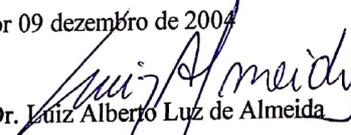


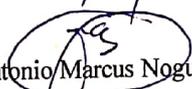


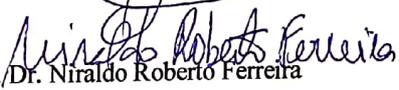
Ata de Defesa de Dissertação Colegiado do Curso de Mestrado em Engenharia Elétrica da UFBA

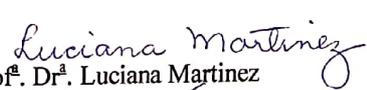
Ata da sessão especial do Colegiado do Curso de Mestrado em Engenharia Elétrica da Escola Politécnica da Universidade Federal da Bahia, realizada em 09 de dezembro de dois mil e quatro, para instalação da Banca Examinadora de Dissertação de Mestrado em Engenharia Elétrica do candidato **Alberto Silva Betzler** intitulada “**Estimação de Parâmetros do Modelo de Histerese L^2P Através de Técnicas Variadas de Otimização**”. Às quatorze horas do citado dia, na Sala Nildo Peixoto do Departamento de Engenharia Elétrica da Escola Politécnica da UFBA, realizou-se a Sessão Pública do Colegiado do Curso de Mestrado em Engenharia Elétrica para instalação da Banca Examinadora de Dissertação de Mestrado em Engenharia Elétrica do referido candidato. Compareceram à sessão os seguintes membros da Banca Examinadora: **Prof. Dr. Luiz Alberto Luz de Almeida (Presidente)**, **Prof. Dr. Antonio Marcus Nogueira Lima (membro externo)**, **Prof. Dr. Niraldo Roberto Ferreira (membro interno)** e a **Prof.ª Dr.ª Luciana Martinez (Membro Interno)**. Instalada a Banca Examinadora, foram esclarecidos os procedimentos e foi passada a palavra ao examinado para apresentação do trabalho de Dissertação. Ao final da apresentação, passou-se à arguição por parte da Banca, a qual, em seguida, reuniu-se para a elaboração dos pareceres. No retorno, foram lidos os pareceres a respeito do trabalho apresentado pelo candidato, tendo a Banca Examinadora aprovado por unanimidade a dissertação apresentada. Em seguida, nada mais havendo a tratar, foi encerrada a sessão, pelo Prof. Luiz Alberto Luz de Almeida, Chefe do Departamento de Engenharia Elétrica, tendo sido, logo a seguir, lavrada a presente ata pelo secretário, *ad hoc*, José Batista de Araújo, que assina abaixo, juntamente com os membros da Banca Examinadora.

Salvador 09 dezembro de 2004

Prof. Dr. 
Luiz Alberto Luz de Almeida

Prof. Dr. 
Antonio Marcus Nogueira Lima

Prof. Dr. 
Niraldo Roberto Ferreira

Prof.ª Dr.ª 
Luciana Martinez

Sr. 
José Batista de Araújo

Coordenação do Curso de Mestrado em Engenharia Elétrica
Escola Politécnica - Rua Aristides Novis n.º 02 4º andar
Federação - CEP:40210-630 - Salvador - Bahia
Tel: (xx71) 203-9760 Fax: 203-9779
e-mail: cmee@ufba.br

Estimação de Parâmetros do Modelo de Histerese L^2P
Através de Técnicas Variadas de Otimização

Alberto Silva Betzler

Dissertação de mestrado submetida à Coordenação dos Cursos de Pós-Graduação em Engenharia Elétrica da Universidade Federal da Bahia, como parte dos requisitos para obtenção do grau de Mestre em Ciências no domínio da Engenharia Elétrica.

Área de Concentração: Processamento da Informação

Luiz Alberto Luz de Almeida, Dr.

Orientador

Luciana Martinez, Dra.

Orientadora

Salvador, Bahia, Brasil

©Alberto Silva Betzler, Dezembro de 2004

Estimação de Parâmetros do Modelo de Histerese L^2P
Através de Técnicas Variadas de Otimização

Alberto Silva Betzler

Dissertação de Mestrado

Luiz Alberto Luz de Almeida, Dr.

Orientador

Luciana Martinez, Dra.

Orientadora

Luiz Alberto Luz de Almeida, Dr.

Componente da Banca

Nivaldo Roberto Ferreira, Dr.

Componente da Banca

Antonio Marcus Nogueira Lima, Dr.Eng.

Componente da Banca

Salvador, Bahia, Brasil, Dezembro de 2004

Dedicatória

Em memória de Alexandre, Dominique, Lindonora e Jorge

Agradecimentos

- Ao meu irmão de coração, Dourival Edgar dos Santos Junior, por sua preciosa ajuda no momento em que mais precisava. Sinceramente eu lhe agradeço.
- À minha mãe, Dona Therezinha, pelo apoio constante.
- Aos professores do DEE-UFBA pelos cursos de alta qualidade ministrados.
- Aos professores Luiz Alberto Luz de Almeida e Luciana Martinez pela orientação amigável e paciente no presente trabalho.
- Às famílias de Dourival e Rosângela Karine pela calorosa acolhida que me deram em terras baianas.
- Aos professores Antônio Sérgio Esperidião e Julian Celedon pela valiosa ajuda com o FORTRAN 77 e MATLAB 6.5.
- À Dra. D. Lazzaro e Dr. A. Kleber, pesquisadores do MCT/Observatório Nacional, por terem me "empurrado" para esta maravilhosa nova vida.
- A todos os (inúmeros!) amigos que fiz aqui na Bahia, em especial, a Amilton André por seus valiosos conselhos.

Resumo

Muitos sensores e atuadores exibem histerese em pelo menos uma de suas características de interesse prático. Frequentemente, a histerese ocorre em atuadores mecânicos, através do atrito, podendo gerar dificuldades em aplicações de controle. Como um exemplo adicional, transformadores elétricos são simples sensores de corrente que apresentam histerese associada à magnetização e ao campo magnético. Neste casos, vários modelos são aplicados para descrever o comportamento da histerese, cada um destes tendo seus aspectos positivos e desvantagens. Recentemente, quando estudando a característica resistência-temperatura de detectores de infravermelhos de Dióxido de Vanádio, Almeida et al. (2003b) propuseram um operador de histerese algébrico simples, o modelo "Limiting Loop Proximity" (L^2P): Este modelo possui uma reduzida complexidade matemática e é capaz de descrever o fenômeno de histerese sem simetria. Analogamente ao demais operadores de histerese, o modelo L^2P necessita do uso de técnicas de busca e otimização para ajustar-se a dados experimentais.

