  $\beta = 0,99$ , e tamanho do passo adaptativo para (a)  $\mathbf{o} = (10; 10)$  e (b)  $\mathbf{o} = (0; 0)$ .

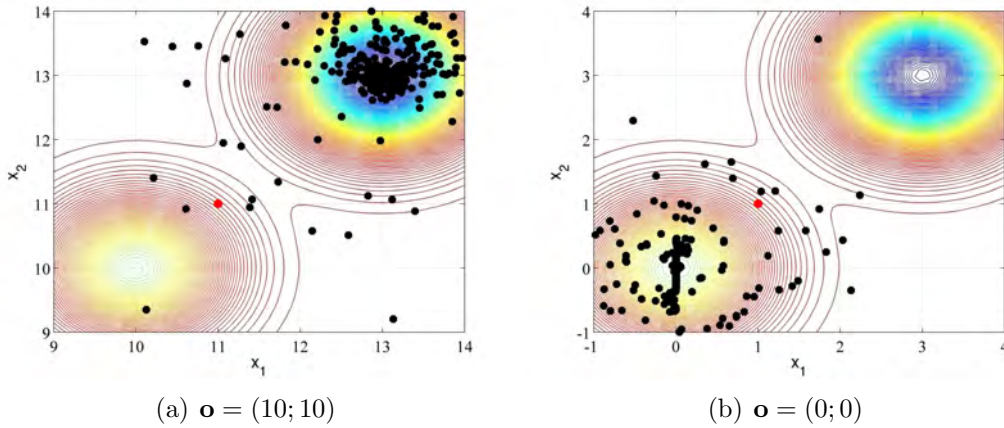


Figura 3.20 - Sequência de pontos  $\{\mathbf{x}^k\}$  obtida para a estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição log-normal com  $\mu = 1$ ,  $\sigma^0 = 3$  e  $\beta = 0,99$ , e controle do passo por redução com  $\alpha^0 = 3,75$  para (a)  $\mathbf{o} = (10; 10)$  e (b)  $\mathbf{o} = (0; 0)$ .

maior exploração do espaço de busca e uma intensificação na bacia de atração do mínimo global  $\mathbf{x}^* = (13; 13)$ , em relação à função sem deslocamento (Figuras 3.19b e 3.20b). Neste último caso, verifica-se a ocorrência do fenômeno da densidade não-uniforme de pontos e o método converge para o mínimo local localizado na origem  $\mathbf{x} = (0; 0)$ , tanto para o passo adaptativo quanto para o controle do passo por redução. No entanto, a ocorrência desse fenômeno parece ser de menor intensidade para a estratégia de controle do passo por redução, pois menos pontos sobrepõem a origem.

As Figuras 3.21 e 3.22 ilustram a sequência de pontos  $\{\mathbf{x}^k\}$  gerada pelo Algoritmo 6 para a estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição gaussiana com  $\mu = \mathbf{x}^k$ ,  $\sigma^0 = 3$  e  $\beta = 0,99$ . A combinação desta estratégia de obtenção dos parâmetros  $\mathbf{q}$ 's com a estratégia de passo adaptativo é ilustrada na Figura 3.21 e a combinação com a estratégia de controle do passo por redução, com  $\alpha^0 = 3,75$ , pode ser observada na Figura 3.22. Novamente, considerou-se a função 3.22 com e sem deslocamento.

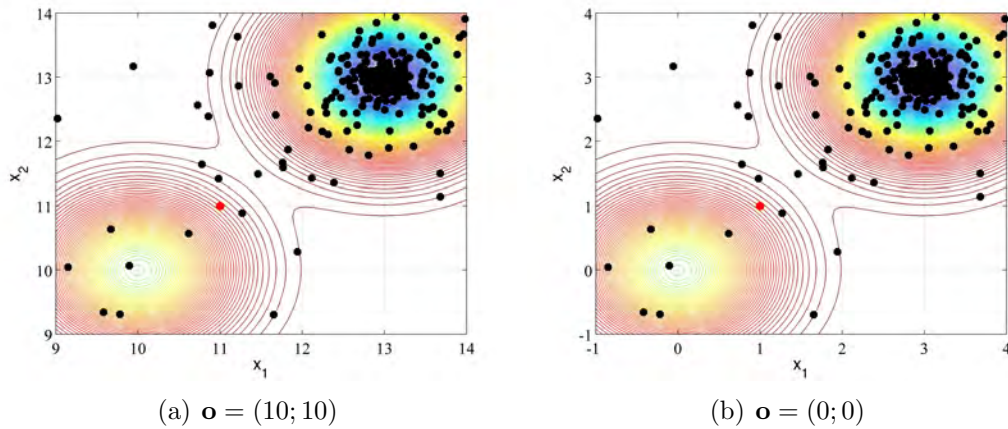


Figura 3.21 - Sequência de pontos  $\{\mathbf{x}^k\}$  obtida para a estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição gaussiana com  $\mu = \mathbf{x}^k$ ,  $\sigma^0 = 3$  e  $\beta = 0,99$ , e tamanho do passo adaptativo para (a)  $\mathbf{o} = (10; 10)$  e (b)  $\mathbf{o} = (0; 0)$ .

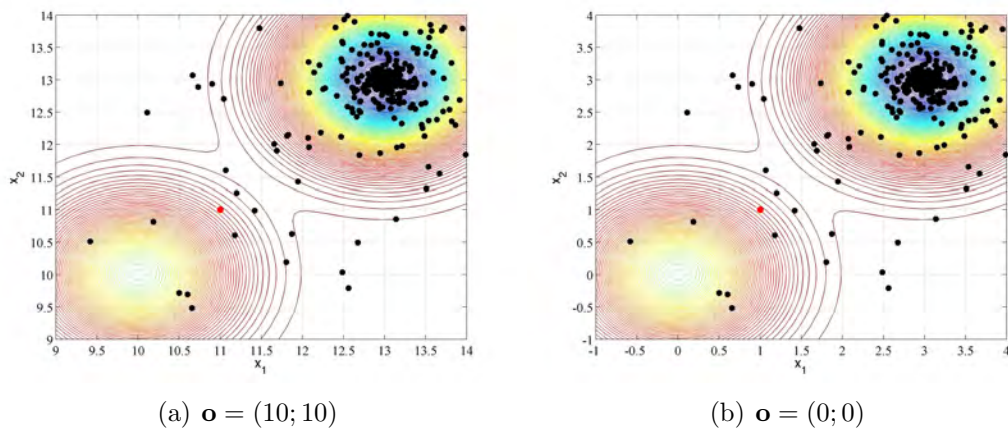


Figura 3.22 - Sequência de pontos  $\{\mathbf{x}^k\}$  obtida para a estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição gaussiana com  $\mu = \mathbf{x}^k$ ,  $\sigma^0 = 3$  e  $\beta = 0,99$ , e controle do passo por redução com  $\alpha^0 = 3,75$  para (a)  $\mathbf{o} = (10; 10)$  e (b)  $\mathbf{o} = (0; 0)$ .



Tanto na Figura 3.21 quanto na Figura 3.22, a estratégia de geração dos parâmetros  $\mathbf{q}$ 's é a da distribuição gaussiana e o que muda é apenas a estratégia de cálculo do tamanho do passo. Os parâmetros  $\mathbf{q}$ 's foram gerados para a mesma sequência de números pseudo-aleatórios. Note que a sequência de pontos  $\{\mathbf{x}^k\}$  ilustrada na Figura 3.21a é a mesma ilustrada na Figura 3.21b. O mesmo comportamento é verificado na Figura 3.22a em relação à Figura 3.22b. Isso mostra que, o uso da distribuição gaussiana na geração dos parâmetros  $\mathbf{q}$ 's não foi influenciado pelo fenômeno da densidade não-uniforme de pontos, independente das estratégias para a obtenção do tamanho do passo.

Diante de todas essas ilustrações, as estratégias que apresentaram os melhores resultados foram a da geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição gaussiana, que não foi influenciada pelo fenômeno da densidade não-uniforme, e a do controle do tamanho do passo por redução. Dessa forma, o algoritmo de referência para o método do  $q$ -gradiente e que foi utilizado nas validações sobre funções teste da área de otimização é o Algoritmo 7 apresentado a seguir.

---

**Algoritmo 7:** Algoritmo para o método do  $q$ -gradiente

---

**Dados**  $f(\mathbf{x})$  contínua e diferenciável com  $\mathbf{x} \in \mathbb{R}^n$ , um ponto inicial  $\mathbf{x}^0$  e os parâmetros livres  $\sigma^0 > 0$ ,  $\alpha^0 > 0$  e  $0 < \beta < 1$

- 1: **Faça**  $k = 0$
  - 2: **Faça**  $\mathbf{x}_{melhor} = \mathbf{x}^0$
  - 3: **Enquanto** não atingir um critério de parada, **faça**
  - 4:   Obtenha  $\mathbf{q}^k$  segundo uma distribuição gaussiana com  $\mu = \mathbf{x}^k$  e  $\sigma^k$
  - 5:    $\mathbf{d}^k = -\nabla_{\mathbf{q}} f(\mathbf{x}^k) / \|\nabla_{\mathbf{q}} f(\mathbf{x}^k)\|$
  - 6:    $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
  - 7:   **Se**  $f(\mathbf{x}^{k+1}) < f(\mathbf{x}_{melhor})$  **então**  $\mathbf{x}_{melhor} = \mathbf{x}^{k+1}$
  - 8:    $\sigma^{k+1} = \beta \cdot \sigma^k$
  - 9:    $\alpha^{k+1} = \beta \cdot \alpha^k$
  - 10:    $k = k + 1$
  - 11: **Retorne**  $\mathbf{x}_{melhor}$
- 

O critério de parada pode seguir as regras descritas na Seção 2.4. Como a estratégia escolhida para a geração dos parâmetros  $\mathbf{q}$ 's possui uma fase de exploração, é importante salvar o ponto com menor valor da função objetivo visitado durante a busca e retorná-lo como a melhor solução.

Em suma, o Algoritmo 7 do método do  $q$ -gradiente pode ser considerado uma generalização do Algoritmo 1 do método da máxima descida. Basicamente, a direção

de busca dada pelo anti-gradiente é substituída pela direção de anti- $q$ -gradiente. Essa alteração, por sua vez, requer o cálculo do parâmetro  $\mathbf{q}$ . A estratégia utilizada pelo Algoritmo 7 é a de geração de números aleatórios segundo uma distribuição de probabilidades gaussianas com desvio-padrão  $\sigma^0 > 0$ , que decresce ao longo do procedimento iterativo ( $\sigma^k \rightarrow 0$ ).

O primeiro impacto desta estratégia é a presença de aleatoriedade na busca, que transforma o método do  $q$ -gradiente em um método estocástico. O segundo impacto é a presença de uma transição suave entre busca global e busca local, controlada pelo parâmetro  $\sigma$ . E o terceiro impacto desta estratégia de geração dos parâmetros  $\mathbf{q}$ 's, aliada à estratégia de obtenção do tamanho do passo  $\alpha$ , é a aceitação de soluções piores, isto é,  $\mathbf{x}^{k+1}$  é sempre aceito mesmo que  $f(\mathbf{x}^{k+1}) > f(\mathbf{x}^k)$ . Esse mecanismo de aceitação de soluções piores na direção do anti- $q$ -gradiente permite que o método escape de mínimos locais. Embora o algoritmo do método do  $q$ -gradiente sempre aceite soluções piores, as estratégias de obtenção do parâmetro  $\mathbf{q}$  e do tamanho do passo, quando bem empregadas, introduzem mecanismos que controlam de forma satisfatória este aceite.

Por fim, o Algoritmo 7 do método do  $q$ -gradiente é de fácil implementação e possui apenas três parâmetros de ajuste. A escolha desses parâmetros é descrita no próximo capítulo.

## 4 RESULTADOS NUMÉRICOS

Neste capítulo, são apresentados os resultados de experimentos numéricos do método do  $q$ -gradiente (Algoritmo 7) em funções teste da literatura de otimização contínua, irrestrita e não-linear para as dimensões 2, 10, 20, 30 e 50. Além de funções unimodais, também foram selecionadas funções multimodais com o objetivo de comprovar a capacidade que o método do  $q$ -gradiente exhibe de escapar de mínimos locais e de caminhar na direção do mínimo global, viabilizando sua aplicação em problemas de otimização global contínua.

Foram realizadas comparações de desempenho do método do  $q$ -gradiente com métodos determinísticos e estocásticos para otimização global. No caso determinístico, foram considerados o método da máxima descida e o *solver* GAMS/LINDOGlobal para otimização global determinística disponível no servidor NEOS (CZYZYK et al., 1998). No caso estocástico, o método do  $q$ -gradiente foi comparado extensivamente com os Algoritmos Genéticos (AGs) desenvolvidos por Deb et al. (2002), Ballester e Carter (2004), e os 11 Algoritmos Evolutivos (AEs) que participaram da competição do *IEEE Congress on Evolutionary Computation* em 2005 (CEC-2005).

Os algoritmos para o método do  $q$ -gradiente e o método da máxima descida foram implementados em Fortran 90 para o compilador IFORT (*Intel Fortran Compiler*), versão 10.1.018, e executados em uma máquina com processador Intel®Core™2 Duo T7250 com 4GB RAM e sistema operacional Ubuntu Linux 8.04 LTS 32-bit. Este capítulo também inclui uma discussão sobre os parâmetros de ajuste utilizados pelo método do  $q$ -gradiente.

### 4.1 Parâmetros de ajuste

A escolha adequada dos parâmetros de ajuste é fundamental para a eficiência de um algoritmo. Aperfeiçoamentos de algoritmos bem conhecidos se baseiam em parâmetros da literatura, como no caso dos AGs e do Recozimento Simulado, que possuem faixas de valores estabelecidas para cada um de seus parâmetros. Mesmo que se tenha domínio sobre o problema e sobre o método de otimização que será aplicado, a escolha dos parâmetros de ajuste para algoritmos, novos ou consolidados, sempre envolve um pouco de intuição e empirismo.

O método do  $q$ -gradiente, com a estratégia de obtenção do parâmetro  $\mathbf{q}$  segundo uma distribuição gaussiana e a estratégia de tamanho do passo por redução, requer três parâmetros de ajuste: desvio-padrão inicial ( $\sigma^0$ ), fator de redução ( $\beta$ ) e tamanho

do passo inicial ( $\alpha^0$ ).

O desvio-padrão inicial determina o quanto a busca é global ou local. Como descrito na Seção 3.4, a estratégia de geração dos parâmetros  $\mathbf{q}$ 's segundo uma distribuição gaussiana utiliza, na iteração  $k$ , o desvio-padrão  $\sigma^k$  e a média  $\mathbf{x}^k$ . Dessa forma, o desvio-padrão desta distribuição pode ser definido em função dos valores das variáveis de decisão  $\mathbf{x} \in \mathbb{R}^n$ . Pode-se calcular a amplitude do espaço de busca  $\Delta\mathbf{x}$  pela distância euclidiana entre os limites inferiores  $\mathbf{x}_I$  e superiores  $\mathbf{x}_S$  dessas variáveis (quando disponíveis), ou seja,  $\Delta\mathbf{x} = \|\mathbf{x}_S - \mathbf{x}_I\|$ , e considerar porcentagens desse valor para o desvio-padrão inicial  $\sigma^0$ .

Quando  $\sigma^0$  é grande, existe uma maior probabilidade das direções determinadas pelo anti- $q$ -gradiente serem quaisquer, mas à medida que  $\sigma^0$  tende a zero, os parâmetros  $\mathbf{q}$ 's tendem a  $\mathbf{1}$  e as direções do anti- $q$ -gradiente tendem à direção de máxima descida. Em funções multimodais, é interessante que as direções do anti- $q$ -gradiente apontem para diferentes direções do espaço de busca, incluindo direções de subida, para que o algoritmo, eventualmente, escape de mínimos locais. Logo,  $\sigma^0$  deve ser suficientemente grande no início do procedimento iterativo. Em contrapartida, o desvio-padrão  $\sigma^k$  deve diminuir ao longo das iterações para que o algoritmo migre suavemente de uma busca global para uma busca local. Já para funções unimodais,  $\sigma^0$  deve ser pequeno de forma que garanta direções de descida e que tendam para a direção de máxima descida, desde o início do procedimento iterativo.

O fator de redução  $\beta$  controla a velocidade de transição entre busca global e busca local. Valores pequenos de  $\beta$  reduzem rapidamente o valor de  $\sigma^0$  e as direções de busca, em poucas iterações, se transformam em direções de máxima descida. As simulações realizadas sobre funções multimodais mostraram que valores de  $\beta$  próximos de 1 reduzem o risco do método do  $q$ -gradiente ficar preso em um mínimo local. Logo, em funções multimodais espera-se que valores de  $\beta$  maiores que os utilizados em funções unimodais sejam mais eficientes. Os valores utilizados para o fator de redução  $\beta$  variaram de 0,80 até valores próximos de 1.

O tamanho do passo inicial  $\alpha^0$  está diretamente relacionado às características do espaço de busca e também deve ser definido com base no  $\Delta\mathbf{x}$ . Ao longo das iterações,  $\alpha^0$  deve diminuir de tamanho e ser suficientemente pequeno quando as direções de busca são de descida e próximas da direção de máxima descida. Por simplicidade, tanto  $\sigma^0$  quanto  $\alpha^0$  são reduzidos pelo mesmo fator de redução  $\beta$ .

Em geral, o desvio-padrão inicial e o passo inicial podem assumir porcentagens do



$\Delta \mathbf{x}$  como 1%, 25%, 50%, 75% e 100%. Experimentos numéricos mostraram que, em funções unimodais, é interessante escolher valores de  $\sigma^0$  menores, com porcentagens do  $\Delta \mathbf{x}$  abaixo de 50%. E, em funções multimodais, é aconselhável utilizar porcentagens maiores e acima de 50% do  $\Delta \mathbf{x}$ , para o desvio-padrão inicial, combinadas com fatores de redução próximos de 1.

A partir destas observações, o procedimento utilizado neste trabalho para a escolha dos parâmetros de ajuste seguiu os seguintes passos. Inicialmente foram definidas as porcentagens do  $\Delta \mathbf{x}$  para  $\sigma^0$  e  $\alpha^0$ , e valores para o fator de redução  $\beta \in [0, 80; 1)$ , de acordo com as características da função objetivo (unimodal ou multimodal). Depois, foram realizadas rodadas preliminares combinando os valores pré-definidos para um número máximo de avaliações da função objetivo, em geral, menor que o número total de avaliações disponíveis. As combinações que geraram resultados satisfatórios foram levadas em consideração na seleção final dos parâmetros de ajuste. Para algumas funções teste, realizou-se novas rodadas para combinações de valores estabelecidos dentro da faixa de valores mais promissoras.

## 4.2 Validações em funções teste

Nesta seção, a avaliação de desempenho do método do  $q$ -gradiente será realizada para um conjunto de quatro funções teste contínuas e não-lineares  $F : \mathbb{R}^n \rightarrow \mathbb{R}$ . As figuras a seguir ilustram apenas o caso bidimensional, em que  $n = 2$ .

A função De Jong, também conhecida como função *Sphere*, é dada por

$$F(\mathbf{x}) = \sum_{i=1}^n \mathbf{x}_i^2, \quad (4.1)$$

e ilustrada na Figura 4.1, sendo separável e unimodal (SOUZA, 2002).

A função *Ellipsoidal* dada por

$$F(\mathbf{x}) = \sum_{i=1}^n i \mathbf{x}_i^2, \quad (4.2)$$

também é unimodal e cada variável possui um peso diferente. As curvas de nível desta função são elipsóides concêntricos e alongados, como mostra a Figura 4.2 e, portanto, é considerada uma função mal escalonada (BALLESTER; CARTER, 2004).

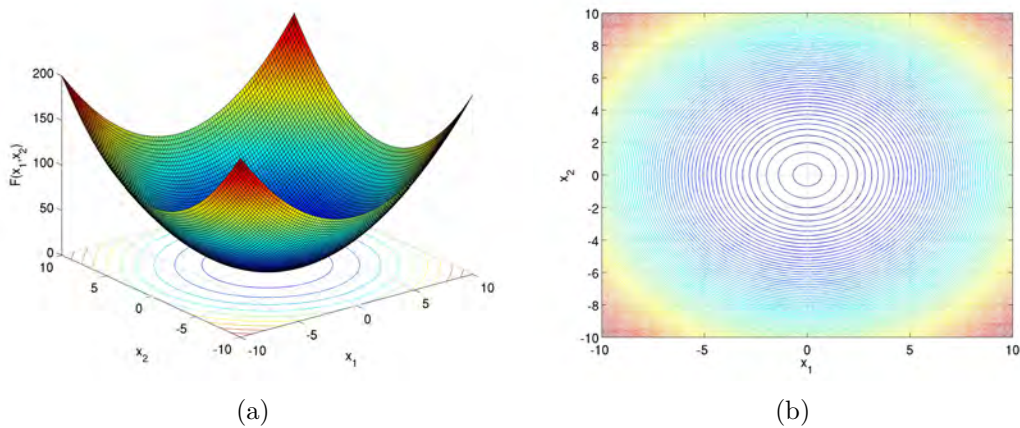


Figura 4.1 - Função De Jong

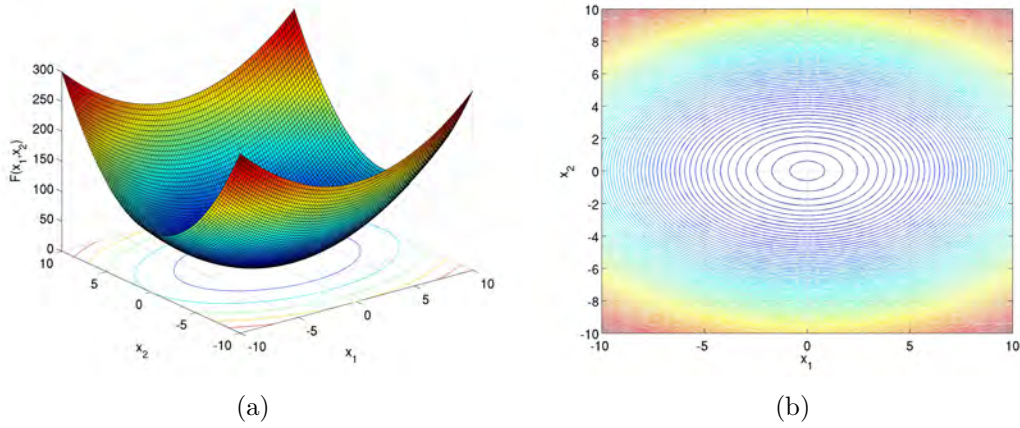


Figura 4.2 - Função *Ellipsoidal*

A função Rosenbrock dada por

$$F(\mathbf{x}) = \sum_{i=1}^{n-1} [100 \cdot (\mathbf{x}_i^2 - \mathbf{x}_{i+1})^2 + (1 - \mathbf{x}_i)^2], \quad (4.3)$$

possui mínimo global localizado dentro de um vale longo, estreito e com formato parabólico aplanado, como ilustra a Figura 4.3, e é conhecida por ser um caso clássico de convergência prematura de métodos de otimização.