Até este momento, não estavam disponíveis procedimentos matemáticos autônomos para extração de parâmetros do L^2P . Esta dissertação é centrada no estudo de algumas técnicas tradicionais de otimização e suas aplicações no ajuste do modelo L^2P com respeito a dados experimentais da histerese de filmes finos de Dióxido de Vanádio. Dois tipos principais de métodos de busca foram estudados: determinístico e randômico. Os métodos determinísticos buscam um ponto ótimo na vizinhança de um ponto inicial e são classificados em dois tipos: diretos e indiretos. O primeiro busca o extremo local solucionando um conjunto de equações não lineares enquanto o outro busca tal extremo ao longo de uma direção permitida usando o gradiente local. Os dois métodos determinísticos estudados são: Gauss-Newton e Levenberg-Marquardt. Algoritmos de busca randômica não são especializados e eficientes como os métodos determinísticos, mas estes são mais robustos e usualmente capazes de explorar um espaço de busca maior. Os algoritmos genéticos constituem um exemplo de método de busca randômica, baseado na mecânica da seleção natural. A extração dos parâmetros do modelo L^2P é também estudada usando um algoritmo genético. Os desempenhos destes três métodos de otimização são comparados e apresentados em termos do erro quadrático médio normalizado (e).

Abstract

Many sensors and actuators exhibit hysteresis in one of their features of practical interest. Frequently, hysteresis appears in mechanical actuators via friction and may cause many difficulties in the design of control applications. As an additional example, electrical transformers are simple current sensors which exhibit hysteresis between the magnetization and the magnetic field. There are several models employed to describe the hysteretic behavior, each of them having its own positive aspects and drawbacks. Recently, when studying the resistance-temperature characteristic of vanadium dioxide infrared detectors, Almeida et al. (2003b) have proposed a simple algebraic hysteresis operator, called Limiting Loop Proximity (L^2P) model. This model has reduced mathematical complexity and is capable of describing nonsymmetrical hysteresis phenomena. Similarly to most hysteresis operators, the L^2P model requires the use of search and optimization techniques to adjust it to experimental data. Up to this date, there is no mathematical autonomous procedure available to extract the L^2P parameters. This dissertation is focused on the study of some traditional optimization techniques and their application to adjust the L^2P model to hysteretic experimental data of vanadium dioxide thin-films. Two main types of search methods are investigated: calculus-based and random. Calculus-based methods seek the optima in a neighborhood of the initial point and are classified as two types: indirect and direct. The former seeks local extrema solving a set of nonlinear equation while the latter seeks local optima in the steepest permissible direction using local gradient. The two calculus-based methods investigated are: Gauss-Newton and Levenberg-Marquardt. Random search algorithms are not as specialized and efficient as calculus-based methods but are more robust and usually capable of exploring a much larger search space. The genetic algorithm is an example of random search which is based on the mechanics of natural selection. The parameter extraction of the L^2P model is also investigated using genetic algorithm. The performance of these optimization methods are compared and presented in terms of normalized root-mean square error (e).

Conteúdo

1	Introdução	5
1.1	Caraterização de um Bolômetro	6
1.2	Extração de Parâmetros em Modelos de Histerese	8
1.3	Objetivos da Investigação	9
1.4	Organização do Texto	9
2	O Fenômeno de Histerese	10
2.1	Introdução	10
2.2	Modelos de Histerese	13
2.2.1	O Modelo de Jiles-Atherton	15
2.2.2	O Modelo de Preisach	16
2.3	Resistência Elétrica nos Filmes Finos de VO_2	18
2.4	Histerese no Filme Fino de VO_2	19
2.4.1	Evolução da fração volumétrica	19
2.5	Modelagem da Histerese na Característica $g \times T$	21
2.6	Conclusões	22
3	Fundamentos e Métodos de Otimização	23
3.1	Introdução	23
3.2	Terminologia	23
3.3	Classes de Problemas de Otimização	25
3.3.1	Programação Linear	26
3.3.2	Programação Não-Linear	26
3.4	Otimização Sem Restrições	26
3.5	Otimização com Restrição	27
3.5.1	Busca ao Longo de uma Linha	28
3.6	O Problema de Mínimos Quadrados Não Linear	29
3.6.1	Cálculo do gradiente e Hessiana	31
3.7	Métodos de Otimização	32

3.7.1	Métodos de Otimização Sem Restrição	32
3.7.2	Método de Newton	32
3.7.3	Métodos Quase-Newton	33
3.7.4	Método Broyden-Fletcher-Goldfarb-Shanno (BFGS)	34
3.7.5	Método Davidon-Fletcher-Powell (DFP)	35
3.7.6	Método de Gauss-Newton (GN)	36
3.7.7	Método de Levenberg-Marquardt (LM)	38
3.8	Conclusões	40
4	Algoritmos Genéticos	42
4.1	Introdução	42
4.2	Terminologia Biológica	44
4.3	Propriedades dos Algoritmos Genéticos	44
4.4	Esquemas de Representação dos Cromossomos	45
4.4.1	Representação Binária	45
4.4.2	Representação Real	46
4.5	Operadores Genéticos	46
4.5.1	Operador de Seleção	47
4.5.2	Operador de Cruzamento ("Crossover")	48
4.5.3	Operador de Mutação	50
4.6	Elitismo	50
4.7	Codificação Real \times Binária	51
4.8	PIKAIA - Descrição da Estrutura Utilizada	51
4.9	Conclusões	53
5	Processamento e Análise	55
5.1	O Experimento	55
5.2	Problema de Otimização	57
5.3	Processamento de Dados Experimentais	58
5.3.1	Discussão e Conclusões	60
5.4	Processamento de Dados Gerados pelo Modelo L^2P	64
5.4.1	Primeiro Processamento	65
5.4.2	Segundo Processamento	65
5.4.3	Terceiro Processamento	68
5.4.4	Conclusões	69
5.5	Otimização de Outras Curvas de Primeira Ordem	70
6	Conclusões e Perspectivas Futuras	72