Embora a função Rosenbrock seja considerada unimodal, experimentos numéricos mostram que, para  $4 \leq n \leq 30$ , a função possui dois mínimos: o global em  $\mathbf{x}^* = \mathbf{1}$  e um local que muda com a dimensionalidade (SHANG; QIU, 2006).

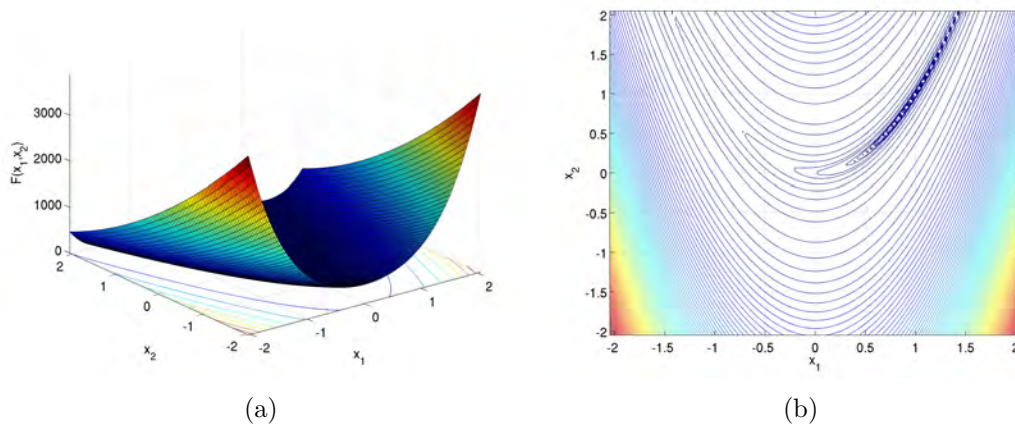


Figura 4.3 - Função Rosenbrock

Por fim, a função Rastrigin dada por

$$F(\mathbf{x}) = 10n + \sum_{i=1}^n (\mathbf{x}_i^2 - 10 \cos(2\pi\mathbf{x}_i)), \quad (4.4)$$

é multimodal e possui um envoltório parabólico com múltiplos mínimos locais, como ilustra a Figura 4.4 (SOUZA, 2002; BALLESTER; CARTER, 2004).

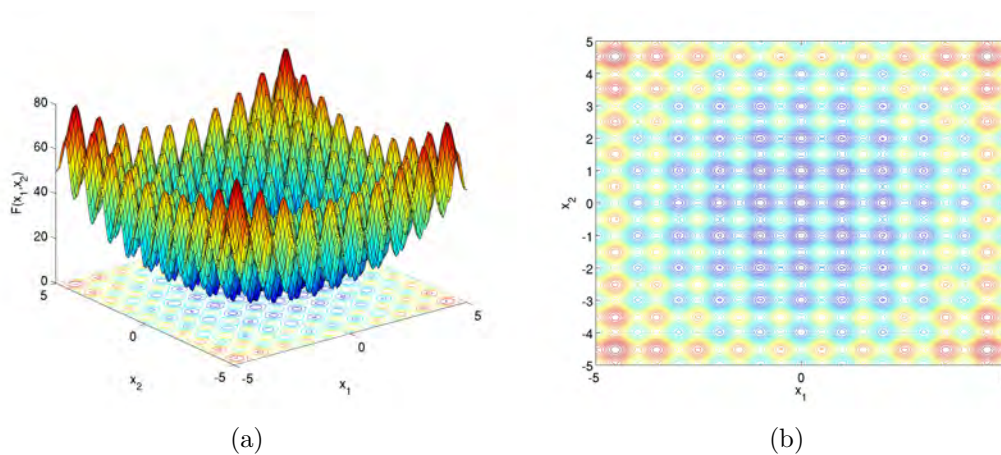


Figura 4.4 - Função Rastrigin

Para todas estas funções, o mínimo global é  $\mathbf{x}^* = \mathbf{0}$  com  $F(\mathbf{x}^*) = 0$ , exceto para a função Rosenbrock, em que  $\mathbf{x}^* = \mathbf{1}$ . Com o objetivo de verificar a eficiência do método do  $q$ -gradiente, o Algoritmo 7 foi aplicado sobre as funções teste com dimensão

$n = 2$  e os parâmetros ilustrados na Tabela 4.1. Esses parâmetros foram selecionados de acordo com as considerações apresentadas na Seção 4.1. Em seguida, aplicou-se o método da máxima descida (Algoritmo 1) sobre o mesmo conjunto de funções e com a mesma estratégia para o cálculo do tamanho do passo. Dessa forma, o algoritmo para o método da máxima descida é basicamente o algoritmo para o método do  $q$ -gradiente com  $\mathbf{q} = \mathbf{1}$  em todas as iterações.

Foram realizadas 50 execuções independentes dos algoritmos, a partir de 50 pontos iniciais distintos gerados no intervalo  $[-10; -5]$ . A escolha deste intervalo previne a geração de pontos iniciais próximos da origem que, em geral, é também o mínimo global de funções teste. O mesmo conjunto de pontos iniciais foi utilizado pelos dois métodos.

Tabela 4.1 - Parâmetros usados pelo método do  $q$ -gradiente.

Função	Parâmetros		
	$\sigma^0$	$\alpha^0$	$\beta$
De Jong	0,1	5	0,80
<i>Ellipsoidal</i>	0,1	5	0,80
Rosenbrock	0,1	0,1	0,9999
Rastrigin	20	0,1	0,9999

Para as funções unimodais De Jong e *Ellipsoidal*, verificou-se a velocidade utilizada por cada método para atingir a precisão 0,001 em termos do número de avaliações da função objetivo (NAFO). O critério de parada é satisfeito quando o erro da função é menor que a precisão desejada, ou seja,  $|F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)| \leq 0,001$  onde  $F(\mathbf{x}_{melhor})$  é o melhor valor da  $F$  obtido pelo algoritmo e  $F(\mathbf{x}^*)$  é o valor da função no mínimo global. A partir das 50 execuções independentes, foram obtidos o valor médio, melhor (ou mínimo), mediano e pior (ou máximo) do NAFO necessário para satisfazer o critério de parada sobre cada função e método, conforme ilustra a Tabela 4.2.

Para as funções Rosenbrock e Rastrigin, que requerem mais iterações para atingir a precisão estabelecida, o critério de parada utilizado foi um número máximo para o NAFO igual a  $10^5$ . A Tabela 4.3 ilustra a média dos melhores valores da função objetivo obtidos quando NAFO atingiu o NAFO máximo permitido em cada uma das 50 execuções independentes.

Tabela 4.2 - Número de avaliações da função objetivo necessário para satisfazer a condição  $|F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)| \leq 0,001$ .

Função	Método	NAFO			
		Média	Mínimo	Mediano	Máximo
De Jong	q-gradiente	<b>33.80</b>	10	37	58
	Máxima Descida	56.80	16	61	96
<i>Ellipsoidal</i>	q-gradiente	<b>49.16</b>	22	47	71
	Máxima Descida	83.00	56	86	111

Tabela 4.3 - Média dos melhores valores da função objetivo ( $F(\mathbf{x}_{melhor})$  Médio) obtidos em 50 execuções independentes dos algoritmos com NAFO máximo igual a  $10^5$ .

Função	Método	$F(\mathbf{x}_{melhor})$ Médio
Rosenbrock	q-gradiente	<b>2.0774E-07</b>
	Máxima Descida	1.4783E-02
Rastrigin	q-gradiente	<b>3.1362E-05</b>
	Máxima Descida	92.0512

As Figuras 4.5, 4.6, 4.7 e 4.8 ilustram o desempenho de cada função por meio da evolução da média dos melhores valores da função objetivo em 50 execuções independentes,  $F(\mathbf{x}_{melhor})$  Médio, versus o NAFO. Para gerar essas curvas, foram considerados  $10^3$  avaliações da função objetivo como critério de parada para as funções De Jong e *Ellipsoidal*, e  $10^5$  avaliações para as funções Rosenbrock e Rastrigin. As Figuras 4.7 e 4.8, no entanto, ilustram apenas as  $10^3$  primeiras avaliações com o intuito de melhorar a visualização. As curvas que não “tocam” o eixo horizontal indicam que o método atingiu o valor de mínimo global da função igual a zero ( $F(\mathbf{x}_{melhor}) = 0$ ) em todas as 50 execuções independentes e, neste caso, as curvas são interrompidas abruptamente porque não é possível representar o número zero em escala logarítmica. Alguns gráficos estão em escala logarítmica para auxiliar a visualização.

Os resultados da Tabela 4.2 e as Figuras 4.5 e 4.6 mostram que, para as funções unimodais De Jong e *Ellipsoidal*, o método do  $q$ -gradiente apresenta resultados ligeiramente melhores que os obtidos pelo método da máxima descida. Note, nas figuras,

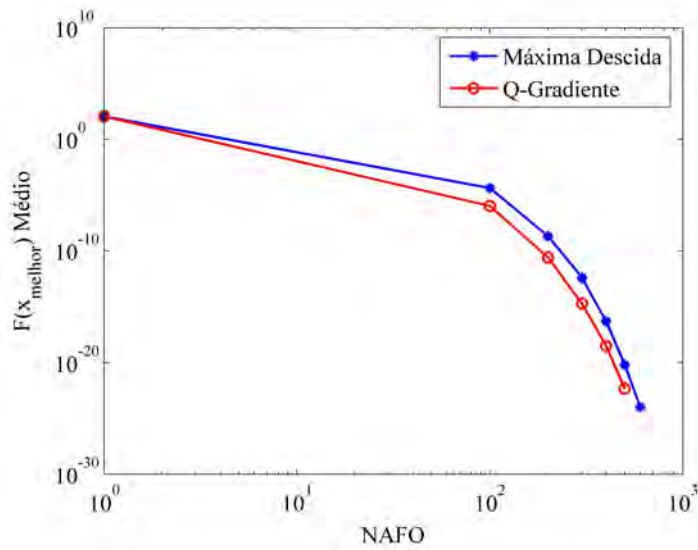


Figura 4.5 - Evolução de  $F(\mathbf{x}_{melhor})$  médio em função do NAFO para a função teste De Jong e os métodos do  $q$ -gradiente e da máxima descida.

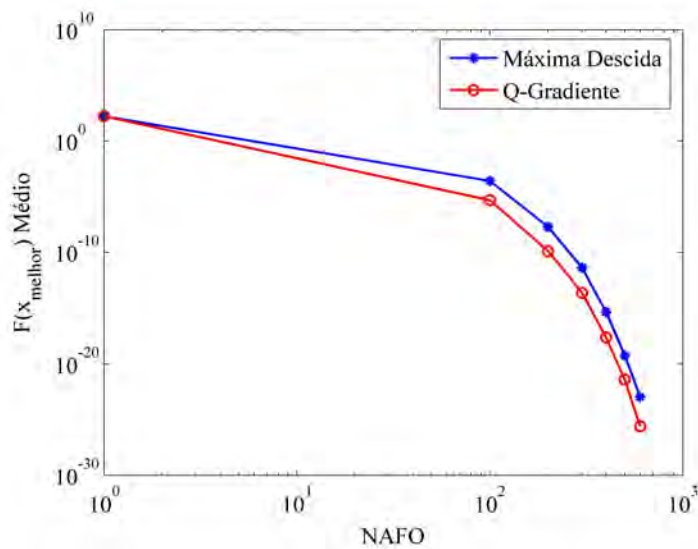


Figura 4.6 - Evolução de  $F(\mathbf{x}_{melhor})$  médio em função do NAFO para a função teste *Elipsoidal* e os métodos do  $q$ -gradiente e da máxima descida.

que o método do  $q$ -gradiente e o método da máxima descida apresentam curvas muito próximas. Os baixos valores selecionados para o desvio-padrão inicial  $\sigma^0$  e para o fator de redução  $\beta$  contribuíram para que o método do  $q$ -gradiente migrasse rapidamente para o método da máxima descida. Esses valores foram selecionados a partir de rodadas preliminares do Algoritmo 7 e foram utilizados porque outros valores, incluindo valores mais altos desses parâmetros, não geraram resultados melhores. O



valor  $\sigma^0 = 0,1$  foi responsável, no início do procedimento iterativo, pela geração de direções de descida, mas não de máxima descida, que garantiram o pequeno ganho do método do  $q$ -gradiente. Esse cenário muda quando as funções são a Rosenbrock, conhecida por ser de difícil minimização, e a função multimodal Rastrigin como ilustrado na Tabela 4.3 e nas Figuras 4.7 e 4.8.

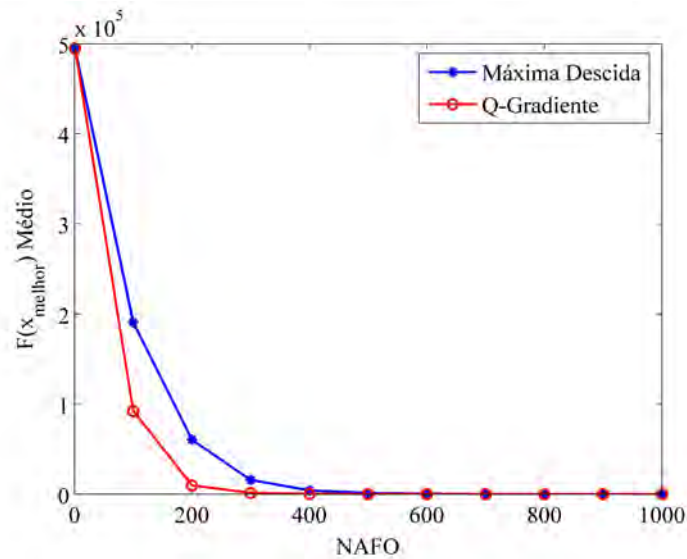


Figura 4.7 - Evolução de  $F(\mathbf{x}_{melhor})$  médio em função do NAFO para a função teste Rosenbrock e os métodos do  $q$ -gradiente e da máxima descida.

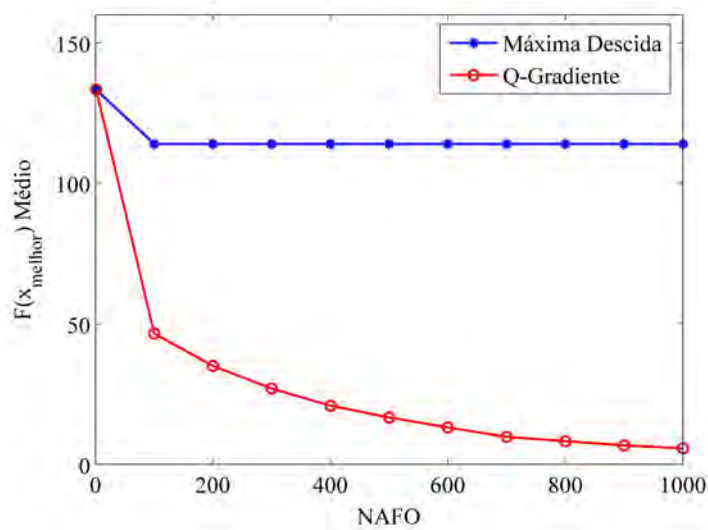


Figura 4.8 - Evolução de  $F(\mathbf{x}_{melhor})$  médio em função do NAFO para a função teste Rastrigin e os métodos do  $q$ -gradiente e da máxima descida.

Para a função Rosenbrock, o método do  $q$ -gradiente atinge uma precisão bem melhor que a obtida pelo método da máxima descida em  $10^5$  avaliações da função objetivo. Para a função Rastrigin, o mau desempenho do método da máxima descida já era esperado, pois o método é local e a função possui múltiplos mínimos locais. No entanto, o método do  $q$ -gradiente atingiu a bacia de atração do mínimo global em todas as 50 execuções independentes, enquanto o método da máxima descida, com a mesma estratégia de obtenção do tamanho do passo, não atingiu o mínimo global em nenhuma das execuções. Ao considerar  $4 \cdot 10^5$  avaliações da função objetivo como critério de parada, o método do  $q$ -gradiente converge para o mínimo global  $\mathbf{x}^* = \mathbf{0}$  com precisão dupla em todas as 50 execuções independentes. Logo, o  $F(\mathbf{x}_{melhor})$  médio também converge para zero e a curva termina abruptamente antes de tocar o eixo horizontal, como ilustra a Figura 4.9.

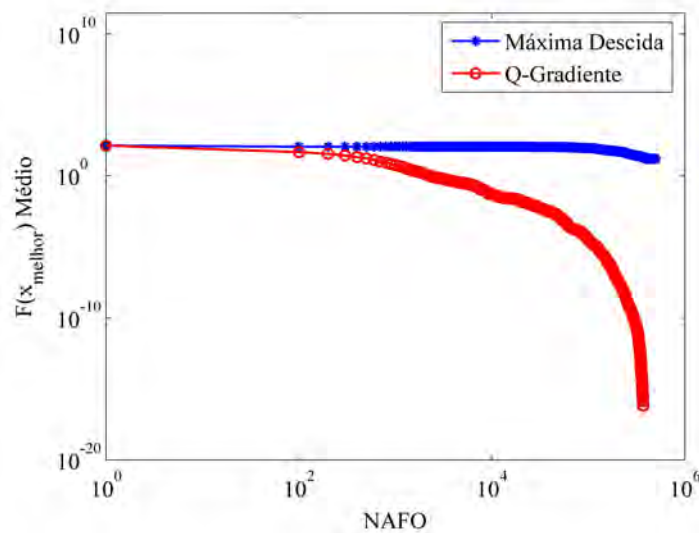


Figura 4.9 - Evolução de  $F(\mathbf{x}_{melhor})$  médio em função de  $4 \cdot 10^5$  avaliações da função objetivo para a função teste Rastrigin e os métodos do  $q$ -gradiente e da máxima descida.

Pode-se dizer que, sobre esse conjunto de funções teste, o desempenho do método do  $q$ -gradiente foi melhor que o do método da máxima descida. Para as funções teste unimodais, os dois métodos possuem comportamento similar. Já para a função Rosenbrock o método do  $q$ -gradiente foi mais eficaz que a sua versão clássica e, finalmente, para a função multimodal Rastrigin o método do  $q$ -gradiente foi capaz de escapar dos múltiplos mínimos locais e atingir o mínimo global em todas as execuções independentes, enquanto o método da máxima descida ficou preso em



mínimos locais e não atingiu a bacia de atração do mínimo global em nenhuma execução. A seguir, o método do  $q$ -gradiente é comparado com outros algoritmos considerados eficientes na resolução de problemas de otimização global.

### 4.3 Primeiro conjunto de funções teste

Com o objetivo de verificar a competitividade do método do  $q$ -gradiente, nesta seção são realizadas comparações de desempenho do método do  $q$ -gradiente com algoritmos estocásticos e determinísticos para otimização global sobre funções teste da literatura de otimização.

O primeiro conjunto de funções teste corresponde a três funções unimodais (*Ellipsoidal*, Schwefel e Rosenbrock) e três funções multimodais (Ackley, Rastrigin e Rastrigin Rotacionada). Todas as funções são contínuas, não-lineares e de  $n$  variáveis. Essas funções foram extraídas dos trabalhos de Deb et al. (2002) e Ballester e Carter (2004).

As funções *Ellipsoidal*, Rosenbrock e Rastrigin já foram descritas na Seção 4.2 e suas expressões são dadas, respectivamente, pelas Equações 4.2, 4.3 e 4.4. A função unimodal Schwefel é dada pela expressão

$$F(\mathbf{x}) = \sum_{i=1}^n \left( \sum_{j=1}^i \mathbf{x}_j \right)^2, \quad (4.5)$$

e o seu gráfico para duas dimensões é ilustrado na Figura 4.10. Note que as curvas de nível são elipsóides concêntricos, alongados e rotacionados.

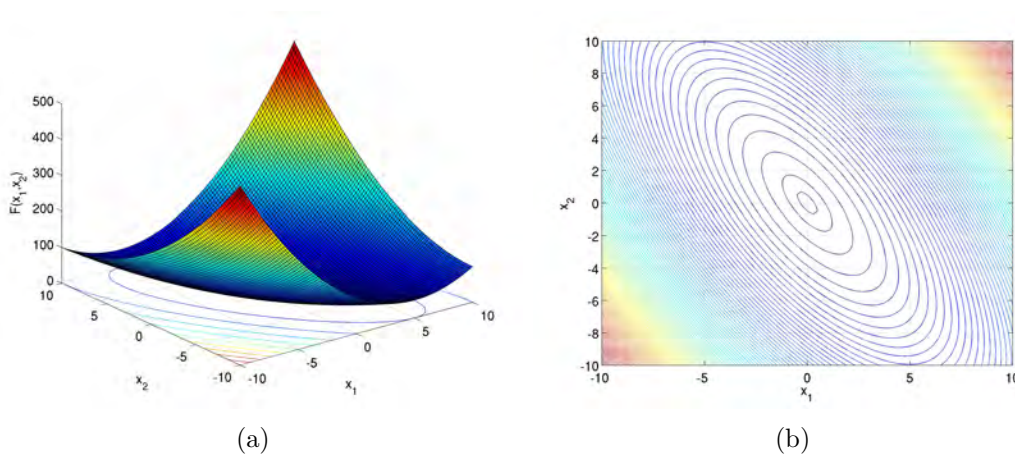


Figura 4.10 - Função Schwefel

A função multimodal Ackley é dada por

$$F(\mathbf{x}) = 20 + e - 20 \exp \left( -0,2 \sqrt{\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^2} \right) - \exp \left( \frac{1}{n} \sum_{i=1}^n \cos(2\pi \mathbf{x}_i) \right), \quad (4.6)$$

e o seu gráfico para  $n = 2$  é ilustrado na Figura 4.11. A função Ackley é altamente multimodal e o número de mínimos locais aumenta à medida que se afasta do mínimo global (BALLESTER; CARTER, 2004).