7 Bibliografía

73

Lista de Figuras

1.1	Curvas experimentais decrescentes de primeira ordem $R \times T$ (círculos sólidos) as curvas correspondentes obtidas através do modelo L^2P (linhas contínuas), com $w = 6,5^0C$, $T_c = 47,6^0C$, $\beta = 0,2^0C^{-1}$ e $\gamma = 0,85$. Tais valores são os parâmetros de referência (PR) a serem apresentados no capítulo 5. Fonte: Almeida (2003)	7
2.1	Diagrama esquemático dos domínios magnéticos em um material ferromagnético. No interior de um domínio, todos os dipolos magnéticos estão alinhados, mas como a direção de alinhamento varia de domínio para domínio, o momento magnético total da amostra pode ser pequeno ou mesmo nulo. Um pequeno campo magnético externo faz aumentar o tamanho dos domínios alinhados no sentido do campo, o que resulta em um aumento do momento magnético total.	11
2.2	Gráfico de B ou M em função do campo aplicado B_{ap} ou H . A curva fechada é conhecida como curva de histerese. O campo B_r é chamado de campo remanente, no qual a amostra de material ferromagnético tornou-se um imã permanente.	12
2.3	(a) Curva característica típica da histerese de materiais ferromagnéticos. (b) Operador de elemento de Preisach para materiais magnéticos e (c) o correspondente triângulo de Preisach. Fonte: Almeida (2003)	17
2.4	Características experimentais $R \times T$ do filme fino de dióxido de vanádio, incluindo laços menores. Fonte: Almeida (2003)	20
3.1	Extremos da função $x^2 \text{sen}(x)$ entre -6 e 6 radianos. O ponto C é máximo enquanto B é um mínimo, ambos locais. O máximo global ocorre em A. O mínimo global é D. . .	24
3.2	Características do vetor gradiente.	27
5.1	Dispositivos empregados e detalhes de montagem da câmara térmica: (a) módulo termoeletrico; (b) dimensões dos filmes utilizados nos experimentos; (c) montagem do filme e termistor sobre o módulo termoeletrico; e d) conjunto da câmara térmica. Fonte: Almeida (2003)	56

5.2	Representação esquemática da histerese na evolução da fração volumétrica com a temperatura. Esta construção geométrica de uma curva reversa ilustra o conceito de proximidade de um ponto nesta curva ao laço principal. Fonte Almeida (2003)	57
5.3	Curva de histerese do filme fino de Dióxido de Vanádio para uma amostragem de 684 pares de resistência e temperatura experimentais. Observa-se que o laço principal e as duas curvas de primeira ordem estão praticamente superpostas.	59
5.4	Mapa de contorno da superfície $e(\beta, \gamma)$ para 75 pares resistência-temperatura. O círculo (o) o ponto inicial para a otimização com o AG enquanto a cruz (+) o ponto ótimo para onde houve convergência. Nota-se que ambos encontram-se no interior de um curva de nível de um possível mínimo local.	61
5.5	Curva $R \times T$ do filme fino de Dióxido de Vanádio. As cruzes (+) representam o modelo com parâmetros otimizados pelo AG enquanto os círculos (o), os dados experimentais, para uma amostra de 75 pontos.	62
5.6	Curva $R \times T$ do filme fino de Dióxido de Vanádio. As cruzes (+) representam o modelo com parâmetros otimizados pelos métodos de GN e LM enquanto os círculos (o), os dados experimentais, para uma amostra de 75 pontos.	63
5.7	Curva $R \times T$ do filme fino de Dióxido de Vanádio. As cruzes (+) representam o modelo com os Parâmetros de Referência enquanto os círculos (o), os dados experimentais, para uma amostra de 75 pontos.	63
5.8	Curvas de histerese do filme fino de Dióxido de Vanádio com 198 pares resistência-temperatura. Os círculos (o) e as cruzes (+) representam, respectivamente, as resistências elétricas geradas pelo modelo L^2P sem e com um desvio relativo percentual de 20%.	64
5.9	Superfície de erro $e(\beta, \gamma)$ para os dados artificiais. Tal superfície é analogia a sua equivalente gerada a partir dos dados experimentais com uma amostragem de 75 pontos.	65
5.10	Mapa de contorno da superfície de $e(\beta, \gamma)$ mostrando as iterações do método LM a partir de $\beta = 1,8^0C^{-1}$ e $\gamma = 0,2$	66
5.11	Superfície de erro $e(w, \gamma)$ para os dados artificiais.	67
5.12	Mapa de contorno da superfície de $e(\beta, \gamma)$ mostrando as iterações do método LM a partir de $\beta = 0,3^0C^{-1}$ e $\gamma = 1,95$	67
5.13	Superfície de erro $e(w, T_c)$ para os dados artificiais.	68
5.14	Contorno da superfície de erro $e(w, T_c)$ para os dados artificiais mostrando as iterações do método LM a partir de $w = 6^0C$ e $T_c = 50^0C$	69
5.15	Detalhe da superfície $e(w, T_c)$ mostrando as iterações do método LM a partir de $w = 6^0C$ e $T_c = 50^0C$	70
5.16	Curvas de primeira ordem para o filme fino de VO_2 . Os círculos (o) representam os dados experimentais e as cruzes (+) o modelo com parâmetros estimados pelo AG.	71

Lista de Tabelas

5.1	Erro quadrático médio normalizado para o AG	58
5.2	PO,s fornecidos pelo AG	59
5.3	PO,s fornecidos pelos Métodos de GN e LM	60
5.4	ϵ e Iterações para os Métodos de GN e LM	61

Lista de Símbolos e Abreviaturas

Símbolos

A_ρ	semiplano definido como a região na qual $\rho^- < \rho^+$
A_ρ^T	triângulo de Preisach
A_ρ^{T-}	região no triângulo de Preisach para qual $\gamma_\rho[\cdot] = -1$
A_ρ^{T+}	região no triângulo de Preisach para qual $\gamma_\rho[\cdot] = +1$
$\mathcal{D}[\cdot]$	operador de Duhem
$\mathcal{D}_J[\cdot]$	operador de Jiles
e	erro médio quadrático normalizado
\mathbf{e}	vetor de erro entre o modelo e dados experimentais
E_a	energia de ativação
$f(\mathbf{x})$	função objetivo
$F_L(T)$	descreve as curvas ascendentes e descendentes principais de g
$g(\mathbf{x})$	vetor gradiente
g	fração volumétrica de microcristais na fase semicondutora do filme de VO_2
\tilde{g}	fração volumétrica experimental
G	condutância térmica de um corpo
$G(\mathbf{x})$	matriz hessiana (matriz de segundas derivadas)
H	campo magnético excitador
I_p	corrente de polarização
k	constante de Boltzman
M	magnetização
M_{irr}	componente irreversível de magnetização
M_{rev}	componente reversível magnetização
$P(\cdot)$	função de proximidade
P	fluxo de calor gerado por efeito Joule no filme
P_e	fluxo de calor transportado para dentro de um corpo
P_r	potência de radiação incidente no filme
R	resistência elétrica

R_m	resistência da região metálica no VO ₂
R_0	resistência para $T \rightarrow \infty$ em R_s
R_s	resistência da região semicondutora no VO ₂
$R - T$	plano de fase resistência-temperatura
$R \times T$	característica resistência-temperatura no VO ₂
t	tempo
T	temperatura
T_c	temperatura crítica ou limiar de percolação
T_s	temperatura do substrato
T_{pr}	temperatura de proximidade
(T_0, g_0)	ponto de interesse no plano $R - T$
(T_r, g_r)	ponto de reversão qualquer no plano $g - T$
V_p	tensão de polarização
i, j, k	índices quaisquer
x	variável em um problema de otimização
x^*	solução local
$x^{(k)}$	variável em um passo $k = 1, 2, \dots$ de um processo iterativo
w	parâmetro de $F_L(T)$ - largura da histerese
δ	sinal de $\frac{d(\cdot)}{dt}$
$\gamma_\rho[\cdot]$	operador elementar de Preisach
$\rho = (\alpha, \beta)$	sistema de coordenadas do modelo de Preisach
$\rho = (r, s)$	sistema de coordenadas do modelo de Preisach
ρ^-	limiar inferior de transição do operador elementar de Preisach
ρ^+	limiar superior de transição do operador elementar de Preisach
$\mu(\rho)$	função de distribuição de Preisach
β	parâmetro de $F_L(T)$ - inclinação da histerese
ρ_s	resistividade da fase semicondutora no filme de VO ₂
ρ_m	resistividade da fase metálica no filme de VO ₂
ΔR	diferença de resistência elétrica
ΔT	diferença de temperatura

Abreviaturas

No caso de abreviaturas internacionalmente conhecidas, optou-se por manter as mesmas em inglês.

EMA	Effective Medium Approximation
TCR	Temperature Coefficient of Resistance
TEM	Thermoelectric Module.

Capítulo 1

Introdução

Detectores de radiação, em várias bandas do espectro eletromagnético, possuem inúmeras aplicações em ciência e tecnologia. Dentre estes detectores, existem os sensores térmicos que são constituídos de um material que absorve radiação incidente e a converte em variação de temperatura (ΔT) (Almeida, 2003). O material sensor é colocado sobre um suporte que propicia isolamento térmico entre este e o dissipador ou banho térmico. Os diversos sensores térmicos existentes diferem no mecanismo de conversão de temperatura em grandezas físicas que possibilitam a determinação da potência absorvida pelo conjunto suporte/sensor.

Sensores pneumáticos convertem ΔT em variação de volume de um gás contido em um recipiente. Esta expansão é medida a partir da deflexão de um feixe de raios LASER em um micro-espelho ligado ao recipiente. Nos sensores piroelétricos, o material sensor possui uma constante dielétrica que varia com a temperatura. Outro sensor térmico é o bolômetro. Neste detector, a variação de temperatura ocasiona mudança de resistência elétrica do material sensor.

Em todos estes sensores, características devem ser adaptadas para uma operação eficiente, em frequências de interesse. Em especial, os bolômetros que empregam como material sensor o Dióxido de Vanádio (VO_2) possuem uma taxa elevada de absorção de radiação na faixa do infravermelho próximo. Experimentos realizados por Malyarov et al. (1999) com sensores constituídos por um "sandwich" de SiH e VO_2 , atuando com uma cavidade óptica, absorveram cerca de 80% da radiação em comprimento de onda de $8\mu\text{m}$. No filme de VO_2 , a absorção foi de 50% a 80% na banda espectral compreendida entre $8,5\mu\text{m}$ e $10\mu\text{m}$. Por estas características, o VO_2 é utilizado na fabricação de sensores e transdutores com aplicações em detecção, modulação e chaveamento óptico (Jerominek et al., 1996). Quando depositado na forma de um filme fino¹, apresenta transição de fase, o que implica na mudança da característica condutora para semicondutora. Tal fenômeno gera mudanças significativas nas propriedades elétricas do filme (Choi et al., 1996).

¹Filmes finos de Dióxido de Vanádio são policristalinos, com espessura em torno de $1\mu\text{m}$, nos quais os microcristais têm diâmetro médio de $0,3\mu\text{m}$.

Dentre os materiais normalmente utilizados em bolômetros refrigerados, os óxidos supercondutores possuem um coeficiente de variação de resistência (R) com a temperatura (T)² (Temperature Coefficient of Resistance - $TCR = R^{-1}dR/dT$) em torno de $40\%/C^{-1}$. Mas estes têm que operar em temperaturas muito baixas tais como as de condensação do Hélio e Nitrogênio. O VO_2 é uma alternativa atraente na construção de bolômetros não-refrigerados³ pois possuem um TCR de $3\%/C^{-1}$ na região semicondutora sem necessitarem de criogenia. Como consequência, os bolômetros empregando o VO_2 como material sensor demonstram ser uma solução promissora para sistemas de imagem infravermelho em aplicações variadas.

1.1 Caracterização de um Bolômetro

O principal critério quando se projeta um bolômetro é avaliar seu desempenho baseando-se em algumas figuras de mérito. O desempenho de um bolômetro é calculado em dois passos. No primeiro passo, deve-se determinar como a temperatura do sensor se eleva quando o mesmo é submetido a uma radiação incidente. No segundo passo, este aumento de temperatura é utilizado para determinar a variação de resistência elétrica (ΔR) do sensor e, por conseguinte, o correspondente sinal elétrico proporcional a esta variação. O efeito da radiação incidente sobre a temperatura do material sensor pode ser descrito mediante uma equação de equilíbrio termodinâmico (Brandão et al., 2001b):

$$C \frac{dT}{dt} = P_e + P_r - G(T - T_s) \quad (1.1)$$

Na qual, P_e é a potência elétrica gerada por efeito Joule devido à polarização do sensor, P_r é a potência absorvida pela área do filme exposta a radiação eletromagnética, T_s é a temperatura do substrato, C é a capacidade térmica do sensor e G é a condutância térmica entre o sensor e o substrato.

O termo $G(T - T_s)$, na equação (1.1), representa a perda de calor para o substrato, enquanto a potência P_e está relacionada com o modo de operação do bolômetro: modo de corrente constante, em que $P_e = I_p^2 R$ e I_p é a corrente de polarização, e modo de tensão constante, no qual $P_e = V_p^2 / R$, sendo V_p a tensão de polarização. Para analisar o funcionamento do

²O coeficiente de variação da resistência com a temperatura é a figura de mérito mais significativa na indicação da sensibilidade de um bolômetro.

³Em bolômetros não refrigerados usualmente um módulo termoeletrico é utilizado para possibilitar a polarização do detector em uma temperatura fixa. Os bolômetros baseados em VO_2 são geralmente polarizados na região semicondutora, em torno de $25^\circ C$, fora da região de transição.