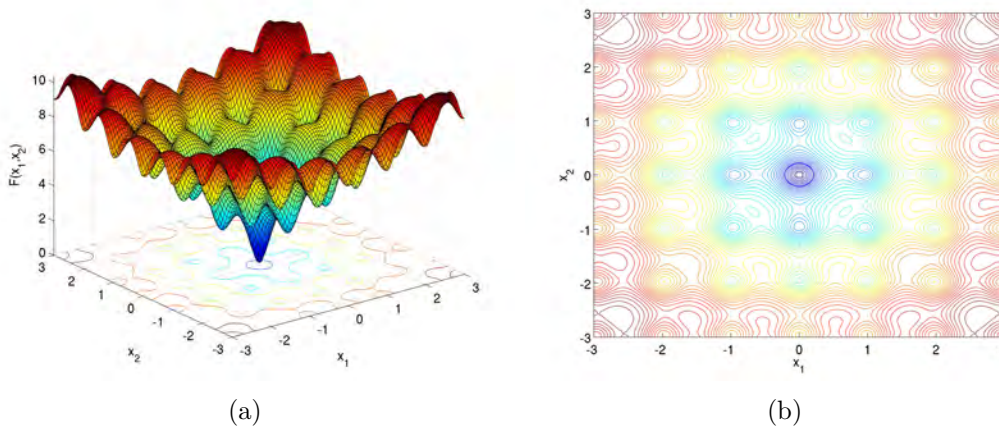


Figura 4.11 - Função Ackley

A função multimodal Rastrigin Rotacionada é dada por

$$F(\mathbf{x}) = 10n + \sum_{i=1}^n (\mathbf{y}_i^2 - 10 \cos(2\pi \mathbf{y}_i)), \quad \mathbf{y} = A \cdot \mathbf{x} \quad (4.7)$$

$$A_{i,i} = 4/5$$

$$A_{i,i+1} = 3/5 \text{ (} i \text{ ímpar)}$$

$$A_{i,i-1} = -3/5 \text{ (} i \text{ par)}$$

$$A_{i,j} = 0 \text{ (caso contrário)}$$

com gráfico para  $n = 2$  ilustrado na Figura 4.12. Essa função é similar à função Rastrigin, porém a rotação aplicada sobre ela faz com que os mínimos locais não estejam mais dispostos regularmente ao longo dos eixos coordenados (BALLESTER; CARTER, 2004).

Foram consideradas  $n = 20$  variáveis de decisão para cada função que possuem

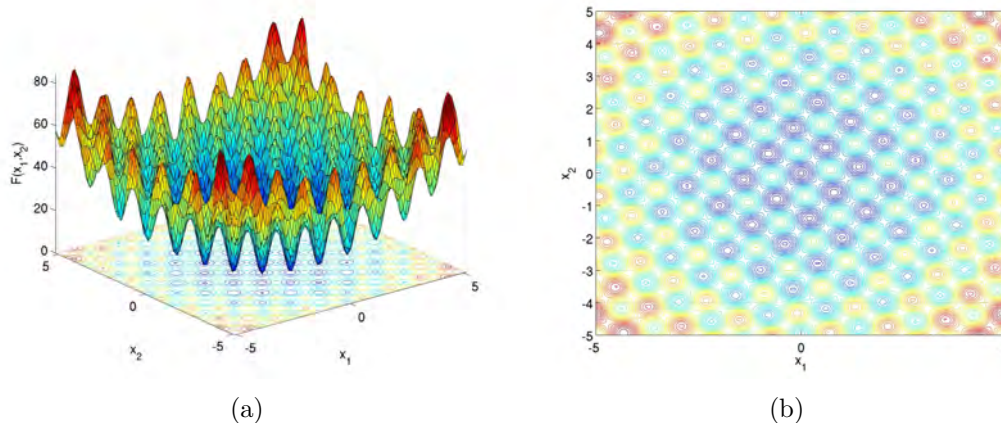


Figura 4.12 - Função Rastrigin Rotacionada

$F(\mathbf{x}^*) = 0$  com mínimo global  $\mathbf{x}^* = \mathbf{0}$ , exceto para a Rosenbrock em que  $\mathbf{x}^* = \mathbf{1}$ . Além de serem utilizadas com frequência na área de otimização contínua, permitindo a comparação direta com outros algoritmos, esse conjunto de funções teste possui características que tornam a sua minimização difícil e que estão presentes em muitos problemas reais (BALLESTER; CARTER, 2004).

#### 4.3.1 Comparação com algoritmos genéticos

A seguir, o método do  $q$ -gradiente é comparado com os AGs G3-PCX, SPC-vSBX e SPC-PNX, e o algoritmo evolutivo híbrido GEOvar+EE modificado para o primeiro conjunto de funções teste. O G3-PCX é uma versão de AG com codificação real e foi apresentado por Deb et al. (2002), exibindo bons resultados para as funções *Ellipsoidal*, Schwefel, Rosenbrock e Rastrigin. Os resultados do G3-PCX serviram de referência para os AGs com codificação real SPC-vSBX e SPC-PNX, desenvolvidos por Ballester e Carter (2004). Os resultados do G3-PCX para as funções Ackley e Rastrigin Rotacionada foram obtidos por Ballester e Carter (2004) a partir do código do G3-PCX disponível no site KanGAL (KANPUR GENETIC ALGORITHMS LABORATORY, 2005). Por fim, o GEOvar+EE modificado é uma hibridização do GEO com Estratégias Evolutivas (EE) e foi desenvolvido por Galski (2006). A Otimização Extrema Generalizada ou GEO (*Generalized Extremal Optimization*) e sua variante GEOvar são meta-heurísticas de otimização global desenvolvidas por Souza (2002) como uma proposta de generalização do algoritmo  $\tau$ -EO de Boettcher e Percus (2001).

Os resultados do método do  $q$ -gradiente foram obtidos para o melhor ajuste de

parâmetros encontrados sobre cada função teste, de acordo com o procedimento descrito da Seção 4.1. Logo, os melhores valores para  $\sigma^0$ ,  $\alpha^0$  e  $\beta$  estão ilustrados na Tabela 4.4.

Tabela 4.4 - Parâmetros usados pelo método do  $q$ -gradiente para o primeiro conjunto de funções teste.

Funções	$\sigma^0$	$\alpha^0$	$\beta$
<i>Ellipsoidal</i>	0,4	38	0,86
Schwefel	0,1	1	0,997
Rosenbrock	0,1	0,1	0,9995
Ackley	20	12	0,90
Rastrigin	21	0,3	0,9995
Rastrigin Rotacionada	30	0,5	0,999

Foram realizadas 50 execuções independentes do algoritmo do método do  $q$ -gradiente para 50 pontos iniciais distintos. Assim como em Deb et al. (2002) e Ballester e Carter (2004), o intervalo de inicialização utilizado pelo método do  $q$ -gradiente foi  $[-10; -5]$ . O GEOvar+EE modificado utiliza, para cada função, um intervalo de inicialização diferente, conforme descrito em Galski (2006). Os critérios de parada utilizados por todos os algoritmos foram: (i) atingir a precisão desejada  $10^{-20}$ , ou seja, o procedimento iterativo termina quando o erro da função é menor que a precisão desejada ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*) \leq 10^{-20}$ ); ou (ii) atingir um NAFO máximo, ou seja, o algoritmo para se um número máximo de avaliações da função objetivo igual a  $10^6$  é alcançado.

As Tabelas 4.5 e 4.6<sup>1</sup> ilustram uma comparação entre os algoritmos sobre o primeiro conjunto de funções teste. Cada execução independente que atingir a precisão desejada  $10^{-20}$  antes de completar as  $10^6$  avaliações da função objetivo tem o NAFO armazenado. Logo, apenas os valores de NAFO das execuções de sucesso são ordenados e exibidos nas colunas “Melhor” (menor), “Mediano” e “Pior” (maior). Se a precisão desejada não é atingida antes de  $10^6$  avaliações, então o melhor valor da função objetivo encontrado dentre todas as execuções independentes é informado na coluna “ $F(\mathbf{x}_{melhor})$ ”. A coluna “Sucesso” exibe o número de execuções independentes

<sup>1</sup>O algoritmo GEOvar+EE modificado aparece nas Tabelas 4.5 e 4.6 apenas como GEOvar+EE por simplicidade de notação.

que atingiram a precisão desejada (para as funções unimodais) ou que alcançaram a bacia de atração do mínimo global (para as funções multimodais). Os melhores valores estão em negrito.

Tabela 4.5 - Desempenho comparativo, em termos no número de avaliações da função objetivo (NAFO), entre o algoritmo para o método do  $q$ -gradiente, os AGs G3-PCX, SPC-vSBX e SPC-PNX, e a meta-heurística GEOvar+EE modificado para as funções unimodais *Ellipsoidal* e Schwefel, e a função Rosenbrock.

Função	Algoritmo	Melhor	Mediano	Pior	$F(\mathbf{x}_{melhor})$	Sucesso
Ellipsoidal	<b>G3-PCX</b>	<b>5.826</b>	<b>6.800</b>	<b>7.728</b>	<b><math>10^{-20}</math></b>	<b>10/10</b>
	SPC-vSBX	49.084	50.952	57.479	$10^{-20}$	10/10
	SPC-PNX	36.360	39.360	40.905	$10^{-20}$	10/10
	GEOvar+EE	16.808	33.315	197.309	$10^{-20}$	50/50
	<b><math>q</math>-Gradiente</b>	<b>5.905</b>	<b>7.053</b>	<b>7.381</b>	<b><math>10^{-20}</math></b>	<b>50/50</b>
Schwefel	<b>G3-PCX</b>	<b>13.988</b>	<b>15.602</b>	<b>17.188</b>	<b><math>10^{-20}</math></b>	<b>10/10</b>
	SPC-vSBX	260.442	294.231	334.743	$10^{-20}$	10/10
	SPC-PNX	236.342	283.321	299.301	$10^{-20}$	10/10
	GEOvar+EE	477.887	685.512	957.552	$10^{-20}$	49/50
	$q$ -Gradiente	289.174	296.103	299.178	$10^{-20}$	50/50
Rosenbrock	<b>G3-PCX</b>	<b>16.508</b>	<b>21.452</b>	<b>25.520</b>	<b><math>10^{-20}</math></b>	<b>36/50</b>
	SPC-vSBX	$10^6$	-	-	$10^{-4}$	48/50
	SPC-PNX	$10^6$	-	-	$10^{-10}$	38/50
	GEOvar+EE	$10^6$	-	-	$4 \cdot 10^{-8}$	50/50
	$q$ -Gradiente	$10^6$	-	-	$10^{-10}$	50/50

Na Tabela 4.5, para a função *Ellipsoidal* todos os algoritmos atingiram a precisão desejada ( $10^{-20}$ ) em todas as execuções independentes. Porém, o método do  $q$ -gradiente e o G3-PCX satisfazem esse critério de parada para os menores números de avaliações da função objetivo e com desempenho similar (por isso os dois aparecem em negrito para esta função). Para a função Schwefel, embora o método do  $q$ -gradiente alcance a precisão desejada em todas as execuções, o G3-PCX apresenta o melhor desempenho. Por fim, para a função Rosenbrock, conhecida por ser um exemplo de convergência prematura para métodos de otimização, o único algoritmo que atinge a precisão desejada é o G3-PCX. O método do  $q$ -gradiente fica apenas atrás do G3-PCX, atingindo a segunda melhor precisão ( $10^{-10}$ ) em mais execuções

independentes em comparação com os demais algoritmos.

Tabela 4.6 - Desempenho comparativo, em termos no número de avaliações da função objetivo (NAFO), entre o algoritmo para o método do  $q$ -gradiente, os AGs G3-PCX, SPC-vSBX e SPC-PNX, e a meta-heurística GEOvar+EE modificado para as funções multimodais Ackley, Rastrigin e Rastrigin Rotacionada.

Função	Algoritmo	Melhor	Mediano	Pior	$F(\mathbf{x}_{melhor})$	Sucesso
Ackley	G3-PCX	$10^6$	-	-	3.959	0
	SPC-vSBX	57.463	63.899	65.902	$10^{-10}$	10/10
	SPC-PNX	45.736	48.095	49.392	$10^{-10}$	10/10
	GEOvar+EE	17.914	23.616	39.308	$10^{-10}$	50/50
	<b><math>q</math>-Gradiente</b>	<b>11.850</b>	<b>12.465</b>	<b>13.039</b>	<b><math>10^{-15}</math></b>	<b>50/50</b>
Rastrigin	G3-PCX	$10^6$	-	-	15.936	0
	SPC-vSBX	260.685	306.819	418.482	$10^{-20}$	6/10
	SPC-PNX	$10^6$	-	-	4.975	0
	<b>GEOvar+EE</b>	<b>18.584</b>	<b>46.655</b>	<b>166.877</b>	<b><math>10^{-20}</math></b>	<b>50/50</b>
	$q$ -Gradiente	676.050	692.450	705.037	$10^{-20}$	48/50
Rastrigin Rotacionada	G3-PCX	$10^6$	-	-	309,429	0
	SPC-vSBX	$10^6$	-	-	8,955	0
	SPC-PNX	$10^6$	-	-	3,980	0
	GEOvar+EE	$10^6$	-	-	4,62	0/50
	<b><math>q</math>-Gradiente</b>	<b>541.857</b>	<b>545.957</b>	<b>549.114</b>	<b><math>10^{-20}</math></b>	<b>20/50</b>

Na Tabela 4.6, para a função Ackley a precisão desejada é  $10^{-15}$  para o método do  $q$ -gradiente e  $10^{-10}$  para os demais algoritmos<sup>2</sup>. Todos os algoritmos, exceto o G3-PCX, atingiram a bacia de atração do mínimo global e a precisão desejada de acordo com seus limites computacionais. No entanto, o método do  $q$ -gradiente satisfaz essa condição com o menor número de avaliações da função objetivo. Para a função Rastrigin, novamente o G3-PCX e agora o SPC-PNX não alcançaram a bacia de atração do mínimo global em nenhuma execução. O método do  $q$ -gradiente alcançou a bacia de atração do mínimo global em 96% das execuções, porém o GEOvar+EE modificado apresenta o melhor desempenho para esta função. Finalmente, para a função Rastrigin Rotacionada o método do  $q$ -gradiente foi o único a atingir a bacia

<sup>2</sup>A função Ackley avaliada em  $\mathbf{x}^* = \mathbf{0}$  com precisão dupla não é zero. Para o compilador utilizado neste trabalho, tem-se  $F(\mathbf{x}^*) = -0.444089209850062E - 15$ .



de atração do mínimo global, e fez isso em 40% das execuções.

Comparando os resultados para as funções unimodais ou quase unimodais (*Ellipsoidal*, Schwefel e Rosenbrock) o escore é favorável ao G3-PCX, com duas vitórias e um empate. Se o algoritmo G3-PCX não for considerado sobre estas funções, o método do  $q$ -gradiente é o vitorioso com duas vitórias (funções *Ellipsoidal* e Rosenbrock) e um empate com os métodos SPC-vSBX e SPC-PNX (função Schwefel). Já para as funções multimodais (Ackley, Rastrigin e Rastrigin Rotacionada) o G3-PCX é claramente inferior. Para essas funções, o escore é favorável para o método do  $q$ -gradiente com duas vitórias (funções Ackley e Rastrigin Rotacionada), seguido do GEOvar+EE modificado com uma vitória (função Rastrigin). Vale dizer que o intervalo de inicialização do GEOvar+EE modificado para a função Rastrigin é o intervalo  $[-5, 12; 5, 12]$  simétrico em relação ao mínimo global desta função, e não o intervalo  $[-10; -5]$  utilizado pelos outros métodos.

Em resumo, para esse primeiro conjunto de funções teste, o método do  $q$ -gradiente empata com o G3-PCX com um escore de duas vitórias e um empate para cada um. No entanto, sobre as funções multimodais o método do  $q$ -gradiente apresenta o melhor desempenho geral em relação aos algoritmos G3-PCX, SPC-vSBX, SPC-PNX e GEOvar+EE.

#### 4.3.2 Comparação com otimização global determinística

Embora a proposta do método do  $q$ -gradiente seja a de um método estocástico para otimização global, a comparação com métodos determinísticos é inevitável. Isso porque o método do  $q$ -gradiente é uma generalização de um dos mais conhecidos métodos determinísticos: o método da máxima descida.

Um pacote ou *solver* de otimização determinística considerado eficiente na resolução de problemas de otimização global é o LINDOGlobal, pertencente ao software comercial de otimização GAMS (*General Algebraic Modeling System*). O GAMS é uma linguagem de modelagem algébrica de alto nível para problemas de otimização complexos e de grande escala. A linguagem GAMS possui uma sintaxe simples e com foco na modelagem de problemas de otimização lineares, não-lineares e de programação inteira mista (ROENTHAL, 2012). Uma lista completa dos tipos de modelos e *solvers* suportados pelo GAMS pode ser encontrada na página oficial do software (GAMS DEVELOPMENT CORPORATION, 2001).

O *solver* LINDOGlobal disponível em linguagem GAMS (GAMS/LINDOGlobal) é

um pacote para problemas de otimização global não-lineares com variáveis contínuas ou discretas. Esse *solver* suporta a maioria das funções matemáticas, incluindo funções trigonométricas. O procedimento de otimização global do GAMS/LINDOGlobal emprega algoritmos do tipo *branch-and-bound* para particionar um problema de otimização não-linear em vários subproblemas e resolver cada subproblema como um problema linear ou convexo relaxado. Esses subproblemas podem ser particionados novamente em outros subproblemas, e assim sucessivamente. Dadas tolerâncias adequadas e depois de um número grande e finito de passos, a solução, provavelmente global, é encontrada. Problemas multimodais não são os mais apropriados para a abordagem do GAMS/LINDOGlobal e, neste caso, o *solver* pode utilizar reinicializações para auxiliar na identificação dos vários mínimos locais. O melhor mínimo encontrado durante o procedimento de otimização é reportado como solução do problema (LINDO SYSTEMS INC, 2001).

A versão gratuita do GAMS/LINDOGlobal é limitada a problemas com 10 variáveis. No entanto, o servidor NEOS (CZYZYK et al., 1998) fornece acesso a esse e outros *solvers* sem nenhum custo. Para resolver um problema de otimização, o usuário acessa a página do NEOS, seleciona o *solver* LINDOGlobal (ou outro), submete o problema modelado em linguagem GAMS (ou outra) e aguarda enquanto o problema é resolvido remotamente. A solução é apresentada na tela ou enviada por correio eletrônico.

O primeiro conjunto de funções teste descrito na Seção 4.3 foi modelado em linguagem GAMS para o *solver* LINDOGlobal, com todas as opções de ajuste padrão. Para o método do  $q$ -gradiente, foram considerados os mesmos parâmetros da Tabela 4.4 em 10 execuções independentes com critérios de parada: (i) atingir um número máximo de avaliações da função objetivo igual a  $10^6$ , ou (ii)  $\|F(\mathbf{x}^k)\| \leq 10^{-20}$ , isto é, vetor gradiente clássico menor que  $10^{-20}$ . Os mesmos 10 pontos iniciais gerados segundo sorteio uniforme no intervalo  $[-10, -5]$  foram utilizados tanto pelo método do  $q$ -gradiente quanto pelo *solver* GAMS/LINDOGlobal.

Infelizmente, o GAMS/LINDOGlobal não fornece o número de avaliações da função objetivo utilizadas durante o procedimento de otimização. Dessa forma, a comparação foi realizada apenas em relação aos melhores valores encontrados para cada função teste,  $F(\mathbf{x}_{melhor})$ , em 10 execuções independentes, tanto para o método do  $q$ -gradiente quanto para o GAMS/LINDOGlobal, como ilustra a Tabela 4.7.

Note que, para as funções unimodais convexas *Ellipsoidal* e Schwefel, o *solver* GAMS/LINDOGlobal, como esperado, apresenta os melhores resultados em termos



Tabela 4.7 -  $F(\mathbf{x}_{melhor})$  encontrado pelo *solver* GAMS/LINDOGlobal e pelo método do  $q$ -gradiente sobre o primeiro conjunto de funções teste.

Funções	$q$ -gradiente	LINDOGlobal
Elipsoidal	3,4710E-28	0
Schwefel	5,7429E-28	0
Rosenbrock	9,3348E-11	6,9877E-08
Ackley	-4,4409E-16	-6,0396E-14
Rastrigin	0	0,99495905
Rastrigin Rotacionada	0	-1.2506E-12

da precisão alcançada. No entanto, para funções multimodais Rosenbrock, Ackley, Rastrigin e Rastrigin Rotacionada, o método do  $q$ -gradiente apresenta desempenho superior ao do GAMS/LINDOGlobal.

Para as funções Rosenbrock e Ackley, o método do  $q$ -gradiente atinge uma melhor precisão. Para a função Rastrigin, o *solver* não alcança a bacia de atração do mínimo global para nenhum ponto inicial distinto e o melhor valor encontrado é  $F(\mathbf{x}_{melhor}) = 0,99495905$ . Já o método do  $q$ -gradiente atinge a bacia do mínimo global em 9 de 10 execuções independentes com  $F(\mathbf{x}_{melhor}) = 0$ . Para a função Rastrigin Rotacionada, o *solver* atinge a bacia de atração do mínimo global para 3 pontos distintos, com precisão menor que a obtida pelo método do  $q$ -gradiente, que alcança a bacia de atração do mínimo global em 8 de 10 execuções independentes com  $F(\mathbf{x}_{melhor}) = 0$ .