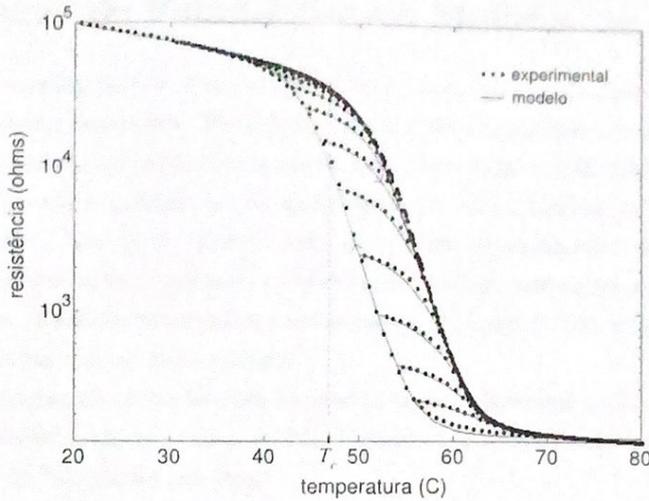


Figura 1.1: Curvas experimentais decrescentes de primeira ordem $R \times T$ (círculos sólidos) as curvas correspondentes obtidas através do modelo L^2P (linhas contínuas), com $w = 6,5^{\circ}C$, $T_c = 47,6^{\circ}C$, $\beta = 0,2^{\circ}C^{-1}$ e $\gamma = 0,85$. Tais valores são os parâmetros de referência (PR) a serem apresentados no capítulo 5. Fonte: Almeida (2003)

bolômetro, utilizando a equação (1.1), é necessário conhecer a dependência da resistência do sensor com a temperatura.

Uma curva característica $R \times T$ do filme de VO_2 é exibida na Figura 1.1. Nesta curva, observa-se o efeito de transição de fase semicondutor-metal na resistência do filme, que sempre exibe o fenômeno de histerese. A histerese está ligada à memória de uma propriedade de um sistema ou material, cujo comportamento independe da taxa de variação da excitação com o tempo. Isto implica que a curva de histerese é estável, apresentando persistência temporal.

Em temperaturas abaixo de $30^{\circ}C$, o filme de VO_2 apresenta características de um semicondutor, enquanto que em temperaturas superiores a $60^{\circ}C$, o mesmo se comporta como um metal. Nas regiões semicondutora e metálica, a variação na resistência do filme é completamente reversível. Durante a transição de fase, mais acentuadamente entre $35^{\circ}C$ e $65^{\circ}C$, os microcristais de VO_2 exibem transformação cristalográfica e a resistência do filme se altera em três ordens de grandeza. Isto resulta em valores elevados de TCR, tipicamente em torno de $60\%^{\circ}C^{-1}$. Este valor é significativamente maior do que os $3\%^{\circ}C^{-1}$ observado na região semicondutora.

1.2 Extração de Parâmetros em Modelos de Histerese

Transformadores e indutores são componentes essenciais em uma grande variedade de aplicações em eletrônica e telecomunicações. Modelagens precisas do comportamento destes dispositivos em circuitos são fundamentais para prever seu desempenho (Wilson et al., 2001). Um modelo de material magnético muito utilizado é o da aproximação de Jiles-Atherton (Wilson et al., 2002). Usando este modelo, Jiles et al. (1992) mostraram como os parâmetros desta aproximação poderiam ser extraídos de um conjunto de dados experimentais, associados ao laço principal da curva de histerese. Baseada nesta mesma amostragem, Prozygy (1993) estabeleceu os efeitos da variação de parâmetros no laço principal.

Métodos de otimização aplicados para ajustar os laços de histerese de Jiles-Atherton foram implementados por Schmidt & Guldner (1996), Lederer et al. (1999) e Hernandez et al. (2000) usando a técnica de "simulated annealing".

Otimização por algoritmos genéticos, pioneiramente desenvolvidos por Holland (1975) e Goldberg (1986), forneceram uma alternativa para otimização, especialmente quando se considera a complexidade do ajuste de vários laços simultaneamente (Wilson et al., 2002), sendo esta metodologia utilizada anteriormente por Wilson et al. (2001). Neste último trabalho, foi proposto ainda uma modificação em k , a energia média necessária para quebrar o confinamento dos domínios, o qual foi concebido originalmente como uma função linear do campo excitador H ou da densidade de fluxo. A curva de histerese do modelo ajustou-se melhor aos dados experimentais quando k tem a forma de uma função gaussiana do tipo $k = k_0 e^{-\frac{H^2}{2\sigma^2}}$, sendo k_0 um valor qualquer. Esta função de k possui a vantagem de não apresentar descontinuidade ao redor de zero quando o campo excitador muda sua orientação. Outra tentativa de estimação de parâmetros do modelo de Jiles-Atherton foi efetuada por Salvini & Fulginei (2002). Neste, foram obtidos os cinco parâmetros que descrevem o modelo (M_s , a , α_L , k e c) das equações (2.3) e (2.7) e considerados como funções da frequência do campo excitador variável $H(t)$. A estimação foi executada a partir de uma combinação de redes neurais, para estimação inicial dos parâmetros, e algoritmos genéticos para seu refinamento. A otimização genética também foi empregada por Sena et al. (2000) para a estimação dos parâmetros do modelo de Presaich.

Nas otimizações feitas por Schmidt & Guldner (1996) e Lederer et al. (1999) houve a restrição da amostra ao laço principal da curva de histerese. Esta limitação na amostragem pode ser inadequada quando o parâmetro a ser obtido indiretamente, tal como a perda de energia, necessitar de uma modelagem muito precisa; quando o modelo é utilizado em condições variadas de operação ou a partir de dimensões variadas de laços. Wilson et al. (2001) demonstraram que atribuindo pesos aos parâmetros do modelo de Jiles-Atherton e uma função objetivo de mínimos quadrados poderia resultar em melhorias nos resultados extraídos. Também, através de uma amostragem que envolvia múltiplos laços, os parâmetros do modelo foram obtidos em

uma grande variedade de condições de operação. Tal metodologia foi aplicada por Almeida (2003), para estimação dos parâmetros do modelo L^2P para o filme fino de VO_2 , se valendo apenas de inspeção de curvas e algebrismo. Mais recentemente, Betzler & Almeida (2003) e Betzler et al. (2004) obtiveram parâmetros do modelo algébrico valendo-se de um algoritmo genético e métodos de busca linear. Tais resultados são apresentados mais detalhadamente no quinto capítulo desta dissertação.