#### 4.4 Segundo conjunto de funções teste

Uma vez definido que o método do  $q$ -gradiente seria comparado com AEs, pensou-se em um conjunto de funções teste abrangente e bem difundido. Logo, optou-se pelos problemas da competição *IEEE Congress on Evolutionary Computation* (CEC), organizado anualmente pela *IEEE Computational Intelligence Society*.

Cada ano uma classe de problemas é selecionada para a competição. Em 2005, foi definido um conjunto de 25 funções com restrições laterais para as dimensões 10, 30 e 50, sendo 5 unimodais e 20 multimodais. As funções unimodais (F1 a F5) e as sete primeiras funções multimodais (F6 a F12) são funções básicas da área de otimização contínua (como as funções *Sphere*, Schwefel, Rosenbrock, Ackley e Rastrigin já definidas na Subseção 4.3, entre outras) com a adição de deslocamento, rotação e/ou

ruído. As demais funções são expansões (F13 e F14) ou composições híbridas (F15 a F25) das funções anteriores. A Tabela 4.8 ilustra as principais características de F1 a F25. A definição completa das funções teste pode ser encontrada no relatório técnico elaborado por Suganthan et al. (2005).

Tabela 4.8 - Principais características das funções teste do CEC-2005.

Unimodal	F1 a F5
Multimodal	F6 a F25
Deslocada	F1 a F4; F6 a F14
Rotacionada	F3; F7; F8; F10; F11; F16 a F18; F21; F24
Separável	F1; F9; F10; F15 (próximo de $\mathbf{x}^*$ )
Não separável	F2 a F8; F10 a F14; F16 a F25
Escalonável	F1 a F3; F5 a F9; F11 a F25
Com ruído	F4; F17
Descontínua	F23 a F25

Participaram da competição promovida pelo CEC-2005 11 AEs com codificação real: BLX-GL50 (GARCIA-MARTINEZ; LOZANO, 2005), BLX-MA (MOLINA et al., 2005), CoEVO (POSIK, 2005), DE (RONKKONEN et al., 2005), DMS-L-PSO (LIANG; SUGANTHAN, 2005), EDA (YUAN; GALLAGHER, 2005), G-CMA-ES (AUGER; HANSEN, 2005b), K-PCX (SINHA et al., 2005), L-CMA-ES (AUGER; HANSEN, 2005a), L-SaDE (QIN; SUGANTHAN, 2005) e SPC-PNX<sup>3</sup> (BALLESTER et al., 2005). Nem todos os algoritmos consideraram a dimensão  $n = 50$ . Para a dimensão  $n = 10$ , nenhum algoritmo conseguiu resolver de forma satisfatória (i.e., atingir um dos critérios de parada) as funções F8, F13, F14, e F16 a F25. Essas funções mais a função F15 também não foram resolvidas por nenhum algoritmo para  $n = 30$ .

Para o segundo conjunto de funções teste utilizado na análise de desempenho do método do  $q$ -gradiente, selecionou-se as 15 primeiras funções teste do CEC-2005. Esse conjunto de funções foi resolvido pela maioria dos AEs e é representativo das características expostas na Tabela 4.8. As 15 funções podem ser divididas em 5 funções unimodais (F1 a F5) e 10 multimodais (F6 a F15). Dentre as multimodais, 7 são funções definidas a partir de funções básicas (F6 a F12), 2 são expansões de funções básicas (F13 e F14) e 1 é uma composição híbrida das funções básicas (F15).

<sup>3</sup>O SPC-PNX também aparece nas comparações do primeiro conjunto de funções teste.

Foram consideradas as dimensões 10, 30 e 50 para cada função teste.

Seguindo as definições da competição, foram realizadas 25 execuções independentes do Algoritmo 7 para o método do  $q$ -gradiente sobre as 15 funções teste. Os pontos iniciais foram obtidos por meio de sorteio uniforme dentro dos intervalos de inicialização especificados para cada função. Os critérios de parada utilizados foram (i) atingir um NAFO máximo igual a  $10.000 \cdot n$ , em que  $n$  é o número de dimensões (10, 30 ou 50); ou (ii) atingir a precisão final  $10^{-8}$ , ou seja, o procedimento iterativo termina se o erro da função dado por  $(F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*))$  for menor que  $10^{-8}$ .

Para cada função, dimensão e execução distinta, o valor do erro da função  $(F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*))$  é armazenado após 1.000, 10.000, 100.000 e  $10.000 \cdot n$  avaliações da função objetivo ou até que a precisão final  $10^{-8}$  seja alcançada. Após 25 execuções independentes, os valores armazenados são ordenados do menor (melhor) para o maior (pior) e apresentados na sequência: 1° (melhor), 7°, 13° (mediano), 19° e 25° (pior), juntamente com a média (ME) e o desvio-padrão (DP) considerando as 25 execuções.

Um raciocínio análogo é empregado sobre o NAFO. Em cada função, dimensão e execução distinta, o NAFO necessário para que uma “precisão estabelecida” seja alcançada é armazenado. Após 25 execuções distintas esses valores são ordenados do menor (melhor) para o maior (pior) e apresentados na sequência: 1° (melhor), 7°, 13° (mediano), 19° e 25° (pior). A média (ME) e o desvio-padrão (DP) são calculados apenas sobre as execuções bem sucedidas. Uma execução é considerada bem sucedida se a “precisão estabelecida” for atingida antes do número de avaliações atingir o NAFO máximo ( $10.000 \cdot n$ ). A “precisão estabelecida” para cada função é diferente da precisão final  $10^{-8}$  e pode ser encontrada no Apêndice A .

Taxa de sucesso (TS) e uma medida de análise de desempenho (AD) são calculadas de acordo com as expressões

$$TS = \frac{\text{número de execuções bem sucedidas}}{\text{número total de execuções}}, \quad (4.8)$$

e

$$AD = \frac{\text{média do NAFO das execuções bem sucedidas} \times \text{número total de execuções}}{\text{número de execuções bem sucedidas}}. \quad (4.9)$$

A taxa de sucesso (TS) dá uma medida entre 0 a 1 do número de execuções bem

sucedidas. Se  $TS = 0$ , então nenhuma execução independente foi bem sucedida. E se  $TS = 1$ , então todas as execuções independentes foram bem sucedidas. Já a análise de desempenho (AD) dá uma medida do NAFO utilizado por cada método sobre uma função teste. Quanto maior for o valor para AD, pior é o desempenho do método.

A Tabela 4.9 ilustra os parâmetros de ajuste utilizados pelo método do  $q$ -gradiente para as funções F1 a F15 e dimensões 10, 30 e 50. Esses parâmetros foram obtidos conforme procedimento descrito na Seção 4.1. Os valores do NAFO necessário para atingir a precisão estabelecida, a média (ME), o desvio-padrão (DP), a taxa de sucesso (TS) e a medida de análise de desempenho (AD) para cada função estão ilustradas nas Tabelas 4.10, 4.11 e 4.12, respectivamente, para as dimensões 10, 30 e 50. Nessas tabelas, o símbolo “-” representa que o valor do NAFO atingiu o NAFO máximo igual a 100.000, 300.000 e 500.000 avaliações, respectivamente, para as dimensões 10, 30 e 50. Os valores do erro da função ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) sobre cada função teste e dimensão estão no Apêndice A.

Tabela 4.9 - Parâmetros de ajuste utilizados pelo método do  $q$ -gradiente para as funções F1 a F15 com as dimensões 10, 30 e 50.

Funções	$n = 10$			$n = 30$			$n = 50$		
	$\sigma^0$	$\alpha^0$	$\beta$	$\sigma^0$	$\alpha^0$	$\beta$	$\sigma^0$	$\alpha^0$	$\beta$
F1	10	60	0,80	13,33	120	0,80	30	140	0,8
F2	15	150	0,98	11	190	0,997	4	190	0,998
F3	5	170	0,999	16,67	200	0,999	1,67	160	0,999
F4	40	12	0,998	175	38	0,9994	175	16	0,9998
F5	0,33	160	0,99	8,33	20	0,999	1,67	40	0,999
F6	1,67	10	0,998	1	8	0,999	0,30	18	0,998
F7	490	19	0,999	75	70	0,99	50	20	0,998
F8	1	10	0,995	1	20	0,998	1	5	0,998
F9	16	0,40	0,998	23	0,40	0,999	39	0,70	0,999
F10	50	1,10	0,996	45	0,90	0,999	40	1,50	0,995
F11	0,50	1	0,995	0,50	0,70	0,995	0,5	1	0,999
F12	0,60	0,90	0,999	0,30	0,50	0,999	0,10	1	0,995
F13	0,90	0,90	0,995	0,20	0,10	0,999	0,30	0,10	0,999
F14	10	10	0,999	5	10	0,999	5	15	0,999
F15	5	0,60	0,996	5	0,50	0,998	5	1	0,995

Tabela 4.10 - Valores do NAFO para as funções F1 a F15 e  $n = 10$ .

Função	1°	7°	13°	19°	25°	ME	DP	TS	AD
F1	485	504	518	536	563	5,2200E+02	2,0160E+01	1,00	5,2200E+02
F2	7.484	7.957	8.289	8.596	9.490	8,3090E+03	5,0594E+02	1,00	8,3090E+03
F3	-	-	-	-	-	-	-	0	-
F4	53.186	56.133	57.529	58.986	62.799	5,7754E+04	2,4260E+03	1,00	5,7754E+04
F5	-	-	-	-	-	-	-	0	-
F6	68.837	70.205	71.070	71.585	73.604	7,0992E+04	1,0729E+03	1,00	7,0992E+04
F7	59.775	64.230	66.848	85.937	94.352	7,2876E+04	1,1871E+04	1,00	7,2876E+04
F8	-	-	-	-	-	-	-	0	-
F9	27.064	29.390	30.500	31.437	-	2,9836E+04	1,2030E+03	0,76	3,9258E+04
F10	19.872	-	-	-	-	2,0049E+04	1,5690E+02	0,12	1,6708E+05
F11	-	-	-	-	-	-	-	0	-
F12	72.241	74.628	-	-	-	7,4013E+04	1,2170E+03	0,44	1,6821E+05
F13	7.143	-	-	-	-	7,1430E+03	0,0000E+00	0,04	1,7858E+05
F14	-	-	-	-	-	-	-	0	-
F15	16.165	-	-	-	-	1,5860E+04	4,6174E+02	0,08	1,9824E+05

Tabela 4.11 - Valores do NAFO para as funções F1 a F15 e  $n = 30$ .

Função	1°	7°	13°	19°	25°	ME	DP	TS	AD
F1	1.645	1.689	1.744	1.966	2.035	1,8210E+03	1,4070E+02	1,00	1,8210E+03
F2	20.9833	222.825	231.545	245.282	252.979	2,3229E+05	1,3318E+04	1,00	2,3229E+05
F3	-	-	-	-	-	-	-	0	-
F4	-	-	-	-	-	-	-	0	-
F5	-	-	-	-	-	-	-	0	-
F6	221.606	-	-	-	-	2,2101E+05	0,0000E+00	0,04	2,2101E+05
F7	17.889	18.447	18.787	20.462	25.338	1,9479E+04	1,5663E+03	1,00	1,9479E+04
F8	-	-	-	-	-	-	-	0	-
F9	187.994	196.627	-	-	-	1,9243E+05	3,6129E+03	0,32	6,0134E+05
F10	-	-	-	-	-	-	-	0	-
F11	-	-	-	-	-	-	-	0	-
F12	224.751	-	-	-	-	2,3590E+05	1,5763E+04	0,08	2,9487E+06
F13	-	-	-	-	-	-	-	0	-
F14	-	-	-	-	-	-	-	0	-
F15	-	-	-	-	-	-	-	0	-

Tabela 4.12 - Valores do NAFO para as funções F1 a F15 e  $n = 50$ .

Função	1°	7°	13°	19°	25°	ME	DP	TS	AD
F1	2.765	2.839	2.848	2.857	2.921	2,8500E+03	4,6733E+01	1,00	2,8500E+03
F2	-	-	-	-	-	-	-	0	-
F3	-	-	-	-	-	-	-	0	-
F4	-	-	-	-	-	-	-	0	-
F5	-	-	-	-	-	-	-	0	-
F6	-	-	-	-	-	-	-	0	-
F7	124.499	127.505	129.651	131.791	184.864	1,3445E+05	1,5747E+04	1,00	1,3445E+05
F8	-	-	-	-	-	-	-	0	-
F9	352.410	-	-	-	-	3,5388E+05	1,8328E+03	0,20	1,7694E+06
F10	-	-	-	-	-	-	-	0	-
F11	-	-	-	-	-	-	-	0	-
F12	-	-	-	-	-	-	-	0	-
F13	-	-	-	-	-	-	-	0	-
F14	-	-	-	-	-	-	-	0	-
F15	-	-	-	-	-	-	-	0	-

Com base apenas nas Tabelas 4.10, 4.11 e 4.12, observa-se que, das 15 funções teste analisadas, o método do  $q$ -gradiente resolveu com pelo uma execução bem sucedida 10 funções para  $n = 10$ , 6 funções para  $n = 30$  e 3 funções para  $n = 50$ . Para algumas funções, o método do  $q$ -gradiente obteve taxa de sucesso (TS) máxima. Pode-se dizer que o aumento das dimensões influenciou o desempenho do método do  $q$ -gradiente. No entanto, esse aumento de dimensionalidade influenciou também os demais AEs. Apenas dois AEs (G-CMA-ES e L-CMA-ES) apresentaram resultados para  $n = 50$ .

A eficiência do método do  $q$ -gradiente sobre esse segundo conjunto de funções teste é melhor analisada por meio das Tabelas 4.13 a 4.16, que ilustram uma comparação entre os 11 AEs que participaram do CEC-2005 mais o método do  $q$ -gradiente. Os resultados originais da comparação entre todos os AEs podem ser encontrados na compilação feita por Hansen (2005).

A primeira linha dessas tabelas exhibe, para cada função, a melhor medida de análise de desempenho<sup>4</sup> (AD) entre os 11 AEs. Para cada método e função, são exibidos um valor em decimal seguido de um valor entre parênteses e, em alguns casos, um valor entre colchetes. O valor em decimal corresponde à medida de análise de desempenho relativa, e é dada pela análise de desempenho (AD) de um determinado algoritmo dividida pela melhor medida de análise de desempenho dentre todos os AEs (valores abaixo de cada função na primeira linha da tabela). O valor entre parênteses representa o número de execuções independentes bem sucedidas. Por fim, os valores entre colchetes só aparecem quando um algoritmo não atinge nenhum critério de parada para nenhuma execução independente. Neste caso é feita uma ordenação dos métodos com base no valor mediano do erro da função ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) obtido em  $n \cdot 10^5$  avaliações da função objetivo, em que  $n$  é a dimensão. Logo, para os 11 AEs mais o método do  $q$ -gradiente, [1] representa o algoritmo que obteve o melhor valor mediano para o erro da função e [12] representa o algoritmo que obteve o pior valor mediano para o erro da função dentre os algoritmos que não obtiveram nenhuma execução independente bem sucedida. A coluna “NFR” representa o número de funções que cada algoritmo ou método conseguiu resolver em pelo menos uma das 25 execuções independentes. E a coluna “TSA” representa a taxa de sucesso do algoritmo sobre o conjunto de funções analisadas. O valor de TSA é dado pela soma do número de execuções bem sucedidas (valores entre parênteses variando de 1 a 25) dividido pela soma de todas as execuções independentes para todas as funções.

---

<sup>4</sup>Esses valores são os mesmos exibidos na comparação original (HANSEN, 2005).



Os 11 AEs mais o método do  $q$ -gradiente são ordenados do melhor para o pior de acordo com o maior TSA, seguido do maior NFR e, por fim, pela média da medida de análise de desempenho relativa para o conjunto de funções em questão. As tabelas de comparação exibidas a seguir são muito próximas das tabelas originais encontradas em Hansen (2005). As únicas modificações são a classificação dos métodos (valores entre colchetes) e a inclusão do método do  $q$ -gradiente.

Tabela 4.13 - Taxas de desempenho normalizadas para as funções teste unimodais com  $n = 10$ .

Método	NFR	TSA	F1	F2	F3	F4	F5	F6
			1.000	2.400	6.500	2.900	5.900	7.100
G-CMA-ES	6	100%	1,6(25)	1,0(25)	1,0(25)	1,0(25)	1,0(25)	1,5(25)
EDA	6	97%	10,0(25)	4,6(25)	2,5(23)	4,1(25)	4,2(25)	9,6(22)
DE	6	96%	29,0(25)	19,2(25)	18,5(20)	17,9(25)	6,9(25)	6,6(24)
L-CMA-ES	6	88%	1,7(25)	1,1(25)	1,0(25)	65,5(7)	1,0(25)	1,3(25)
BLX-GL50	5	83%	19,0(25)	17,1(25)	[9]	14,5(25)	4,7(25)	7,3(25)
DMS-L-PSO	5	80%	12,0(25)	5,0(25)	1,8(25)	[11]	18,6(20)	7,7(25)
L-SaDE	5	77%	10,0(25)	4,2(25)	8,0(16)	15,9(24)	[9]	6,9(25)
<b>q-gradiente</b>	<b>4</b>	<b>67%</b>	<b>0,5(25)</b>	<b>3,5(25)</b>	<b>[11]</b>	<b>19,9(25)</b>	<b>[11]</b>	<b>10,0(25)</b>
SPC-PNX	4	67%	6,7(25)	12,9(25)	[10]	10,7(25)	6,8(25)	[11]
CoEVO	4	67%	23,0(25)	11,3(25)	6,8(25)	16,2(25)	[10]	[12]
K-PCX	4	62%	1,0(25)	1,0(25)	[8]	19,7(21)	[12]	1,0(22)
BLX-MA	3	49%	12,0(25)	15,4(25)	[12]	25,9(24)	[8]	[10]

Para as funções unimodais (F1 a F6) e dimensão  $n = 10$  (Tabela 4.13), o método do  $q$ -gradiente apresenta uma taxa de sucesso (TSA) de 67% e ocupa a 8ª posição da competição em um total de 12 algoritmos (11 AEs mais o método do  $q$ -gradiente). Note que a medida de análise de desempenho relativa do método do  $q$ -gradiente para a função F1 é 0,5 e não 1,0, pois a medida de análise de desempenho do método (522, 16) é menor que a melhor medida da análise de desempenho obtida dentre os 11 AEs (1.000).

O aumento das dimensões afetou o desempenho de todos os algoritmos. Note na Tabela 4.14 que, para  $n = 30$ , nenhum algoritmo atingiu 100% para TSA. Sobre esse conjunto de funções, o método do  $q$ -gradiente apresenta uma taxa de sucesso (TSA) de apenas 34% e ocupa a 9ª posição da competição. Para as funções multimodais, o método do  $q$ -gradiente apresenta melhores resultados, como ilustram as Tabelas 4.15 e 4.16.

Tabela 4.14 - Taxas de desempenho normalizadas para as funções teste unimodais com  $n = 30$ .

Método	NFR	TSA	F1	F2	F3	F4	F5	F6
			2.700	12.000	43.000	59.000	66.000	60.000
G-CMA-ES	6	90%	1,7(25)	1,1(25)	1,0(25)	1,0(10)	1,0(25)	1,0(25)
L-CMA-ES	5	83%	1,8(25)	1,2(25)	1,0(25)	[12]	1,1(25)	1,1(25)
EDA	4	67%	55,6(25)	13,3(23)	5,1(25)	3,4(25)	[3]	[11]
DMS-L-PSO	4	63%	1,9(25)	10,8(25)	7,9(25)	[9]	[10]	5,5(24)
BLX-GL50	3	50%	21,5(25)	13,3(25)	[7]	[5]	[7]	3,7(25)
SPC-PNX	4	45%	11,1(25)	26,7(22)	[12]	6.1(19)	[11]	86,7(1)
K-PCX	3	43%	1,0(25)	1,0(25)	[6]	[8]	[8]	1,0(14)
L-SaDE	3	41%	7,4(25)	12,5(24)	[8]	9,2(13)	[5]	[10]
<b>q-gradiente</b>	<b>3</b>	<b>34%</b>	<b>0,7(25)</b>	<b>19,4(25)</b>	<b>[9]</b>	<b>[11]</b>	<b>[4]</b>	<b>3,7(1)</b>
BLX-MA	1	17%	11,9(25)	[11]	[11]	[7]	[9]	[9]
DE	1	17%	51,9(25)	[10]	[10]	[4]	[6]	[7]
CoEVO	2	7%	51,9(3)	70,0(8)	[5]	[10]	[12]	[12]

Tabela 4.15 - Taxas de desempenho normalizadas para as funções multimodais com  $n = 10$ .

Método	NFR	TSA	F7	F9	F10	F11	F12	F15
			4.700	17.000	55.000	190.000	8.200	33.000
G-CMA-ES	5	63%	1,0(25)	4,5(19)	1,2(23)	1,4(6)	4,0(22)	[3]
L-SaDE	4	53%	36,2(6)	1,0(25)	[6]	[9]	3,9(25)	1,0(23)
DMS-L-PSO	4	47%	126(4)	2,1(25)	[3]	[8]	6,6(19)	1,7(22)
<b>q-gradiente</b>	<b>5</b>	<b>41%</b>	<b>15,5(25)</b>	<b>2,3(19)</b>	<b>3,0(3)</b>	<b>[2]</b>	<b>20,5(11)</b>	<b>6,0(3)</b>
K-PCX	3	40%	[10]	2,9(24)	1,0(22)	[11]	1,0(14)	[12]
DE	5	30%	255(2)	10,6(11)	[10]	1,0(12)	8,8(19)	75,8(1)
L-CMA-ES	2	25%	1,2(25)	[12]	[11]	[6]	11,6(12)	[5]
BLX-GL50	3	17%	12,3(9)	10,0(3)	[7]	[5]	12,1(13)	[9]
BLX-MA	2	15%	[11]	5,7(18)	[8]	[10]	[10]	8,5(5)
EDA	3	9%	404(1)	[10]	[5]	2,9(3)	4,3(10)	[10]
SPC-PNX	2	1%	383(1)	[9]	[9]	5,8(1)	[12]	[4]
CoEVO	0	0%	[6]	[11]	[12]	[12]	[12]	[7]

Para as funções multimodais e dimensão  $n = 10$ , os AEs não obtiveram nenhuma execução bem sucedida dentre as 25 execuções independentes para as funções F8, F13 e F14. O método do  $q$ -gradiente também não obteve sucesso para as funções F8 e F14. Porém, para a F13 o método foi o único a obter 1 (de 25) execução bem

Tabela 4.16 - Taxas de desempenho normalizadas para as funções multimodais com  $n = 30$ .