1.3 Objetivos da Investigação

A compreensão das não-linearidades na relação de histerese $R \times T$ se apresenta como sendo fundamental na análise de qualquer aplicação que utiliza filmes finos de VO_2 na região de transição, especialmente em microbolômetros. Todavia, ainda não estão disponíveis métodos detalhados de estimação de parâmetros para o modelo L^2P . O principal objetivo desta investigação é estudar comparativamente o desempenho dos métodos de otimização Levenberg-Marquardt, Gauss-Newton e Algoritmos Genéticos na estimação de parâmetros do modelo algébrico.

1.4 Organização do Texto

Abaixo será apresentada de maneira sucinta como é organizada esta dissertação:

No Capítulo 2 é conceituado o fenômeno de histerese e apresentados os modelos L^2P , Jiles-Atherton e Preisach.

No Capítulo 3 são apresentados os fundamentos e métodos determinísticos de otimização com ênfase nos utilizados neste trabalho, com enfoque em suas principais características.

No Capítulo 4 são apresentadas a origem e fundamentos dos algoritmos genéticos e uma descrição do programa de otimização genética utilizado nesta dissertação.

No Capítulo 5 é apresentada a formulação do problema de otimização do modelo L^2P para o VO_2 , envolvendo a determinação dos parâmetros T_c , β , γ e w . São também apresentados os parâmetros do modelo L^2P obtidos das iterações com os métodos de otimização aplicados.

No Capítulo 6 são apresentadas algumas considerações em relação aos resultados obtidos no Capítulo 5. Também são apresentadas perspectivas futuras deste trabalho.

Capítulo 2

O Fenômeno de Histerese

2.1 Introdução

Quando uma barra de material ferromagnético, como uma amostra de ferro, níquel, cobalto ou suas ligas, é submetido a um campo magnético aplicado, B_{ap} , os momentos de dipolo magnéticos atômicos da amostra tendem se alinhar a B_{ap} (Figura 2.1). Na Figura 2.2, observa-se que se a magnetização total da barra for inicialmente nula e B_{ap} aumentar gradualmente, o campo B no material segue a linha OP_1 . O achatamento da curva próximo ao ponto P_1 indica que a magnetização M do material está próximo do valor de saturação M_s , no qual todos os dipolos estão alinhados. Após a saturação, B apresenta um incremento apenas em função do aumento de B_{ap} . Se B_{ap} sofre uma redução, a partir do ponto P_1 , a magnetização da barra diminui, porém mais lentamente. Como o deslocamento das fronteiras dos domínios de um material ferromagnético não é totalmente reversível, a barra permanece magnetizada mesmo quando B_{ap} é totalmente removido. Este fenômeno é conhecido como histerese que vem da palavra grega "hysteros", que significa atrás ou posterior (Tipler, 2000). Segundo Visintin (1994), o termo histerese está ligado com a memória que uma certa propriedade, que descreve um sistema ou material, possui e cujo comportamento independe de sua taxa de variação temporal de excitação. Isto implicaria que a curva de histerese é estável e sua memória possui persistência temporal. Observa-se, entretanto, em certos materiais que apresentam histerese originárias de fenômenos variados como plasticidade, ferromagnetismo e ferroeletricidade, que este efeito não é puramente independente da taxa de excitação. De fato, existe um efeito viscoso ou inercial nos fenômenos de histerese (Ohta et al., 1991). Este efeito de dependência se torna muito pequeno quando a taxa de excitação tende a zero. Deste modo, neste caso, a histerese pode ser considerada como independe da taxa.

A histerese também está relacionada com a forma e desenvolvimento de laços principais e parciais e a forma, existência e acomodação em laços menores (Torres, 1994).

Diversos modelos foram criados para descrever a histerese, cada um deles destinado a des-

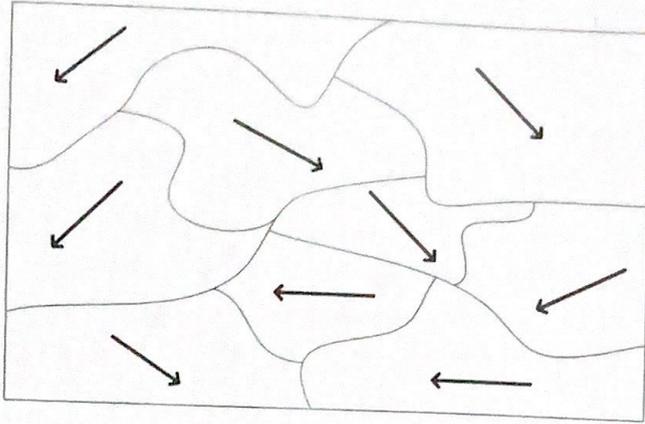


Figura 2.1: Diagrama esquemático dos domínios magnéticos em um material ferromagnético. No interior de um domínio, todos os dipolos magnéticos estão alinhados, mas como a direção de alinhamento varia de domínio para domínio, o momento magnético total da amostra pode ser pequeno ou mesmo nulo. Um pequeno campo magnético externo faz aumentar o tamanho dos domínios alinhados no sentido do campo, o que resulta em um aumento do momento magnético total.

crever o fenômeno em uma certa classe de materiais. O modelo de Prandtl-Reuss permite representar a histerese associada a diversos fenômenos de atrito e viscosidade, sendo um modelo contínuo dos mais elementares (Visintin, 1994). O modelo de Preisach, originalmente proposto para descrever o ferromagnetismo (Torres, 1994; Bertotti, 1999), é capaz de caracterizar a histerese causada pelos mais variados fenômenos (Brokate & Sprekels, 1996), como a presente em certas rochas (Fabian & Von Döbeneck, 1997).

O modelo de Preisach foi implementado numericamente por Mayergoyz (1991) e aplicado em dados experimentais de materiais ferromagnéticos. Este trabalho contribuiu para a aceitação do modelo e demonstrou sua aplicabilidade em outros fenômenos principais geradores da histerese. O modelo de Krasnoselskii-Pokrovskii, ou K-P, é uma generalização do modelo de Preisach, através da introdução de um operador¹ contínuo em contraste ao operador elementar descontínuo do último (Krasnoselskii et al., 1989).

Um modelo diferencial para a histerese magnética, baseado em considerações energéticas, foi proposto por Jiles & Atherton (1983), o qual possui alguma semelhança matemática com o modelo proposto por Duhem (Brokate & Sprekels, 1996).

Em relação aos fenômenos que originam a histerese existem a magnetização de rotação, os movimentos das paredes e nucleação (Bertotti, 1999). Estes mecanismos estão presentes na mai-

¹Um operador é um ente matemático que correlaciona a entrada excitadora x e sua resposta y , na forma $y = H[x]$

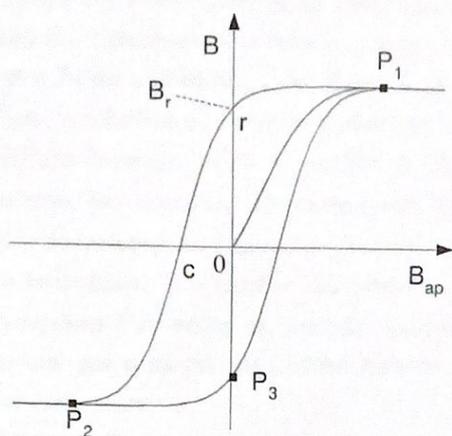


Figura 2.2: Gráfico de B ou M em função do campo aplicado B_{ap} ou H . A curva fechada é conhecida como curva de histerese. O campo B_r é chamado de campo remanente, no qual a amostra de material ferromagnético tornou-se um ímã permanente.

oria dos materiais magnéticos, em proporções diferentes. O isolamento e ponderação do efeito de cada um destes mecanismos, seria o ideal para a construção de um modelo para descrição da histerese. Apesar desta superposição de fenômenos geradores de histerese é possível construir um modelo correlacionado-o com algum mecanismo específico (Jiles & Atherton, 1983). Por outro lado, as expressões matemáticas dos modelos são desenvolvidas para representar de maneira adequada o comportamento associado aos dados experimentais, sem qualquer correlação com os princípios físicos envolvidos (Hodgdon, 1988; Mayergoyz, 1991). Modificações podem ser efetuadas nos modelos anteriores, com base em considerações físicas, visando torná-los mais precisos (Torres, 1991; Basso & Bertotti, 1996; Basso et al., 1998).

Matematicamente, os modelos de histerese podem ser classificados em duas classes: o modelo de Duhem (1897) e Presaich (1935). O modelo de Duhem possui poucas referências nos periódicos de engenharia, bem como em ciência em geral. Tal ausência pode ser justificada pela natureza do operador de histerese na forma $y' = f_1(x, y)(x')^+ - f_2(x, y)(x')^-$, no qual $f_1(x, y)$ e $f_2(x, y)$ são funções arbitrárias. Tais funções não definem a histerese associado a algum mecanismo específico, tampouco é fornecida um metodologia para obtenção destas relações. Deste modo, pode-se concluir que o modelo de Duhem fornece apenas uma estrutura matemática geral para descrever a histerese.

O modelo de Presaich foi adaptado para a descrição da histerese do VO_2 por Almeida (2003), com a criação de um operador elementar específico $\gamma_{\alpha, \beta}[H]$ e de uma função de distribuição $\mu(\alpha, \beta)$. Esta adaptação teve uma concordância razoavelmente precisa com os dados experimentais na região de transição condutor-semicondutor, no centro da curva de histerese. Entretanto para temperaturas menores que 35°C , o modelo não é adequado para descrição do fenômeno. Isto pode ser justificado pelo fato da curva de histerese da característica $R \times T$ do VO_2 ser assimétrica. Como resultado, o modelo de Presaich adaptado pode ser usado para uma avaliação do desempenho de um eventual sensor apenas na região de transição. Isto implica que o modelo de Preisach não é totalmente adequado para a descrição da histerese no filme fino de VO_2 .

2.2 Modelos de Histerese

O estudo teórico do desempenho de um bolômetro, baseado em VO_2 , requer um modelo $R = H[T]$, independente da taxa², que descreva a relação histerética na característica $R \times T$. Este modelo deve ser capaz de descrever as regiões nas quais o filme comporta-se apenas como semicondutor ou somente como metal, bem como deve ser capaz também de descrever a região

²Quando a escala de tempo de excitação de um sistema muda sem produzir mudanças na resposta do mesmo, diz-se que este sistema tem comportamento independente da taxa. Isto é, seja a excitação uma função do tempo t , escalonado por a , definida como $T = y(\frac{t}{a})$ é invariante qualquer que seja o valor de a .

de transição. Por exemplo, para o caso de um bolômetro polarizado em tensão constante na região de transição, a equação (1.1) pode ser escrita na forma:

$$C \frac{dT}{dt} = \frac{V_p^2}{H[T]} + P_r - G(T - T_a) \quad (2.1)$$

O acoplamento de um modelo de histerese $H[T]$ com um sistema dinâmico, descrito por uma equação diferencial, constitui um problema ainda em aberto do ponto de vista matemático (Brokate & Sprekels, 1996).

A análise da dinâmica de um microbolômetro, cujo comportamento é descrito por (2.1) para $H[T]$ sem histerese, demonstra a existência de uma diversidade considerável no comportamento do mesmo (Brandão et al., 2001a). Isto sugere que a utilização de um sensor com histerese na equação (2.1) pode resultar em comportamentos complexos no plano de fase $R - T$. Esta complexidade, conforme observado experimentalmente em Almeida et al. (2000), está associada a evidências de comportamentos caóticos exibidos pelo filme fino de VO_2 .

Os principais modelos de histerese existentes foram originalmente desenvolvidos para o magnetismo (Bertotti, 1999). O modelo de histerese proposto em Jiles & Atherton (1983) é o modelo físico mais difundido para representação da histerese em materiais ferromagnéticos. O modelo clássico de Preisach foi implementado numericamente por Mayergoyz (1991), sendo este trabalho o responsável por tornar o modelo aceito, em função de conectar esta aproximação do problema a dados experimentais de materiais ferromagnéticos. Este modelo foi adaptado para descrever a histerese em filme finos de VO_2 (Almeida, 2003). Devido a forte assimetria na característica $R \times T$ do VO_2 , o modelo de Preisach, concebido originalmente para descrever histereses com simetria, apresenta discrepâncias em relação aos resultados experimentais para este tipo de aplicação. Isto causa algumas dificuldades na obtenção dos parâmetros do modelo, o que suscitou investigações adicionais nas quais, o método numérico obtido em Almeida et al. (2000) foi adaptado em Sena et al. (2000) para estimação de parâmetros do modelo de Preisach.

Para superar a existência de discrepâncias observadas entre as características $R \times T$ experimentais e as obtidas a partir do modelo de Preisach, foi proposta por Almeida (2003) uma nova estrutura de modelo de histerese, inicialmente desenvolvida para a histerese em materiais ferromagnéticos. Este novo modelo baseia-se na utilização do laço principal de histerese como um conjunto invariante, que é definido como sendo a referência para a descrição da histerese. Outra propriedade invariante utilizada pelo modelo é que todas as trajetórias da curva de histerese estão confinadas ao laço principal e sempre tendem a se aproximar do mesmo com o aumento do campo magnético.