Método	NFR	TSA	F7	F9	F10	F11	F12	F15
			6.100	99.000	450.000	5.000.000	180.000	
K-PCX	4	38%	2,5(10)	3,3(18)	1,0(14)	[8]	1,0(5)	[12]
G-CMA-ES	5	37%	1,0(25)	8,0(9)	5,3(3)	1,0(1)	1,3(8)	[1]
L-SaDE	2	36%	21,3(20)	1,0(25)	[5]	[6]	[4]	[6]
<b>q-gradiente</b>	<b>3</b>	<b>28%</b>	<b>3,2(25)</b>	<b>6,1(8)</b>	<b>[3]</b>	<b>[1]</b>	<b>16,4(2)</b>	<b>[7]</b>
DMS-L-PSO	2	22%	9,8(24)	[7]	[6]	[7]	8,3(4)	[4]
EDA	1	20%	21,3(25)	[11]	[10]	[12]	[8]	[8]
BLX-GL50	1	20%	10,2(25)	[6]	[4]	[5]	[9]	[3]
DE	1	20%	32,8(22)	[8]	[7]	[11]	[6]	[10]
L-CMA-ES	1	20%	1,1(25)	[12]	[12]	[4]	[10]	[2]
SPC-PNX	1	13%	60,7(16)	[9]	[8]	[3]	[11]	[9]
CoEVO	1	9%	93,4(11)	[10]	[11]	[10]	[12]	[11]
BLX-MA	1	7%	[8]	6,7(9)	[9]	[9]	[7]	[5]

sucedida. Como a taxa de sucesso do método do  $q$ -gradiente sobre a F13 é muito pequena (TS=0,04), essa função não foi incluída na Tabela 4.15, pois a classificação do método não seria alterada mesmo refazendo os cálculos para todos os AEs.

Dessa forma, para as funções multimodais F7, F9, F10, F11, F12 e F15 com  $n = 10$ , o método do  $q$ -gradiente apresenta uma taxa de sucesso (TSA) de 41% e ocupa a 4ª posição da competição. Levando em consideração apenas o NFR (número de funções que cada algoritmo ou método consegue resolver em pelo menos uma das 25 execuções independentes) o método do  $q$ -gradiente ocuparia a 1ª posição, pois resolve 6 funções multimodais (5 da Tabela 4.15 mais a função F13). Novamente, o aumento das dimensões diminuiu o desempenho de todos os algoritmos. Para as mesmas funções multimodais e  $n = 30$  (Tabela 4.16) o método do  $q$ -gradiente apresenta uma taxa de sucesso de apenas 28%. No entanto, ocupa a 4ª posição da competição.

Para este segundo conjunto de funções teste, o método do  $q$ -gradiente foi comparado com 11 AEs que participaram da competição do CEC-2005. O balanço sobre as funções unimodais é a 8ª posição na competição para a dimensão  $n = 10$  e 9ª posição para  $n = 30$ . Sobre as funções multimodais, o método do  $q$ -gradiente ocupa a 4ª posição, tanto para  $n = 10$  quanto para  $n = 30$ . O aumento das dimensões afetou os algoritmos de modo geral.

Vale ressaltar que os resultados dos AEs foram obtidos considerando um espaço de busca restrito (restrições laterais), o que não foi o caso do método do  $q$ -gradiente. Em outras palavras, os resultados apresentados para o método do  $q$ -gradiente nesta comparação são conservadores e espera-se que uma versão com restrições melhore a colocação do método na competição.

Por fim, podemos observar que, embora o método do  $q$ -gradiente não apresente um bom desempenho sobre as funções unimodais, ele se mostra competitivo sobre as funções multimodais. Dessa forma, o método comprova a sua capacidade de escapar de mínimos locais e caminhar na direção do mínimo global com eficiência comparada a de AEs que são métodos amplamente reconhecidos e utilizados na resolução de problemas de otimização global.

## 5 APLICAÇÃO À ENGENHARIA ESPACIAL

Neste capítulo, são apresentados um problema de otimização da engenharia aeroespacial e os resultados obtidos da sua resolução pelo método do  $q$ -gradiente. O problema consiste na configuração de uma constelação de satélites para um sistema de posicionamento regional cobrindo, inicialmente, o Brasil e, posteriormente, a América do Sul. O resultado obtido pelo método do  $q$ -gradiente para o caso brasileiro é comparado com o obtido por um algoritmo evolutivo híbrido.

### 5.1 Introdução

Satélites artificiais são equipamentos construídos pelo homem e colocados em torno da Terra para diferentes propósitos como monitoramento, espionagem, telecomunicações, meteorologia, agricultura, sensoriamento remoto, navegação terrestre, entre outros. A órbita ou trajetória de um satélite é determinada, principalmente, pela sua inclinação e seu período de revolução (tempo que o satélite necessita para dar uma volta em torno da Terra) que, por sua vez, está diretamente relacionado a sua altitude.

Satélites colocados em órbita alta de 35.800 a 36.000 quilômetros de altitude possuem período de revolução igual a 23h e 56 minutos, que é igual ao período de rotação da Terra. Neste caso, os satélites são denominados geossíncronos e levam 1 dia para dar a volta em torno da Terra. Se o plano da órbita for o mesmo do equador, o satélite parecerá imóvel a um observador terrestre e será denominado geostacionário. Satélites de posicionamento giram em torno da Terra em órbitas intermediárias de 20.000 quilômetros de altitude com período de revolução de 12 horas (MOREIRA, 2012).

Um Sistema Global de Navegação por Satélite ou *Global Navigation Satellite System* (GNSS) fornece informações de posicionamento geoespacial com cobertura global. Esses sistemas são popularmente conhecidos como GPS, *Global Positioning System*, em função do sistema americano *Navigation Satellite with Timing and Ranging Global Positioning System* ou simplesmente NAVSTAR GPS. Além do americano, existe hoje em funcionamento o russo Glonass, e em fase de desenvolvimento o europeu Galileo, o chinês COMPASS, o japonês QZSS e o indiano IRNSS (GALSKI, 2012).

O sistema GPS é constituído por três componentes principais: o segmento espacial (satélites), o segmento terrestre (monitoramento e controle) e o segmento do usuário (receptores GPS e equipamentos associados). A constelação desse sistema é composta

por 24 satélites que executam órbitas circulares ao redor da Terra e estão distribuídos em 6 planos orbitais (A, B, C, D, E e F) com 4 satélites em cada um como ilustra a Figura 5.1. Esses planos possuem uma inclinação de  $55^\circ$  em relação ao Equador, a uma altura de aproximadamente 20.200 quilômetros, completando uma volta circular ao redor da Terra em um período orbital de 11 horas e 58 minutos (MIGUENS, 2012).

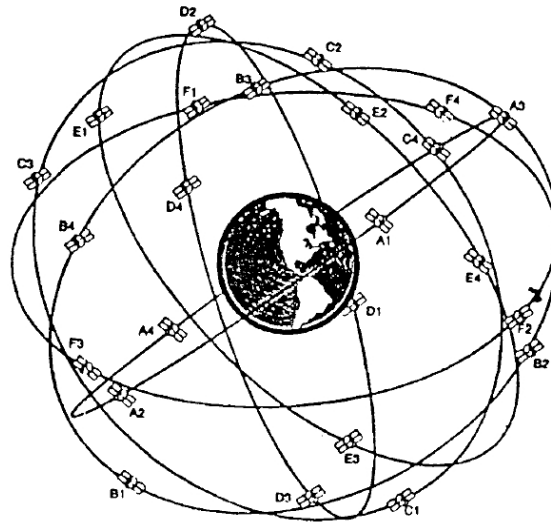


Figura 5.1 - Constelação de satélites GPS.  
Fonte: Extraído de (MIGUENS, 2012).

A constelação de satélites GPS foi colocada em órbita de forma que pelo menos 4 satélites estão sempre visíveis em qualquer ponto da superfície terrestre com uma probabilidade de 95% (MIGUENS, 2012). O sistema GPS permite que receptores móveis civis e militares localizados na terra, no mar ou no ar, determinem suas próprias localizações em termos de longitude, latitude e altitude a partir dos sinais de rádio enviados continuamente pelos satélites.

Os satélites de um GNSS possuem relógios com padrões atômicos e por isso são os pontos de referência do sistema. O cálculo da localização de um receptor é baseado na diferença entre o instante que o sinal de tempo foi emitido pelo satélite e o instante que o receptor recebeu esse sinal. Com esta medida de tempo e conhecendo a velocidade de propagação do sinal, o receptor é capaz de calcular a sua distância ao satélite. A partir de três distâncias conhecidas é possível determinar a localização do receptor, em termos de latitude e longitude, usando um princípio básico de posicionamento conhecido como trilateração. Cada distância é tomada como raio de

uma esfera com centro no satélite e a posição do receptor é definida pela interseção de três esferas (três satélites) com a superfície da Terra. Com quatro medidas de distância (4 satélites) é possível determinar, também, o erro de sincronia entre o relógio do receptor e o relógio de bordo dos satélites (MIGUENS, 2012).

As fontes comuns de erro em GNSS são a falta de sincronia entre os relógios dos satélites e dos receptores, a passagem dos sinais dos satélites pelas camadas da atmosfera (ionosfera e troposfera), os erros de cálculos dos softwares dos receptores, e a disposição geométrica dos satélites. A influência da disposição dos satélites na determinação da posição do receptor é medida pelo princípio geométrico denominado diluição geométrica da precisão (GDOP - *Geometric Dilution of Precision*) (MIGUENS, 2012). A Figura 5.2 ilustra dois tipos de disposição geométrica para 4 satélites, uma boa e uma ruim.

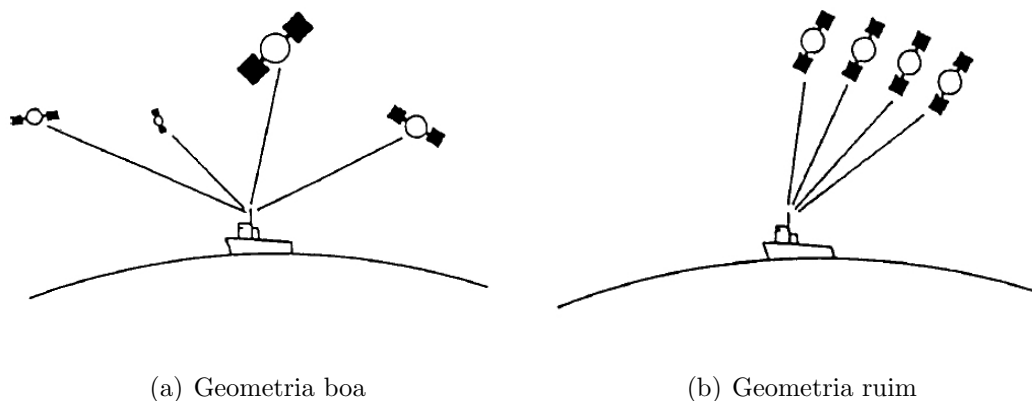


Figura 5.2 - Tipos de geometria que influenciam na precisão de uma posição GPS.  
Fonte: Adaptado de (MIGUENS, 2012).

O GDOP está associado, geometricamente, ao volume do poliedro formado pelos 4 satélites (base) e o receptor (vértice). Quanto maior o volume deste poliedro, melhor (isto é, mais baixo) é o GDOP (veja a Figura 5.2a). Uma situação crítica para o cálculo do GDOP é quando os satélites estão “alinhados”, como ilustra a Figura 5.2b. Neste caso, o volume do poliedro associado será pequeno, mas com valor de GDOP alto e a precisão na localização do receptor será ruim.

Em face a essa informação de que a disposição dos satélites pode afetar a precisão do posicionamento, os receptores podem ser programados de modo a decidir dentre

todos os satélites visíveis aqueles que oferecerão o menor valor de GDOP calculado para uma combinação de 4 satélites. A combinação de 4 satélites que oferecer o menor valor do GDOP oferecerá a solução de navegação com a melhor precisão (LANGLEY, 1999). O valor do GDOP é um número adimensional classificado de acordo com a Tabela 5.1.

Tabela 5.1 - Valores de GDOP.

Valores do GDOP	Classificação
1-2	Excelente
2-5	Boa
5-10	Moderada
10-20	Aceitável
> 20	Ruim

Vale mencionar que, o sistema GPS oferece dois serviços de posicionamento: um preciso e outro padrão. O serviço de posicionamento preciso (PPS - *Precise Positioning Service*) está disponível primeiramente às forças armadas dos EUA e de seus aliados. Os demais usuários possuem acesso ao serviço de posicionamento padrão (SPS - *Standard Positioning Service*) que possui um nível de precisão degradado. Além do serviço ser naturalmente menos preciso, o departamento de defesa dos EUA, por razões de segurança, pode a qualquer momento introduzir erros nos relógios dos satélites diminuindo ainda mais a precisão do GPS (MIGUENS, 2012). Esse procedimento é denominado disponibilidade seletiva (SA - *Selective Availability*) e segundo declaração do governo americano, o seu uso foi suspenso desde o ano 2000 (GPS.GOV, 2012).

Atualmente o Brasil, como inúmeros outros países, não possui um sistema de navegação por satélites próprio e depende da tecnologia americana. Para se livrar da dependência e dessa possível disponibilidade seletiva, outros países estão desenvolvendo sistemas do tipo GPS. Considerando a importância desse serviço, tanto para fins militares quanto civis, é razoável acreditar que o Brasil também terá que desenvolver o seu próprio sistema.

Neste contexto, a aplicação do método do  $q$ -gradiente consiste na determinação da configuração inicial de uma constelação de satélites para um sistema de posicionamento regional (RPS - *Regional Positioning System*) que cubra primordialmente o



território brasileiro com a maior precisão possível no sentido do GDOP. Além do território nacional, também foi considerado um RPS para o território sul-americano. O resultado para o RPS brasileiro obtido pelo método do  $q$ -gradiente é comparado com o obtido pelo GEO+EE apresentado em Galski (2012).

A seguir o problema de otimização é definido.

## 5.2 Problema de otimização

O problema de otimização consiste na determinação da configuração inicial (parâmetros orbitais) de uma constelação composta por 4 satélites geossíncronos através da minimização da média da diluição geométrica da precisão, ou o GDOP médio, para um conjunto de receptores hipotéticos localizados no solo do território alvo. Foram considerados cinco receptores para o RPS brasileiro (R1, R2, R3, R4 e R5) e seis para o RPS sul-americano (R1, R2, R3, R4, R5 e R6), que estão dispostos conforme ilustram as Figuras 5.3 e 5.4.

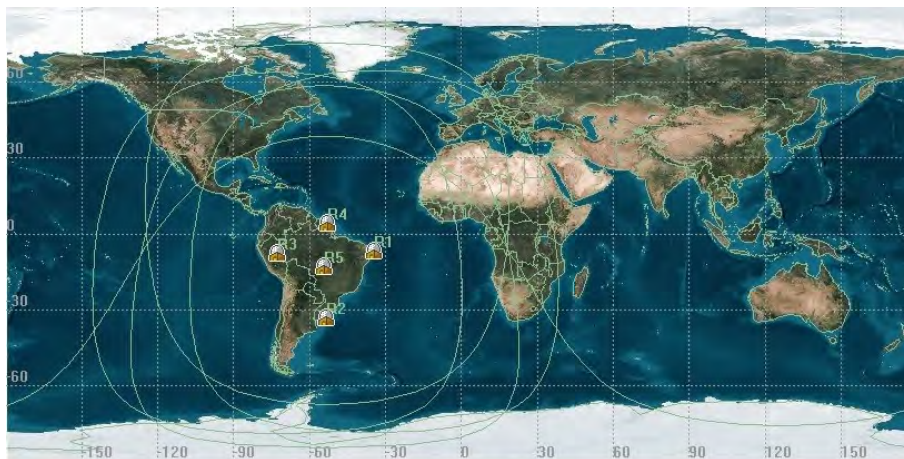


Figura 5.3 - Localização dos 5 receptores sobre o território brasileiro.

As localizações dos receptores foram obtidas de forma empírica, mas com o auxílio do software *Satellite Tool Kit* (STK) (AGI, 2011) de modo que os círculos de visibilidade de todos os receptores cobrissem primordialmente os territórios alvo. A intersecção de todos os círculos de visibilidade representa indiretamente a região do espaço de busca viável, pois somente nesta intersecção todos os receptores possuem visibilidade simultânea a um dado satélite. Note que essa intersecção é menor para o RPS sul-americano em relação ao RPS brasileiro.

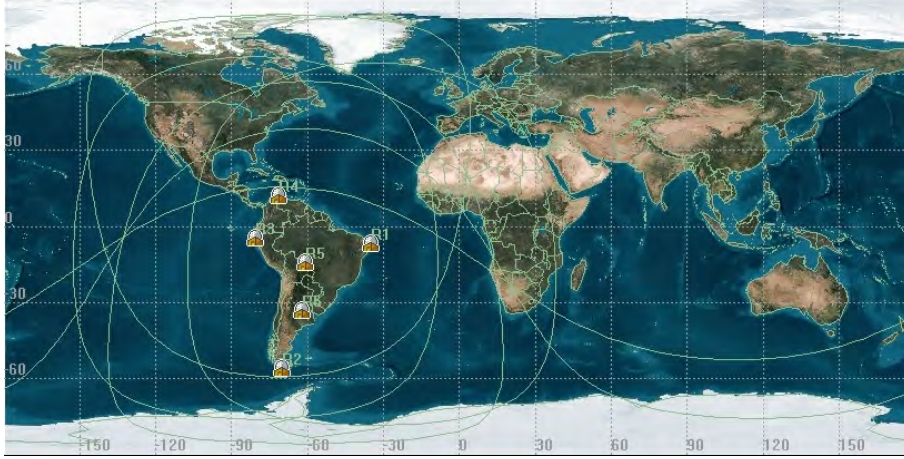


Figura 5.4 - Localização dos 6 receptores sobre o território sul-americano.

A inclinação da órbita em relação ao Equador ( $I_i$ ), a ascensão reta do nó ascendente ( $\Omega_i$ ) e a anomalia média ( $M_i$ ) de cada satélite  $i$  da constelação ( $i = 1, 2, 3, 4$ ) são as variáveis de decisão para o problema de otimização. Os demais elementos keplerianos são constantes para todos os satélites: semi-eixo maior (42.164, 139 km), excentricidade (0,0) e argumento do perigeu (0,0°). O valor definido para o semi-eixo maior corresponde a uma altitude da órbita geossíncrona igual a 35.786 km. Os intervalos de inicialização utilizados para as variáveis de decisão foram  $0^\circ \leq I_i \leq 60^\circ$  para a inclinação,  $0^\circ \leq \Omega_i \leq 360^\circ$  para a ascensão reta e  $0^\circ \leq M_i \leq 360^\circ$  para a anomalia média.

Para avaliar os parâmetros orbitais  $I$ ,  $\Omega$  e  $M$  da constelação candidata, um modelo numérico para a propagação de órbitas de múltiplos satélites é invocado para simular a trajetória de cada satélite da constelação num período de 24 horas, ou 86.400 segundos, com um passo de integração de 30 segundos. A cada passo de integração, o GDOP de cada receptor é calculado de acordo com a geometria dos 4 satélites da constelação e, em seguida, é armazenado. Ao final da simulação da propagação da órbita da constelação candidata, tem-se 2.881 valores de GDOP para cada receptor, sendo um valor de GDOP para o instante inicial e o restante para os demais 2.880 instantes de propagação ( $86.400/30 = 2.880$ ). Matematicamente, o problema é formulado como segue.

$$\text{Minimize} \quad \overline{\text{GDOP}}(\mathbf{x}) = \frac{1}{\text{NRP}} \sum_{i=1}^{\text{NRP}} \left( \frac{1}{\text{NP}} \sum_{k=1}^{\text{NP}} \text{GDOP}_{i,k}(\mathbf{x}) \right) \quad (5.1)$$

em que

$\overline{\text{GDOP}}(\mathbf{x})$	=	GDOP médio considerando todos os receptores em solo e para todos os pontos de propagação de órbita dos satélites
NRP	=	Número de receptores em solo considerados
NP	=	Intervalo de propagação/passo de propagação
$\text{GDOP}_{i,k}(\mathbf{x})$	=	GDOP calculado para o receptor $i$ no instante de propagação $k$
$\mathbf{x}$	=	Variáveis de decisão $[I_{1,\dots,\text{NS}}, \Omega_{1,\dots,\text{NS}}, M_{1,\dots,\text{NS}}]$
$I_j$	=	Inclinação do satélite $j$
$\Omega_j$	=	Ascensão reta do nodo ascendente do satélite $j$
$M_j$	=	Anomalia média do satélite $j$
NS	=	Número de satélites

O cálculo do GDOP para cada receptor em um dado instante de propagação ( $\text{GDOP}_{i,k}(\mathbf{x})$  na Equação 5.1) é formulado a seguir. Inicialmente, considere os vetores unitários dos receptores do satélite  $j$ :  $\left(\frac{(x_j-x)}{R_j}, \frac{(y_j-y)}{R_j}, \frac{(z_j-z)}{R_j}\right)$ , onde  $R_j = \sqrt{(x_j-x)^2 + (y_j-y)^2 + (z_j-z)^2}$ . Tem-se que  $x$ ,  $y$ , e  $z$  são as posições do receptor, e  $x_j$ ,  $y_j$ , e  $z_j$  são as posições do satélite  $j$ . Considere a matriz A, dada por

$$A = \begin{bmatrix} \frac{(x_1-x)}{R_1} & \frac{(y_1-y)}{R_1} & \frac{(z_1-z)}{R_1} & -1 \\ \frac{(x_2-x)}{R_2} & \frac{(y_2-y)}{R_2} & \frac{(z_2-z)}{R_2} & -1 \\ \frac{(x_3-x)}{R_3} & \frac{(y_3-y)}{R_3} & \frac{(z_3-z)}{R_3} & -1 \\ \frac{(x_4-x)}{R_4} & \frac{(y_4-y)}{R_4} & \frac{(z_4-z)}{R_4} & -1 \end{bmatrix}.$$

Os primeiros três elementos de cada linha da matriz A representam os componentes do vetor unitário do receptor para um determinado satélite. Se os elementos na quarta coluna são iguais a  $c$ , o qual representa a velocidade da luz, então o fator  $d_t$  é sempre 1. Se os elementos na quarta coluna são iguais a  $-1$ , então o fator  $d_t$  é calculado de forma apropriada.

Considere a matriz Q como sendo  $Q = (A^T A)^{-1}$ . Então, os elementos de Q são dados por

$$Q = \begin{bmatrix} d_x^2 & d_{xy}^2 & d_{xz}^2 & d_{xt}^2 \\ d_{xy}^2 & d_y^2 & d_{yz}^2 & d_{yt}^2 \\ d_{xz}^2 & d_{yz}^2 & d_z^2 & d_{zt}^2 \\ d_{xt}^2 & d_{yt}^2 & d_{zt}^2 & d_t^2 \end{bmatrix}.$$

Por fim, o PDOP, o TDOP e o GDOP são dados por

$$\begin{aligned}PDOP &= \sqrt{d_x^2 + d_y^2 + d_z^2}, \\TDOP &= \sqrt{d_t^2}, \\GDOP &= \sqrt{PDOP^2 + TDOP^2}.\end{aligned}$$

Quando menos de 4 satélites são visíveis para um determinado receptor, o valor do GDOP deste receptor é arbitrariamente imposto como sendo igual a 100.000, e essa imposição implica em valores altos para o GDOP médio de tais casos.

O modelo de simulação da órbita de múltiplos satélites utilizado neste trabalho é o mesmo utilizado em [Galski \(2012\)](#) e se baseia nas rotinas de dinâmica orbital para a órbita de apenas um satélite desenvolvidas pelo grupo de dinâmica orbital do INPE ([KUGA; ORLANDO, 1988b](#); [KUGA; ORLANDO, 1988a](#); [KUGA et al., 1988](#)). A seguir, são especificados os parâmetros utilizados pelo método do  $q$ -gradiente na resolução deste problema de otimização.

### 5.3 Metodologia

A determinação da configuração inicial de uma constelação de satélites, a partir de seus parâmetros orbitais ( $I$ ,  $\Omega$  e  $M$ ), para um sistema de posicionamento regional (RPS) de modo a obter o menor valor de GDOP médio possível, foi obtida a partir da aplicação do método do  $q$ -gradiente (Algoritmo 7) na minimização do funcional descrito no problema de otimização 5.1.

O método do  $q$ -gradiente e as rotinas de propagação da órbita da constelação foram executadas em uma máquina com processador Intel®Core™2 Quad Q9550 com 2GB RAM e sistema operacional Windows XP Professional 32 bits. O passo de integração das órbitas foi de 30 segundos, e a época considerada para todas as simulações foi o dia 01 de Novembro de 2010 às 0h00m00s.

Resultados para um RPS brasileiro, composto por 4 satélites geossíncronos e cinco receptores em solo, foram obtidos por [Galski \(2012\)](#) aplicando as mesmas rotinas de propagação da órbita da constelação, mas com o cálculo do GDOP médio a partir do algoritmo evolucionário híbrido GEO+EE ([GALSKI, 2006](#)). Seguindo o mesmo critério de parada definido em [Galski \(2012\)](#), o método do  $q$ -gradiente utilizou, tanto para o RPS brasileiro quanto para o RPS sul-americano, um número total de avaliações da função objetivo, ou cálculos do GDOP médio, igual a 180.000 divididos em

10 execuções independentes. Ou seja, para cada execução independente, o método do  $q$ -gradiente dispunha de um número máximo de avaliações da função objetivo igual a 18.000.

Os parâmetros de ajuste do método do  $q$ -gradiente (desvio-padrão inicial  $\sigma^0$ , passo inicial  $\alpha^0$  e taxa de redução  $\beta$ ) para o RPS brasileiro e RPS sul-americano foram obtidos por meio de rodadas preliminares, conforme procedimento descrito na Seção 4.1. Foram considerados 6,25%, 12,5%, 25%, 37,5%, 50%, 62,5%, 75% e 87,5% de  $\Delta \mathbf{x}$ , tanto para o  $\sigma^0$  quanto para o  $\alpha^0$ , em que o  $\Delta \mathbf{x}$  é a distância euclidiana entre os valores mínimos e máximos de todas as variáveis de decisão. Os valores de  $\beta$  considerados foram 0,80, 0,90, 0,99, 0,995 e 0,999. Para cada combinação de parâmetros, foram realizadas cinco execuções distintas do Algoritmo 7, cada uma com um critério de parada igual a 1.000 avaliações da função objetivo ou cálculos do GDOP (em vez de 18.000). Após essa investigação preliminar, escolheu-se os melhores valores dos parâmetros de ajuste, conforme ilustra a Tabela 5.2.

Tabela 5.2 - Parâmetros de ajuste utilizados pelo RPS brasileiro e sul-americano.

RPS	Parâmetros		
	$\sigma^0$	$\alpha^0$	$\beta$
Brasileiro	85,44	384,48	0,995
Sul-americano	128,16	768,96	0,995

#### 5.4 Análises e resultados

Os resultados obtidos pelo método do  $q$ -gradiente, com os parâmetros de ajuste da Tabela 5.2, refletem a melhor solução (ou menor GDOP médio) dentre as 10 execuções independentes. A Tabela 5.3 ilustra a solução encontrada pelo método do  $q$ -gradiente para os valores de melhor GDOP médio e das variáveis de decisão, ou parâmetros orbitais.

Para o RPS brasileiro, o valor de GDOP médio obtido pelo método do  $q$ -gradiente (8,89) é próximo do obtido pelo GEO+EE (7,71) sobre o mesmo problema (GALSKI, 2012). Ambos estão na faixa de valores considerada moderada (entre 5 e 10). A Figura 5.5 ilustra os valores de GDOP calculados para cada receptor (R1, R2, R3, R4 e R5) em cada passo de integração durante 24 horas.

Tabela 5.3 - Parâmetros orbitais (em graus) e melhor GDOP médio (adimensional) para o RPS brasileiro e RPS sul-americano obtidos pelo método do  $q$ -gradiente.

Parâmetros Orbitais		RPS brasileiro	RPS sul-americano
Inclinação	$I_1$	0,94	14,40
	$I_2$	35,02	13,79
	$I_3$	26,13	1,30
	$I_4$	2,97	0,02
Ascensão reta do nó ascendente	$\Omega_1$	338,60	342,53
	$\Omega_2$	18,59	152,28
	$\Omega_3$	188,63	124,74
	$\Omega_3$	126,06	6,69
Anomalia média	$M_1$	50,10	0,46
	$M_2$	318,78	189,66
	$M_3$	154,07	172,81
	$M_4$	170,29	9,45
GDOP médio		8,89	16,02

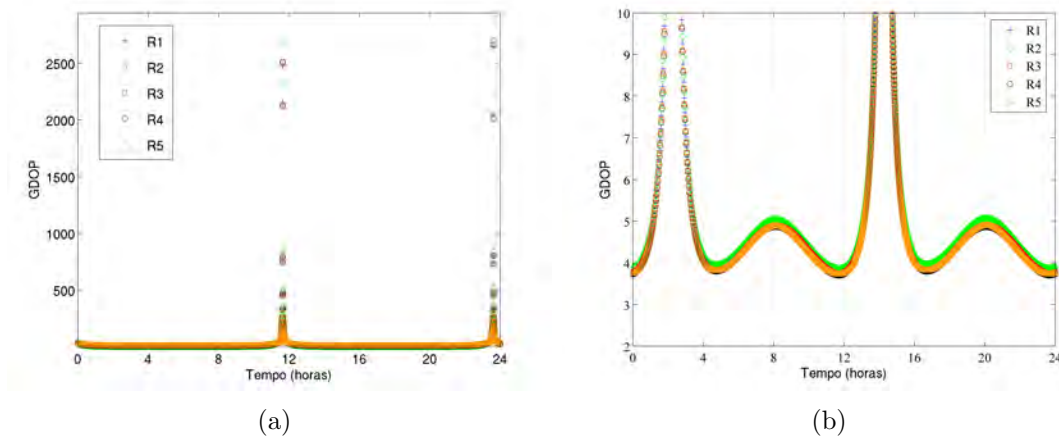


Figura 5.5 - GDOP para o RPS brasileiro com 4 satélites e 5 receptores.

Os comportamentos dos GDOPs dos cinco receptores para o RPS brasileiro são muito próximos. O tempo, em porcentagem, que os valores de GDOP estão abaixo de 10 para os receptores R1, R2, R3, R4 e R5 são, respectivamente, 91,95%, 91,17%,



92,02%, 92,07% e 92,02%. Em 24 horas, os valores de GDOP para os cinco receptores são maiores que 10 durante (aproximadamente) 1 hora e 56 minutos. Note que, todas as curvas apresentam dois grandes picos simultaneamente, sendo o primeiro aproximadamente às 2h18 e o segundo às 14h18. Note, também, que a separação entre os picos é de 12 horas.

Com o auxílio do STK, é possível visualizar a órbita da constelação candidata a partir dos parâmetros orbitais encontrados pelo método do  $q$ -gradiente, e entender o fenômeno por detrás dos picos. As Figuras 5.6 e 5.7 ilustram, respectivamente, os instantes 2h18 e 14h18.



Figura 5.6 - Primeiro pico das curvas GDOP para o RPS brasileiro às 2h18.

Durante os picos das curvas do GDOP, as órbitas dos 4 satélites quase se alinham em um plano. Para estes casos a geometria é considerada ruim e os valores de GDOP são altos. Os cinco receptores estão representados pelas siglas R1, R2, R3, R4 e R5 e os 4 satélites por S1, S2, S3 e S4. As figuras ilustram também as trajetórias dos satélites (formato de 8) e o círculo de visibilidade do R2, localizado no extremo sul do Brasil, que limita indiretamente a intersecção de visibilidade de todos os receptores do RPS brasileiro de forma mais restritiva.

Para o RPS sul-americano, o valor de GDOP médio obtido (16,02) pertence a faixa de valores aceitáveis, mas com um baixo nível de confiabilidade nos resultados. A Figura 5.8 ilustra os valores de GDOP calculados para cada receptor (R1, R2, R3,



Figura 5.7 - Segundo pico das curvas GDOP para o RPS brasileiro às 14h18.

R4, R5 e R6) ao longo da propagação das órbitas dos 4 satélites.

As análises aqui são similares às realizadas para o RPS brasileiro. O tempo, em porcentagem, que os valores de GDOP estão abaixo de 10 para os receptores R1, R2, R3, R4, R5 e R6 são, respectivamente, 82,19%, 80,14%, 82,23%, 82,33%, 82,44% e 81,71%. Em 24 horas, os valores de GDOP para os seis receptores são maiores que 10 durante (aproximadamente) 4 horas e 20 minutos. Novamente, as curvas de GDOP apresentam dois grandes picos separados por uma distância temporal de 12 horas, sendo o primeiro aproximadamente às 11h30 e o segundo às 23h30. As Figuras 5.9 e 5.10 ilustram, respectivamente, os instantes 11h30 e 23h30 em que os satélites S1, S2, S3 e S4 estão alinhados sobre o plano.

Novamente, a curva de visibilidade para o receptor R2 localizado no extremo sul do continente americano delimita a intersecção de visibilidade comum a todos os receptores de forma mais restritiva. Visualmente os satélites S1 e S2 se sobrepõem, porém a Figura 5.11 ilustra que os dois apresentam órbitas próximas.

Os resultados obtidos para o RPS brasileiro são próximos dos obtidos por Galski (2012). Os resultados para o RPS sul-americano são piores quando comparados com os obtidos para o RPS brasileiro. A constelação proposta utiliza o número mínimo de satélites (quatro) necessário para o cálculo do GDOP e, portanto, os valores baixos de GDOP médio serão obtidos apenas para as órbitas que fiquem compreendidas dentro da região do espaço de busca, que corresponde a intersecção de todos os



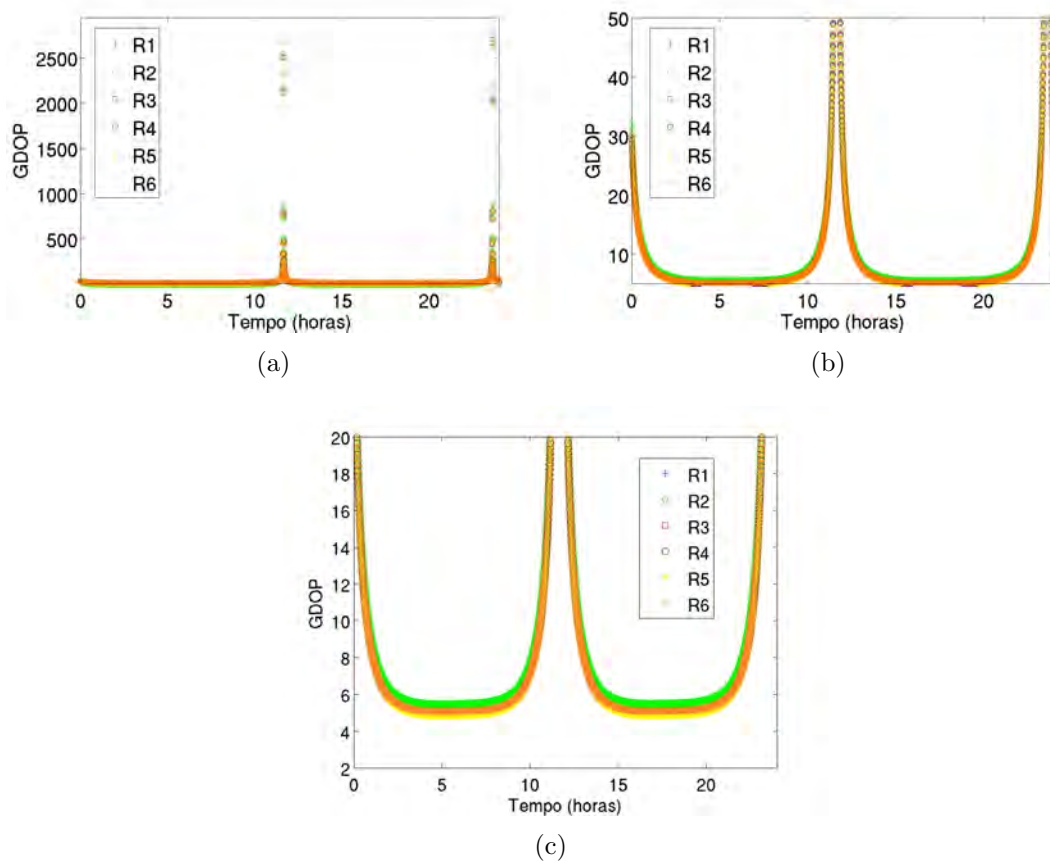


Figura 5.8 - GDOP para o RPS sul-americano com 4 satélites e 6 receptores.

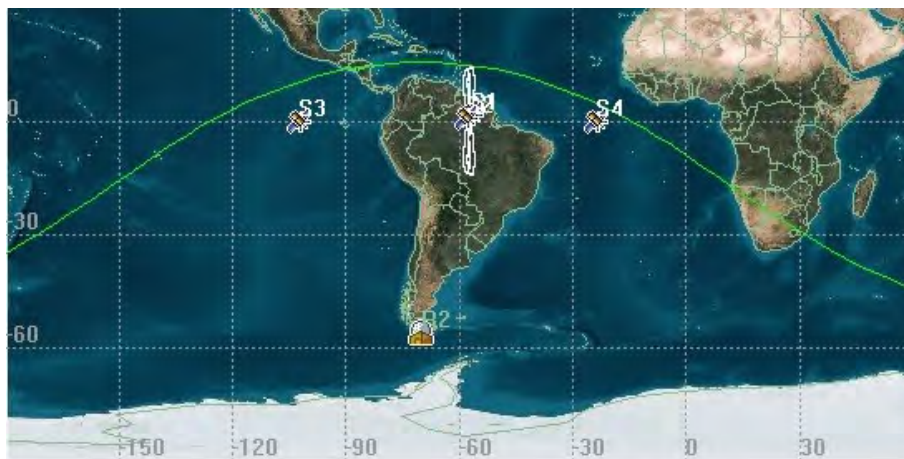


Figura 5.9 - Primeiro pico das curvas GDOP para o RPS sul-americano às 11h30.

círculos de visibilidade dos receptores. No RPS sul-americano essa intersecção, ou seja, essa visibilidade simultânea, é menor em relação ao RPS brasileiro, o que torna o problema de otimização no caso sul-americano mais difícil.



Figura 5.10 - Segundo pico das curvas GDOP para o RPS sul-americando às 23h30.



Figura 5.11 - Zoom da Figura 5.10. Note que os satélites S1 e S2 possuem órbitas próximas.

Embora preliminares, os resultados obtidos mostram a viabilidade de aplicação do método do  $q$ -gradiente na resolução de um problema real. Novas configurações com a adição de mais satélites deverão ser testadas de modo a reduzir ou eliminar os picos no cálculo do GDOP médio.

## 6 CONCLUSÕES

Problemas em diversas áreas podem ser modelados por meio de conceitos e ferramentas de otimização. Dentre eles, os problemas de otimização global constituem a maioria dos problemas reais e são de difícil resolução, pois podem assumir características de não-convexidade, não-linearidade e, em alguns casos, as funções objetivo são determinadas implicitamente e avaliadas somente por meio de simulação numérica. Esses problemas também são caracterizados pela existência de inúmeros mínimos locais e, muitas vezes, a quantidade de mínimos nem é conhecida. Existe uma dificuldade em garantir que a solução encontrada para esses problemas é de fato a melhor solução possível pois, ao contrário das soluções locais, as soluções globais não possuem condições de otimalidade bem definidas. Uma vez que os problemas de otimização global assumem desde características bem conhecidas até outras pouco exploradas, essa área requer o desenvolvimento de paradigmas de modelagem e de resolução que sejam novos, mas ao mesmo tempo levem em consideração os conceitos e as técnicas clássicas de otimização.

Este trabalho estendeu a aplicabilidade do  $q$ -cálculo na área de otimização por meio da definição e do uso do vetor  $q$ -gradiente da função objetivo em métodos clássicos. Foi criada uma generalização do método da máxima descida, denominado método do  $q$ -gradiente, em que a direção de busca dada pelo anti-gradiente é substituída pela direção do anti- $q$ -gradiente. O método do  $q$ -gradiente retoma a sua versão clássica sempre que o parâmetro  $\mathbf{q}$  é igual a  $\mathbf{1}$ . A principal ideia por trás do método é a transição entre busca global, no início, e busca local, no final, do procedimento iterativo, com a presença de mecanismos que permitem que o método escape de mínimos locais e caminhe cada vez mais na direção do mínimo global.

Dentre as estratégias para a obtenção do parâmetro  $\mathbf{q}^k$ , destacou-se a estratégia de geração de fatores  $\mathbf{q}^k \mathbf{x}^k$  segundo uma distribuição gaussiana centrada no ponto atual da busca  $\mathbf{x}^k$  na iteração  $k$  e com desvio-padrão inicial positivo  $\sigma^0 > 0$ , mas que decresce ao longo das iterações de acordo com a regra  $\sigma^{k+1} = \beta \cdot \sigma^k$ , onde  $0 < \beta < 1$  é o fator de redução. A influência do desvio-padrão no método do  $q$ -gradiente é similar a da temperatura no Recozimento Simulado, ou seja, valores altos para  $\sigma^k$  no início do procedimento iterativo implicam em busca global com maior probabilidade, e valores baixos de  $\sigma^k$  no final do procedimento iterativo implicam em busca local com maior probabilidade. É o decréscimo do  $\sigma^0$  ao longo das iterações que permite a transição suave entre busca global e busca local do método do  $q$ -gradiente.

Como a direção de busca dada pela direção do anti- $q$ -gradiente nem sempre é de

máxima descida e pode ser, inclusive, uma direção de subida, estratégias de busca linear que requerem decréscimo no valor da função objetivo não podem ser aplicadas. A estratégia utilizada é a de um tamanho de passo inicial positivo  $\alpha^0 > 0$ , que decresce ao longo das iterações. Por simplicidade, considerou-se o mesmo fator de redução  $\beta$  utilizado no decréscimo do desvio-padrão  $\sigma^k$ .

Portanto, a direção de busca dada pelo anti- $q$ -gradiente da função objetivo, juntamente com as estratégias para obtenção do parâmetro  $\mathbf{q}$  e do tamanho do passo deram origem ao algoritmo do método do  $q$ -gradiente destinado aos problemas de otimização contínuos, irrestritos, não-lineares, não-convexos, unimodais ou multimodais, isto é, problemas de otimização global.

Além da comparação com algoritmos determinísticos, o algoritmo para o método do  $q$ -gradiente foi extensivamente comparado com Algoritmos Evolutivos (AEs) que possuem diversas técnicas de aprimoramento incorporadas e são muito utilizados na resolução de problemas de otimização global. Os resultados mostraram que o método do  $q$ -gradiente é competitivo em relação aos AEs, sobretudo para as funções teste multimodais, comprovando a capacidade do método de escapar de mínimos locais. O método do  $q$ -gradiente também foi aplicado na resolução de um problema da engenharia aeroespacial e os resultados mostraram a viabilidade do seu uso em aplicações práticas. No entanto, o método não se mostrou tão competitivo na minimização de problemas unimodais. Na comparação com o método da máxima descida, por exemplo, o desempenho dos dois métodos é apenas similar. Este resultado aponta para a necessidade de desenvolvimento de novas estratégias de obtenção do tamanho do passo na direção de anti- $q$ -gradiente.

O uso do  $q$ -gradiente em métodos de otimização mostrou que uma modificação simples no método da máxima descida – o método que possui uma das piores taxas de convergência dentre os métodos clássicos – permitiu que este método local adquirisse características de método global e se tornasse competitivo com algoritmos evolutivos junto a problemas de otimização global. Trabalhos futuros incluem o aprimoramentos desta  $q$ -versão, como a introdução restrições e de características de *reannealing* na metodologia de cálculo do parâmetro  $\mathbf{q}$ . O desenvolvimento de novas  $q$ -versões de métodos clássicos mais eficientes, como o método dos gradientes conjugados, o método de Newton e os métodos quase-Newton, surgem naturalmente por meio da substituição do gradiente clássico pelo  $q$ -gradiente. Essas novas direções de busca baseadas em  $q$ -gradiente precisarão também de novas estratégias para a obtenção do tamanho do passo, como a de um passo adaptativo que apresente comportamentos

distintos na direção de anti-gradiente e na direção de anti- $q$ -gradiente.



## REFERÊNCIAS BIBLIOGRÁFICAS

ABE, S. A note on the  $q$ -deformation-theoretic aspect of the generalized entropies in nonextensive physics. **Physics Letters A**, v. 224, n. 6, p. 326–330, 1997. 29

ADJIMAN, C. S.; FLOUDAS, C. A.  $\alpha$ bb algorithm. In: FLOUDAS, C. A.; PARDALOS, P. M. (Ed.). **Encyclopedia of optimization**. 2. ed. New York: Springer, 2009. p. 61–73. 23

AGI. **Satellite Tool Kit**. Analytical Graphics, Inc., 2011. Disponível em: <<http://www.agi.com/product-explorer/>>. Acesso em: 27 fev. 2012. 95

AUGER, A.; HANSEN, N. Performance evaluation of an advanced local search evolutionary algorithm. In: CONGRESS ON EVOLUTIONARY COMPUTATION, 2005. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. p. 1777–1784. Special Session on Real-Parameter Optimization. 80

\_\_\_\_\_. A restart cma evolution strategy with increasing population size. In: CONGRESS ON EVOLUTIONARY COMPUTATION. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. p. 1769–1776. Special Session on Real-Parameter Optimization. 80

BALLESTER, P. J.; CARTER, J. N. An effective real-parameter genetic algorithm with parent centric normal crossover for multimodal optimisation. In: GENETIC AND EVOLUTIONARY COMPUTATION CONFERENCE. **Proceedings...** Seattle, WA: Springer-Verlag, 2004. p. 901–913. Lecture Notes in Computer Science. 4, 61, 63, 65, 71, 72, 73, 74

BALLESTER, P. J.; STEPHENSON, J.; CARTER, J. N.; GALLAGHER, K. Real-parameter optimization performance study on the cec-2005 benchmark with spc-pnx. In: CONGRESS ON EVOLUTIONARY COMPUTATION. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. Special Session on Real-Parameter Optimization. 80

BAZARAA, M. S.; SHETTY, C. M. **Nonlinear programming**: Theory and algorithms. New York-Chichester-Brisbane-Toronto-Singapore: John Wiley & Sons, 1979. 10, 18

BECCENERI, J. C. Computação e matemática aplicadas às ciências e tecnologias espaciais. In: \_\_\_\_\_. [S.l.]: LAC-INPE, 2008. cap. Meta-heurísticas e otimização combinatória: aplicações em problemas ambientais, p. 65–81. 2, 11, 24

- BOETTCHER, S.; PERCUS, A. Optimization with extremal dynamics. **Phys. Rev. Lett.**, v. 86, p. 5211–5214, 2001. 73
- BORGES, E. P. **Manifestações dinâmicas e termodinâmicas de sistemas não-extensivos**. Tese (Doutorado em Ciências Físicas) — Centro Brasileiro de Pesquisas Físicas - CBPF, Rio de Janeiro, 2004. 29
- CHAUNDY, T. W. Frank hilton jackson (obituary). **Journal of the London Mathematical Society**, v. 37, p. 126–128, 1962. 3, 29
- CZYZYK, J.; MESNIER, M.; MORE, J. **The NEOS Server**. IEEE Journal on Computational Science and Engineering, 1998. 68-75 p. Disponível em: <<http://www.neos-server.org/neos/index.html>>. Acesso em: 10 jul. 2012. 4, 61, 78
- DEB, K. **Multi-objective optimization using evolutionary algorithms**. 1. ed. Chichester: John Wiley and Sons, 2001. 11
- DEB, K.; ANAND, A.; JOSHI, D. A computationally efficient evolutionary algorithm for real-parameter optimization. **Evolutionary Computation**, v. 10, p. 345–369, 2002. 4, 61, 71, 73, 74
- ERNST, T. **The history of q-calculus and a new method**. Tese (Licentiate Thesis) — U.U.D.M. Uppsala University Department of Mathematics, Sweden, 2000. Disponível em: <<http://www.math.uu.se/research/pub/Ernst4.pdf>>. Acesso em: 07 mai. 2008. 3, 29
- ERNST, T. A method for q-calculus. **Journal of Nonlinear Mathematical Physics**, v. 10, n. 4, p. 487–525, 2003. 29
- FLETCHER, R.; REEVES, C. M. Function minimization by conjugate gradients. **Computer Journal**, v. 7, p. 149–154, 1964. 20
- FLOUDAS, C. A. Deterministic global optimization and its applications. In: PARDALOS, P. M.; RESENDE, M. G. C. (Ed.). **Handbook of applied optimization**. New York: Oxford University Press, 2002. p. 311–336. 23
- GALSKI, R. L. **Desenvolvimento de versões aprimoradas, híbridas, paralela e multiobjetivo do método da otimização extrema generalizada e sua aplicação no projeto de sistemas espaciais**. 282 p. (INPE-14795-TDI/1238). Tese (Doutorado em Computação Aplicada) — INPE, São José dos Campos, 2006. 73, 74, 98



GALSKI, R. L. Preliminary design of satellite constellations for a brazilian regional positioning system by means of an evolutionary algorithm. **Journal of Aerospace Engineering, Sciences and Applications**, IV, n. 1, p. 52–59, 2012. ISSN 2236-577X. 91, 95, 98, 99, 102

GAMS DEVELOPMENT CORPORATION. 2001. Disponível em: <<http://www.gams.com/default.htm>>. Acesso em: 10 julho 2012. 77

GARCIA-MARTINEZ, C.; LOZANO, M. Hybrid real-coded genetic algorithms with female and male differentiation. In: CONGRESS ON EVOLUTIONARY COMPUTATION. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. p. 896–903. Special Session on Real-Parameter Optimization. 80

GOLDBERG, D. E. **Genetic algorithms in search, optimization, and machine learning**. New York: Addison-Wesley Publishing, 1989. 26

GOLDEN, B. L.; WASIL, E. A. Metaheuristics: introduction. In: PARDALOS, P. M.; RESENDE, M. G. C. (Ed.). **Handbook of applied optimization**. New York: Oxford University Press, 2002. p. 123–138. 24

GOUNARIS, C. E.; FLOUDAS, C. A. Global optimization: tight convex underestimators. In: FLOUDAS, C. A.; PARDALOS, P. M. (Ed.). **Encyclopedia of optimization**. 2. ed. New York: Springer, 2009. p. 1411–1418. 23

GPS.GOV. **Selective availability**. National Coordination Office for Space-Based Positioning, Navigation and Timing, 2012. Disponível em: <<http://www.gps.gov/systems/gps/modernization/sa/>>. Acesso em: 10 julho 2012. 94

HANSEN, N. Compilation of results on the 2005 cec benchmark function set: Technical report. Institute of computational science, ETH Zurich, Switzerland, 2005. Disponível em: <<http://www.ntu.edu.sg/home/epnsugan/>>. Acesso em: 14 jul. 2012. 86, 87

HOLLAND, J. H. **Adaptation in natural and artificial systems**. Cambridge: MIT Press, 1992. 26

IZMAILOV, A.; SOLODOV, M. **Otimização: Condições de otimalidade, elementos de análise convexa e de dualidade**. Rio de Janeiro: IMPA, 2005. 6, 7

\_\_\_\_\_. **Otimização: Métodos computacionais**. Rio de Janeiro: IMPA, 2007. 3, 9, 10, 11, 12, 13, 15, 19, 22, 35, 53

JACKSON, F. H. A generalization of the functions  $\gamma(n)$  and  $x^n$ . **Proc. Roy Soc. London**, p. 64–72, 1904. 3, 29

\_\_\_\_\_. On q-functions and a certain difference operator. **Trans. Roy Soc. Edin.**, v. 46, p. 253–281, 1908. 3, 29

\_\_\_\_\_. On q-definite integrals. **Quart. J. Pure and Appl. Math.**, v. 41, p. 193–203, 1910. 3, 29

\_\_\_\_\_. q-difference equations. *American Journal of Mathematics*, v. 32, p. 307–314, 1910. 3, 29

JOHAL, R. S. q calculus and entropy in nonextensive statistical physics. **Phys. Rev. E**, v. 58, p. 4147–4151, 1998. 29

KAC, V.; CHEUNG, P. **Quantum calculus**. New York: Springer-Verlag, 2002. 3, 29, 30, 31

KANPUR GENETIC ALGORITHMS LABORATORY. 2005. Disponível em: <<http://www.iitk.ac.in/kangal/index.shtml>>. Acesso em: 10 julho 2012. 73

KIRKPATRICK, S.; JR., C. D. G.; VECCHI, M. P. Optimizing by simulated annealing. **Science**, v. 220, n. 4598, p. 671–680, 1983. 24

KUGA, H. K.; ORLANDO, V. **Structure of software for orbit determination and propagation**: Technical report ccs-swr-0024. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 1988. 98

\_\_\_\_\_. **Structure of software for prediction and analysis of orbit and attitude**: Technical report ccs-swr-0023. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 1988. 98

KUGA, H. K.; RAO, K. R.; ORLANDO, V. **Mathematical theory of the software for orbit determination and propagation**: Technical report ccs-swr-0028. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 1988. 98

LACERDA, E. G. M.; CARVALHO, A. C. P. L. F. de. Introdução aos algoritmos genéticos. In: XIX CONGRESSO NACIONAL DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. **Anais...** Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro, 1999. v. 2, p. 51–126. 27

- LANGLEY, R. B. **Dilution of Precision**. GPS World, 1999. Disponível em: <<http://gauss.gge.unb.ca/papers.pdf/gpsworld.may99.pdf>>. 94
- LAVAGNOA, A.; SWAMY, P. N. Thermostatistics of a q-deformed boson gas. **Phys. Rev. E**, v. 61, n. 2, p. 1218–1226, 2000. 29
- LIANG, J. J.; SUGANTHAN, P. N. Dynamic multi-swarm particle swarm optimizer. In: CONGRESS ON EVOLUTIONARY COMPUTATION. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. p. 124–129. Special Session on Real-Parameter Optimization. 80
- LIMPERT, E.; STAHEL, W. A.; ABBT, M. Log-normal distributions across the sciences: keys and clues. **BioScience**, v. 51, p. 341–352, 2001. 40
- LINDO SYSTEMS INC. **LINDO/LINDOGlobal**. GAMS Development Corporation, Washington D.C., 2001. Disponível em: <<http://www.gams.com/dd/docs/solvers/lindoglobal.pdf>>. Acesso em: 10 julho 2012. 78
- LUENBERGER, D. G. **Linear and nonlinear programming**. 2. ed. Massachusetts: Addison-Wesley, 1984. 1, 8, 9, 14, 16, 18, 19, 21
- METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H. Equation of state calculations by fast computing machines. **The Journal of Chemical Physics**, v. 21, n. 6, p. 1087–1092, 1953. 24
- MIGUENS, A. P. **Manual de Navegação**. MARINHA DO BRASIL, 2012. Disponível em: <<https://www.mar.mil.br/dhn/bhmn/download/cap-37.pdf>>. Acesso em: 10 julho 2012. 92, 93, 94
- MOLINA, D.; HERRERA, F.; LOZANO, M. Adaptive local search parameters for real-coded memetic algorithms. In: CONGRESS ON EVOLUTIONARY COMPUTATION. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. p. 888–895. Special Session on Real-Parameter Optimization. 80
- MOREIRA, M. A. **Satélites: conceito, classificação e objetivos**. Instituto Federal do Espírito Santo, 2012. Disponível em: <[ftp://ftp.cefetes.br/Cursos/Geomatica/Adelson/Sensoriamento\\_Remoto/Topicos\\_sobre\\_satelites.pdf](ftp://ftp.cefetes.br/Cursos/Geomatica/Adelson/Sensoriamento_Remoto/Topicos_sobre_satelites.pdf)>. Acesso em: 10 julho 2012. 91
- NOCEDAL, J.; WRIGHT, S. J. **Numerical optimization**. 2. ed. New York: Springer-Verlag New York, Inc, 2006. 2, 17, 35

PARDALOS, P. M.; MAVRIDOU, T. D. Simulated annealing sa. In: FLOUDAS, C. A.; PARDALOS, P. M. (Ed.). **Encyclopedia of optimization**. 2. ed. New York: Springer, 2009. p. 3591–3593. 25

PILLO, G. D.; PALAGI, L. Nonlinear programming: Introduction. In: PARDALOS, P. M.; RESENDE, M. G. C. (Ed.). **Handbook of applied optimization**. New York: Oxford University Press, 2002. p. 263–268. 5, 9, 13

\_\_\_\_\_. Nonlinear programming: Unconstrained nonlinear programming. In: PARDALOS, P. M.; RESENDE, M. G. C. (Ed.). **Handbook of applied optimization**. New York: Oxford University Press, 2002. p. 268–284. 16, 17, 21, 22

PINTER, J. D. Global optimization: Software, teste problems, and applications. In: PARDALOS, P. M.; ROMEIJN, H. E. (Ed.). **Handbook of global optimization**. Dordrecht: Kluwer Academic Publishers, 2002. p. 515–569. 13

POSIK, P. Real-parameter optimization using the mutation step co-evolution. In: CONGRESS ON EVOLUTIONARY COMPUTATION. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. p. 872–879. Special Session on Real-Parameter Optimization. 80

QIN, A. K.; SUGANTHAN, P. N. Self-adaptive differential evolution algorithm for numerical optimization. In: CONGRESS ON EVOLUTIONARY COMPUTATION. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. Special Session on Real-Parameter Optimization. 80

ROENTHAL, R. E. **A User's Guide**. [S.l.]: GAMS Development Corporation, Washington D.C., 2012. 77

RONKKONEN, J.; KUKKONEN, S.; PRICE, K. V. Real-parameter optimization with differential evolution. In: CONGRESS ON EVOLUTIONARY COMPUTATION. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. p. 506–513. Special Session on Real-Parameter Optimization. 80

SHANG, Y.; QIU, Y. A note on the extended rosenbrock function. **Evolutionary Computation**, v. 14, p. 119–126, 2006. 64

SINHA, A.; TIWARI, S.; DEB, K. A population-based, steady-state procedure for real-parameter optimization. In: CONGRESS ON EVOLUTIONARY COMPUTATION. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. Special Session on Real-Parameter Optimization. 80

SOUZA, F. L. de. **Otimização extrema generalizada: um novo algoritmo estocástico para o projeto ótimo**. 144 p. (INPE-9564-TDI/836). Tese (Doutorado em Computação Aplicada) — INPE, São José dos Campos, 2002. 63, 65, 73

SUGANTHAN, P.; HANSEN, N.; LIANG, J. J.; DEB, K.; CHEN, Y. P.; AUGER, A.; TIWARI, S. **Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization**: Technical report. Nanyang Technological University and Kanpur Genetic Algorithms Laboratory, 2005. Disponível em: <<http://web.mysites.ntu.edu.sg/epnsugan/PublicSite/Shared%20Documents/CEC2005/Tech-Report-May-30-05.pdf>>. Acesso em: 10 jul. 2012. 80, 117

TSALLIS, C. Nonextensive physics: a possible connection between generalized statistical mechanics and quantum groups. **Physics Letters A**, v. 195, n. 5-6, p. 329–334, 1994. 29

VANDERPLAATS, G. N. **Numerical optimization techniques for engineering design**: with applications. New York: McGraw-Hill, Inc, 1984. 1, 7, 11, 14, 50

WOLPERT, D. H.; MACREADY, W. G. **No Free Lunch Theorems for Optimization**. 1997. 67–82 p. 3

YUAN, B.; GALLAGHER, M. Experimental results for the special session on real-parameter optimization at cec 2005: a simple, continuous eda. In: CONGRESS ON EVOLUTIONARY COMPUTATION. **Proceedings...** Edinburgh, Scotland: IEEE, 2005. p. 1792–1799. Special Session on Real-Parameter Optimization. 80

ZABINSKY, Z. B. **Stochastic adaptive search for global optimization**. [S.l.]: Kluwer Academics Publishers, 2003. (Nonconvex optimization and its applications, v. 72). 2, 11, 12, 13, 23, 25

ZHU, Y. Qbb global optimization method. In: FLOUDAS, C. A.; PARDALOS, P. M. (Ed.). **Encyclopedia of optimization**. 2. ed. New York: Springer, 2009. p. 3111–3116. 23



## APÊNDICE A - TABELAS COMPLEMENTARES PARA AS FUNÇÕES DO CEC 2005

Neste apêndice são exibidas tabelas complementares referentes ao segundo conjunto de funções teste extraídas do *IEEE Congress on Evolutionary Computation* para o ano de 2005 (CEC-2005), conforme descrito na Subseção 4.4 do Capítulo 4.

A Tabela B.1 ilustra a precisão estabelecida que cada função deve atingir antes do número máximo de avaliações da função objetivo ser alcançado (NAFO máximo) para que uma execução independente seja considerada bem sucedida. Esses valores foram extraídos do relatório elaborado por Suganthan et al. (2005).

Os valores do erro da função ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) obtidos 25 execuções independentes foram ordenados e as Tabelas B.2 a B.7 ilustram o 1° (melhor), 7°, 13° (mediano), 19° e 25° (pior), além da média (ME) e do desvio-padrão (DP), para as funções F1 a F15.

Tabela B.1 - Precisões estabelecidas para as funções F1 a F15 do CEC 2005.

Função	Precisão	Função	Precisão
F1	-450+1E-06	F9	-330+1E-02
F2	-450+1E-06	F10	-330+1E-02
F3	-450+1E-06	F11	90+1E-02
F4	-450+1E-06	F12	-460+1E-02
F5	-310+1E-06	F13	-130+1E-02
F6	390+1E-02	F14	-300+1E-02
F7	-180+1E-02	F15	120+1E-02
F8	-140+1E-02		

Tabela B.2 - Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F1 a F8 com  $n = 10$ .

NAFO		F1	F2	F3	F4	F5	F6	F7	F8
1e3	1°	9,6065E-12	1,4806E+02	2,5377E+08	5,6770E+03	2,3228E+03	1,7394E+02	7,9122E+02	2,0397E+01
	7°	1,0326E-09	2,0047E+02	1,6897E+09	1,0449E+04	3,8056E+03	1,1571E+03	2,1845E+03	2,0718E+01
	13°	4,3793E-09	3,4862E+02	2,5116E+09	1,5147E+04	4,5585E+03	1,8936E+03	2,6362E+03	2,0765E+01
	19°	6,2535E-09	1,2670E+03	4,3064E+09	2,0456E+04	5,5085E+03	5,1067E+03	3,2840E+03	2,0824E+01
	25°	9,8620E-09	3,5157E+03	5,4402E+09	4,1289E+04	6,2143E+03	9,3435E+04	5,6404E+03	2,0932E+01
	ME	4,1154E-09	9,1878E+02	2,8336E+09	1,7052E+04	4,5778E+03	8,3278E+03	2,8767E+03	2,0754E+01
	DP	3,2726E-09	1,1001E+03	1,5300E+09	9,0729E+03	1,0311E+03	1,9308E+04	1,0106E+03	1,1520E-01
1e4	1°	9,6065E-12	8,9512E-09	9,0935E+07	1,1748E+01	3,3120E+00	1,2097E+01	1,1863E+00	2,0212E+01
	7°	1,0326E-09	1,4144E-08	1,8753E+08	2,3080E+01	1,0225E+01	2,4814E+01	1,5738E+00	2,0447E+01
	13°	4,3793E-09	6,8409E-08	2,6517E+08	2,8700E+01	2,2068E+01	2,7132E+01	1,7451E+00	2,0562E+01
	19°	6,2535E-09	1,0730E-07	6,5627E+08	3,4947E+01	3,3216E+01	3,0885E+01	2,2969E+00	2,0607E+01
	25°	9,8620E-09	5,1437E-07	9,9164E+08	3,0601E+02	7,1780E+01	3,6844E+01	2,9276E+00	2,0684E+01
	ME	4,1154E-09	9,2049E-08	4,1714E+08	4,5215E+01	2,3593E+01	2,7189E+01	1,9092E+00	2,0525E+01
	DP	3,2726E-09	1,1432E-07	2,8781E+08	5,9284E+01	1,6943E+01	5,8096E+00	4,8904E-01	1,1618E-01
1e5	1°	9,6065E-12	8,9512E-09	7,8291E+03	6,6926E-09	7,3989E-03	3,3229E-03	1,2686E-05	2,0003E+01
	7°	1,0326E-09	9,6832E-09	1,8149E+04	7,8301E-09	8,9922E+00	3,7533E-03	2,4036E-05	2,0021E+01
	13°	4,3793E-09	9,8081E-09	3,6548E+04	8,8685E-09	1,9043E+01	4,0050E-03	3,0367E-05	2,0055E+01
	19°	6,2535E-09	9,9122E-09	6,9692E+04	9,7343E-09	3,1100E+01	4,2135E-03	9,8696E-03	2,0131E+01
	25°	9,8620E-09	9,9841E-09	2,0720E+05	9,9948E-09	6,9190E+01	5,0231E-03	9,8913E-03	2,0479E+01
	ME	4,1154E-09	9,7238E-09	4,7629E+04	8,7986E-09	2,1068E+01	4,0228E-03	3,3816E-03	2,0114E+01
	DP	3,2726E-09	2,6910E-10	4,5383E+04	9,5761E-10	1,6845E+01	3,7668E-04	4,5975E-03	1,3427E-01



Tabela B.3 - Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F9 a F15 com  $n = 10$ .

NAFO		F9	F10	F11	F12	F13	F14	F15
1e3	1°	5,4186E+01	7,5702E+01	8,7139E+00	1,0840E+03	2,9334E+00	3,9377E+00	3,0422E+02
	7°	6,4307E+01	9,9045E+01	1,1107E+01	2,7285E+03	4,5864E+00	4,4375E+00	5,2371E+02
	13°	8,2249E+01	1,3385E+02	1,1535E+01	4,4089E+03	4,9569E+00	4,6937E+00	6,1130E+02
	19°	9,2818E+01	1,5827E+02	1,2098E+01	6,9723E+03	5,5097E+00	4,9435E+00	6,7069E+02
	25°	1,1137E+02	2,1606E+02	1,3459E+01	1,1355E+04	6,5866E+00	4,9884E+00	9,5877E+02
	ME	8,1204E+01	1,3550E+02	1,1537E+01	5,1745E+03	4,9630E+00	4,6159E+00	6,2390E+03
	DP	1,6149E+01	3,7701E+01	1,0156E+00	3,1380E+03	8,4859E-01	3,4355E-01	1,5540E+03
1e4	1°	3,4030E+00	2,3357E+00	9,0003E-01	2,0229E+02	2,8244E-05	3,9377E+00	5,2198E-01
	7°	4,7788E+00	6,7811E+00	1,2057E+00	2,8042E+02	1,5973E+00	4,2227E+00	1,4213E+02
	13°	8,2428E+00	9,5794E+00	1,4844E+00	4,4677E+02	2,1083E+00	4,4375E+00	2,9017E+02
	19°	1,0633E+01	1,3081E+01	2,5800E+00	5,4095E+02	2,6311E+00	4,5093E+00	4,0811E+02
	25°	1,9128E+01	2,5017E+01	4,9806E+00	3,0897E+03	3,0220E+00	4,9300E+00	7,6368E+02
	ME	8,3274E+00	1,0442E+01	2,0105E+00	6,5890E+02	2,0650E+00	4,3828E+00	3,1262E+02
	DP	4,1698E+00	5,0289E+00	1,1023E+00	6,7116E+02	7,1587E-01	2,5076E-01	2,2088E+02
1e5	1°	5,9969E-09	6,9114E-09	2,9416E-02	1,2111E-04	1,8186E-09	3,9377E+00	9,5459E-10
	7°	8,0769E-09	1,9899E+00	3,7799E-02	2,4490E-04	1,5814E+00	4,1700E+00	1,4113E+02
	13°	9,3013E-09	3,9798E+00	5,7273E-01	8,5290E-01	2,1027E+00	4,4375E+00	2,6675E+02
	19°	9,9012E-09	7,9597E+00	1,5350E+00	1,0003E+01	2,6311E+00	4,5092E+00	4,0626E+02
	25°	2,9849E+00	1,5070E+01	3,7468E+00	1,5566E+03	2,9938E+00	4,9300E+00	7,6351E+02
	ME	4,3274E+00	5,0959E+00	9,3594E-01	9,4809E+01	2,0242E+00	4,3682E+00	3,0541E+02
	DP	8,4459E+00	3,9891E+00	1,0824E+00	3,3583E+02	7,3707E-01	2,5262E-01	2,2277E+02

Tabela B.4 - Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F1 a F8 com  $n = 30$ .

NAFO		F1	F2	F3	F4	F5	F6	F7	F8
1e3	1°	5,2279E-04	1,2661E+05	2,7113E+09	8,0246E+04	1,4402E+04	2,1510E+07	1,8659E+03	2,1010E+02
	7°	9,9599E-04	3,6394E+05	3,6592E+09	1,0565E+05	1,8721E+04	1,3167E+09	2,6041E+03	2,1160E+02
	13°	1,5850E-03	6,4811E+05	5,4649E+09	1,2482E+05	1,9946E+04	5,0726E+09	2,9271E+03	2,1215E+02
	19°	2,0713E-03	1,0401E+06	6,3309E+09	1,6897E+05	2,1890E+04	9,5466E+09	3,3058E+03	2,1229E+0
	25°	2,8852E-03	2,3291E+06	7,4135E+09	3,0525E+05	2,6421E+04	3,0228E+10	4,9661E+03	2,1316E+02
	ME	1,5688E-03	8,5756E+05	5,2336E+09	1,4695E+05	2,0079E+04	6,8928E+09	3,0009E+03	2,1194E+02
	DP	6,4833E-04	6,3946E+05	1,4446E+09	6,0644E+04	3,1411E+03	6,9305E+09	7,0790E+02	6,6718E-01
1e4	1°	5,6843E-14	1,1096E+04	2,2320E+09	6,1236E+04	1,2272E+03	3,2887E+02	6,9904E-01	2,0984E+01
	7°	8,2252E-11	1,7419E+04	2,9583E+09	7,5872E+04	1,5432E+03	4,8646E+02	7,5472E-01	2,1025E+01
	13°	2,0208E-09	2,5834E+04	3,3006E+09	8,5245E+04	1,7710E+03	6,0340E+02	8,0131E-01	2,1092E+01
	19°	5,5543E-09	7,0218E+04	3,5678E+09	9,5894E+04	2,0412E+03	3,8373E+03	8,6429E-01	2,1115E+01
	25°	9,9169E-09	3,7323E+05	3,9437E+09	1,4587E+05	2,7198E+03	3,5217E+04	9,4631E-01	2,1177E+01
	ME	3,1566E-09	7,6226E+04	3,2610E+09	8,9313E+04	1,8020E+03	4,2946E+03	8,1093E-01	2,1077E+01
	DP	3,3961E-09	1,0356E+05	4,6600E+08	2,1408E+04	3,6969E+02	8,5461E+03	6,7899E-02	5,2880E-02
1e5	1°	5,6843E-14	1,2542E-03	7,5062E+06	6,3415E+03	6,1981E+01	2,8669E+01	8,2256E-09	2,0889E+01
	7°	8,2252E-11	1,8859E-03	1,1109E+07	1,0859E+04	9,6550E+01	2,9398E+01	9,3036E-09	2,0974E+01
	13°	2,0208E-09	2,3718E-03	1,1974E+07	1,6262E+04	1,0801E+02	2,9615E+01	9,8089E-09	2,1009E+01
	19°	5,5543E-09	2,9596E-03	1,3092E+07	2,5981E+04	1,5659E+02	2,9951E+01	7,3960E-03	2,1045E+01
	25°	9,9169E-09	4,5142E-03	1,4963E+07	7,9468E+04	2,2998E+02	1,0270E+03	9,8573E-03	2,1124E+01
	ME	3,1566E-09	2,5541E-03	1,1912E+07	2,5068E+04	1,2762E+02	1,1172E+02	2,7610E-03	2,1010E+01
	DP	3,3961E-09	9,5079E-04	1,6666E+06	2,1756E+04	5,0037E+01	2,3284E+02	3,7870E-03	5,4860E-02
3e5	1°	5,6843E-14	1,2299E-08	1,3914E+05	9,0031E+00	6,0112E-01	4,5863E-04	8,2256E-09	2,0001E+01
	7°	8,2252E-11	2,2821E-08	2,9761E+05	2,7205E+03	2,1897E+01	1,2624E-01	9,3036E-09	2,0022E+01
	13°	2,0208E-09	5,2215E-08	3,5858E+05	6,3019E+03	3,7450E+01	1,4688E-01	9,8089E-09	2,0037E+01
	19°	5,5543E-09	1,4939E-07	5,2564E+05	1,3868E+04	8,2673E+01	1,7914E-01	7,3960E-03	2,0051E+01
	25°	9,9169E-09	2,5839E-07	8,5883E+05	7,9468E+04	1,4935E+02	1,0141E+03	9,8573E-03	2,0091E+01
	ME	3,1566E-09	8,3140E-08	4,0210E+05	1,5100E+04	5,7427E+01	7,9914E+01	2,7610E-03	2,0038E+01
	DP	3,3961E-09	7,4008E-08	1,7544E+05	2,0785E+04	4,5425E+01	2,3512E+02	3,7870E-03	2,1964E-02

Tabela B.5 - Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F9 a F15 com  $n = 30$ .

NAFO		F9	F10	F11	F12	F13	F14	F15
1e3	1°	4,7449E+02	6,5199E+02	4,1419E+01	2,5392E+04	5,1193E+01	1,3795E+01	6,7663E+02
	7°	5,7929E+02	8,8771E+02	4,3351E+01	5,1323E+04	1,0498E+02	1,4340E+01	9,0362E+02
	13°	6,2663E+02	1,0134E+03	4,5679E+01	6,8070E+04	1,3482E+02	1,4445E+01	9,9340E+02
	19°	6,6372E+02	1,1202E+03	4,6252E+01	9,4317E+04	2,4175E+02	1,4548E+01	1,0523E+03
	25°	7,3264E+02	1,2848E+03	4,7908E+01	1,5496E+05	6,5129E+02	1,4992E+01	1,1974E+03
	ME	6,2086E+02	9,9412E+02	4,4961E+01	7,4662E+04	1,9756E+02	1,4474E+01	9,5697E+02
	DP	6,1925E+01	1,6602E+02	1,9773E+00	3,4028E+04	1,5727E+02	3,0207E-01	1,4001E+02
1e4	1°	2,0677E+02	2,4017E+02	2,5901E+01	5,6691E+03	1,2218E+01	1,3716E+01	2,4947E+02
	7°	2,3330E+02	2,9225E+02	2,8850E+01	8,4993E+03	1,3745E+01	1,4112E+01	4,3576E+02
	13°	2,4727E+02	3,1298E+02	3,0000E+01	1,0302E+04	1,4760E+01	1,4212E+01	4,6323E+02
	19°	2,6782E+02	3,4909E+02	3,1957E+01	1,3929E+04	1,6353E+01	1,4496E+01	5,6676E+02
	25°	2,8867E+02	3,9431E+02	3,3684E+01	5,3273E+04	2,3795E+01	1,4862E+01	8,8770E+02
	ME	2,4865E+02	3,1703E+02	3,0083E+01	1,2900E+04	1,5448E+01	1,4267E+01	5,0622E+02
	DP	2,1958E+01	3,9951E+01	2,0410E+00	9,4363E+03	2,7536E+00	2,6696E-01	1,5755E+02
1e5	1°	1,3963E+00	1,9153E+01	1,2072E-01	1,9339E+01	1,4043E+00	1,3716E+01	5,1073E+01
	7°	2,2833E+00	4,2439E+01	2,4859E+00	3,5653E+02	2,2831E+00	1,4024E+01	2,7481E+02
	13°	4,9293E+00	5,1641E+01	3,7839E+00	1,2358E+03	2,6586E+00	1,4190E+01	3,0624E+02
	19°	9,0515E+00	6,6981E+01	5,7158E+00	4,9993E+03	3,5712E+00	1,4445E+01	4,9050E+02
	25°	2,1201E+01	8,5683E+01	1,0232E+01	1,7958E+04	7,7727E+00	1,4531E+01	8,8344E+02
	ME	6,9930E+00	5,3274E+01	4,4808E+00	3,1253E+03	2,9771E+00	1,4202E+01	3,7289E+02
	DP	6,0000E+00	1,6376E+01	2,7646E+00	4,3267E+03	1,2711E+00	2,3659E-01	2,2085E+02
3e5	1°	7,3987E-06	2,9850E+00	1,2072E-01	3,3673E-03	1,3482E+00	1,3716E+01	5,0834E+01
	7°	1,2610E-05	1,7610E+01	2,4859E+00	3,2744E+02	2,1920E+00	1,4024E+01	2,7180E+02
	13°	2,1245E+00	2,1599E+01	3,7547E+00	1,2097E+03	2,6245E+00	1,4190E+01	3,0189E+02
	19°	5,0558E+00	3,3164E+01	5,7158E+00	4,9768E+03	3,4998E+00	1,4445E+01	4,9043E+02
	25°	1,8652E+03	4,9432E+01	1,0232E+01	1,7930E+04	7,6409E+00	1,4531E+01	8,8341E+02
	ME	4,5465E+00	2,4633E+01	4,4704E+00	3,0972E+03	2,9059E+00	1,4202E+01	3,7137E+02
	DP	5,6397E+00	1,1965E+01	2,7635E+00	4,3275E+03	1,2542E+00	2,3659E-01	2,2127E+02

Tabela B.6 - Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F1 a F8 com  $n = 50$ .

NAFO		F1	F2	F3	F4	F5	F6	F7	F8
1e3	1°	1,0708E+00	4,4501E+05	2,2086E+09	2,8830E+05	2,3966E+04	2,7134E+08	1,4374E+04	2,1244E+01
	7°	1,4556E+00	1,2920E+06	3,1349E+09	5,3498E+05	3,1468E+04	1,8830E+09	1,5899E+04	2,1295E+01
	13°	1,5943E+00	2,1832E+06	3,6588E+09	8,4446E+05	3,4927E+04	4,7180E+09	1,6877E+04	2,1338E+01
	19°	2,0210E+00	4,3618E+06	4,6948E+09	1,2330E+06	3,6358E+04	8,0790E+09	1,7461E+04	2,1357E+01
	25°	2,6368E+00	6,4179E+06	5,9841E+09	2,8443E+06	4,5672E+04	2,1670E+10	1,9611E+04	2,1399E+01
	ME	1,7056E+00	2,7868E+06	3,7897E+09	1,0408E+06	3,4226E+04	6,2988E+09	1,6784E+04	2,1325E+01
	DP	3,9996E-01	1,8180E+06	1,0429E+09	6,9952E+05	4,8228E+03	5,9071E+09	1,4693E+02	4,5023E-02
1e4	1°	2,0066E-11	5,7778E+04	1,4954E+09	1,2668E+05	5,5736E+03	7,3688E+02	1,2736E+03	2,1182E+01
	7°	1,9433E-09	1,2638E+05	2,3762E+09	1,6502E+05	6,0564E+03	3,4756E+03	1,6342E+03	2,1231E+01
	13°	6,5368E-09	4,8812E+05	2,8924E+09	1,9444E+05	6,3446E+03	4,6253E+03	1,7485E+03	2,1266E+01
	19°	8,1025E-09	1,4262E+06	3,2620E+09	2,1074E+05	6,6286E+03	3,2416E+04	1,9880E+03	2,1284E+01
	25°	9,7878E-09	2,6783E+06	4,1277E+09	3,3928E+05	7,7965E+03	6,1824E+04	2,4145E+03	2,1320E+01
	ME	5,2773E-09	8,3771E+05	2,8586E+09	1,9473E+05	6,3753E+03	1,6064E+04	1,8086E+03	2,1256E+01
	DP	3,3980E-09	8,2891E+05	6,4080E+08	4,5025E+04	5,5469E+02	1,8718E+04	3,2198E+02	4,0722E-02
1e5	1°	2,0066E-11	1,4265E+01	1,1336E+07	5,4995E+04	9,8593E+02	4,8966E+01	6,7835E-02	2,1130E+01
	7°	1,9433E-09	2,2911E+01	1,2998E+07	7,8077E+04	1,1410E+03	4,9385E+01	8,4629E-02	2,1166E+01
	13°	6,5368E-09	3,0116E+01	1,7115E+07	9,0567E+04	1,2475E+03	4,9903E+01	8,9572E-02	2,1182E+01
	19°	8,1025E-09	3,9943E+01	3,7767E+07	1,0311E+05	1,2776E+03	5,0449E+01	9,4855E-02	2,1216E+01
	25°	9,7878E-09	2,6307E+04	9,3022E+07	1,2775E+05	1,4303E+03	3,3445E+02	1,3527E-01	2,1257E+01
	ME	5,2773E-09	4,3003E+01	2,9852E+07	9,1765E+04	1,2183E+03	6,6015E+01	9,0800E-02	2,1189E+01
	DP	3,3980E-09	4,8261E+01	2,3362E+07	1,8321E+04	1,1231E+02	5,9343E+01	1,4204E-02	3,3089E-02
5e5	1°	2,0066E-11	3,3531E-04	1,7739E+05	3,5662E+03	2,4123E+01	4,3809E+00	8,1266E-09	2,0002E+01
	7°	1,9433E-09	7,2690E-04	2,5905E+05	1,2518E+04	1,9803E+02	3,1303E+01	9,0793E-09	2,0010E+01
	13°	6,5368E-09	1,1573E-03	2,9703E+05	1,6322E+04	2,6744E+02	3,1743E+01	9,3711E-09	2,0015E+01
	19°	8,1025E-09	1,6751E-03	3,4624E+05	2,2450E+04	3,4391E+02	3,1962E+01	9,9171E-09	2,0021E+01
	25°	9,7878E-09	4,4858E-01	4,7829E+05	9,2165E+04	4,3691E+02	2,5940E+02	9,8573E-03	2,0053E+01
	ME	5,2773E-09	2,5508E-02	3,1096E+05	2,0546E+04	2,6782E+02	3,9704E+00	1,0844E-03	2,0019E+01
	DP	3,3980E-09	9,0871E-02	7,3904E+04	1,7020E+04	1,0815E+02	4,6097E+00	3,0251E-03	1,4250E-02

Tabela B.7 - Valores do erro ( $F(\mathbf{x}_{melhor}) - F(\mathbf{x}^*)$ ) para as funções F9 a F15 com  $n = 50$ .

NAFO		F9	F10	F11	F12	F13	F14	F15
1e3	1°	1,0040E+03	1,2615E+03	7,5173E+01	2,1740E+05	6,6134E+02	2,3444E+01	6,6406E+02
	7°	1,0610E+03	1,5033E+03	8,0106E+01	3,2797E+05	1,4520E+03	2,4185E+01	7,8103E+02
	13°	1,1125E+03	1,6690E+03	8,0987E+01	4,0239E+05	2,0589E+03	2,4341E+01	9,7549E+02
	19°	1,2555E+03	1,8114E+03	8,3822E+01	4,6880E+05	3,5046E+03	2,4456E+01	1,1639E+03
	25°	1,3797E+03	2,0884E+03	8,5619E+01	7,9109E+05	5,8414E+03	2,4941E+01	1,4250E+03
	ME	1,1650E+03	1,6692E+03	8,1392E+01	4,2159E+05	2,4631E+03	2,4278E+01	9,8291E+02
	DP	1,2466E+02	2,2904E+02	2,8315E+00	1,4141E+05	1,2844E+03	3,5002E-01	2,1200E+02
1e4	1°	5,1255E+02	4,0038E+02	7,1889E+01	2,3914E+04	3,0501E+01	2,3444E+01	2,5206E+02
	7°	5,5758E+02	4,7146E+02	7,5652E+01	3,3027E+04	3,5677E+01	2,4063E+01	3,9523E+02
	13°	5,8630E+02	4,8969E+02	7,7604E+01	4,8083E+04	3,8090E+01	2,4219E+01	4,3934E+02
	19°	6,0224E+02	5,1108E+02	7,9045E+01	6,6745E+04	3,9312E+01	2,4358E+01	4,4669E+02
	25°	6,3618E+02	5,4505E+02	8,0258E+01	9,7113E+04	5,7991E+01	2,4436E+01	5,1496E+02
	ME	5,8119E+02	4,8455E+02	7,7243E+01	5,1571E+04	3,8434E+01	2,4109E+01	4,2233E+02
	DP	3,2508E+01	3,4350E+01	2,1784E+00	2,1787E+04	5,3773E+00	3,1533E-01	6,2515E+01
1e5	1°	4,2553E+01	3,0848E+01	4,0605E+01	6,5000E+02	4,4348E+00	2,3444E+01	4,9271E+01
	7°	7,5198E+01	5,7708E+01	4,2994E+01	4,0860E+03	4,8681E+00	2,3906E+01	1,9792E+02
	13°	8,8940E+01	6,9648E+01	4,5191E+01	8,5244E+03	5,1391E+00	2,4164E+01	2,9923E+02
	19°	1,0439E+02	7,4745E+01	4,8464E+01	1,5538E+04	6,2585E+00	2,4334E+01	3,9199E+02
	25°	1,1977E+02	1,1554E+02	5,2051E+01	3,6154E+04	1,7328E+01	2,4409E+01	4,3237E+02
	ME	8,8397E+01	6,8906E+01	4,5865E+01	1,1458E+04	6,0895E+00	2,4083E+01	2,9039E+02
	DP	2,0328E+01	1,8181E+01	3,2755E+00	9,8076E+03	2,6111E+00	3,0626E-01	1,1119E+02
5e5	1°	2,6301E-05	3,0847E+01	2,0458E+00	5,2663E+03	2,7078E+00	2,3444E+01	4,9269E+01
	7°	2,1230E+00	5,7708E+01	4,6467E+00	4,0714E+04	3,3086E+00	2,3906E+01	1,9745E+02
	13°	3,9906E+00	6,9647E+01	5,9019E+00	8,5144E+04	3,8978E+00	2,4164E+01	2,9887E+02
	19°	2,4204E+01	7,4726E+01	8,3901E+00	1,5538E+05	4,6271E+00	2,4334E+01	3,9199E+02
	25°	3,2331E+01	1,1552E+02	1,2301E+01	3,5572E+05	1,4120E+01	2,4409E+01	4,3208E+02
	ME	1,2129E+01	6,8892E+01	6,5827E+00	1,1240E+05	4,4546E+00	2,4083E+01	2,9018E+02
	DP	1,1783E+01	1,8177E+01	2,8699E+00	9,7059E+04	2,2774E+00	3,0626E-01	1,1121E+02



## PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

### **Teses e Dissertações (TDI)**

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

### **Manuais Técnicos (MAN)**

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

### **Notas Técnico-Científicas (NTC)**

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

### **Relatórios de Pesquisa (RPQ)**

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

### **Propostas e Relatórios de Projetos (PRP)**

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

### **Publicações Didáticas (PUD)**

Incluem apostilas, notas de aula e manuais didáticos.

### **Publicações Seriadas**

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Contam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

### **Programas de Computador (PDC)**

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

### **Pré-publicações (PRE)**

Todos os artigos publicados em periódicos, anais e como capítulos de livros.