Utilizando basicamente as mesmas formulações apresentadas em Almeida et al. (2001) e em Almeida et al. (2003), a histerese nas características óticas de transmitância e refletância

do filme de VO_2 foram descritas por Fontana et al (2001) e Fontana (2001), respectivamente, com pequenos erros de ajuste. Esta nova proposta de modelo foi estendida e modificada para descrever as características $R \times T$ do filme de VO_2 . Entretanto, esta extensão não foi direta, requerendo a interpretação da transição do filme mediante a teoria da percolação (Stauffer & Aharony, 1994), e respectiva descrição matemática em termos da teoria eletromagnética do meio efetivo (Effective Medium Approximation - EMA³; (Noh & Song, 1991)). Este novo modelo de histerese, o L^2P (Limiting Loop Proximity) vem sendo estudado atualmente quanto à sua generalidade em descrever outros fenômenos de histerese. Não obstante as dificuldades associadas à utilização do modelo de Preisach e os resultados satisfatórios obtidos (Almeida et al., 2002), estão sendo realizados estudos para o desenvolvimento de novos operadores elementares para descrever a histerese. Tais operadores são utilizados para caracterizar a histerese originária de fenômenos tão variados como a decorrente da mudança de tensão em um cabo atuador de um servomotor (Galinaitis & Poling, 2003).

2.2.1 O Modelo de Jiles-Atherton

Um modelo de material magnético amplamente utilizado é o da aproximação de Jiles-Atherton (Jiles & Atherton, 1984; 1986). Baseado em considerações termodinâmicas, nestes trabalhos foi proposto que o processo de magnetização, ocasionado pela ação de um campo externo, ocorre através da movimentação das paredes entre os domínios (Almeida, 2003). Esta movimentação das paredes dos domínios magnéticos é impedida pela ocorrência de defeitos nos sólidos. Isto resulta no confinamento dos limites dos domínios aos defeitos até que seja fornecida energia externa suficiente para quebrar o confinamento. Esta energia pode ser originária do campo magnético excitador. O campo externo provoca o crescimento de um certo domínio particular em detrimento de outros, efetuando um reajuste na estrutura total dos domínios, introduzindo um elevado nível de complexidade nestas estruturas. Foi postulado ainda que o processo de magnetização é o resultado médio destas interações, ligando as irregularidades na movimentação dos domínios com as perdas energéticas no processo de magnetização.

Este modelo tem como hipótese que a magnetização M é composta por um componente reversível M_{rev} e um irreversível M_{irr} de modo que:

$$M = M_{rev} + M_{irr} \quad (2.2)$$

O componente reversível está ligado diretamente com a curva de Langevin (Almeida, 2003).

³EMA trata os comportamentos óticos e elétrico do filme em relação à fração volumétrica dos microcristais em cada uma das fases semicondutora ou metálica

$$M_{rev} = M_s \left[\cot g \left(\frac{H + \alpha_L M}{a} \right) - \frac{a}{H + \alpha_L M} \right] \quad (2.3)$$

Na qual, M_s é a magnetização de saturação, a e α_L são constantes que dependem do material e H o campo magnético externo aplicado.

A razão entre as componentes reversível e irreversíveis de M é considerada constante para qualquer valor no plano de fase $H - M$, sendo a magnetização total dada por:

$$M = (1 - c)M_{irr} + cM_{rev} \quad (2.4)$$

Sendo a razão c ligada ao tipo de material e cujos limites são $0 < c < 1$.

Considerando as perdas energéticas no processo de magnetização, a equação da susceptibilidade diferencial irreversível é obtida como sendo:

$$\frac{dM_{irr}}{dH} = \frac{M_{rev} - M_{irr}}{k\delta - \alpha(M_{rev} - M_{irr})} \quad (2.5)$$

sendo, k a energia média requerida para quebrar o confinamento dos domínios. Esta equação é somente válida para $k \neq 0$ e

$$k\delta - \alpha(M_{rev} - M_{irr}) \neq 0 \quad (2.6)$$

A partir de (2.4), em associação com as considerações posteriores apresentadas neste tópico, se obtém a equação final do modelo de Jiles-Atherton (Jiles et al., 1992):

$$\frac{dM}{dH} = (1 - c) \frac{M_{rev} - M_{irr}}{k\delta - \alpha(M_{rev} - M_{irr})} + c \frac{dM_{rev}}{dH} \quad (2.7)$$

2.2.2 O Modelo de Preisach

O modelo de Preisach é largamente utilizado para a descrição da histerese magnética, sendo os parâmetros que descrevem este modelo correlacionados com os de Jiles-Atherton (Dupré et al., 1999).

Na Figura 2.3 (a) é apresentada a forma da curva de histerese ferromagnética, sendo sua principal característica a simetria com relação a origem. Esta simetria também é observada no operador elementar que descreve a histerese $\gamma_{\alpha,\beta}[H]$ (figura 2.3 (b)). O correspondente triângulo de Preisach para a histerese magnética é mostrado na Figura 2.3 (c).

O modelo clássico de Preisach é definido como uma superposição linear de operadores elementares ponderados pela função $\mu(\rho_j)$ na forma

$$y(t) = \lim_{N \rightarrow \infty} \sum_{j=1}^N \mu(\rho_j) \gamma_{\rho_j}[\gamma_0, x(t)], \quad (2.8)$$

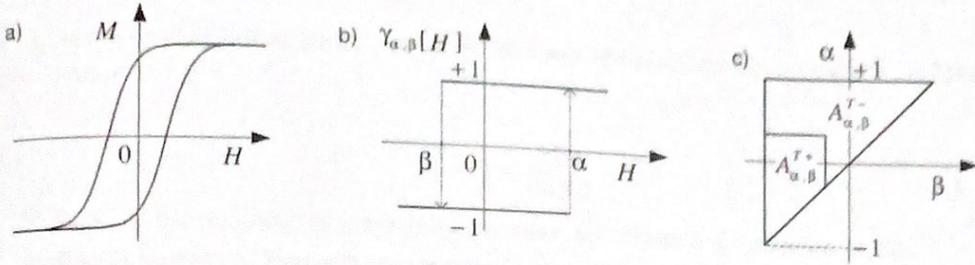


Figura 2.3: (a) Curva característica típica da histerese de materiais ferromagnéticos. (b) Operador de elemento de Preisach para materiais magnéticos e (c) o correspondente triângulo de Preisach. Fonte: Almeida (2003)

Tal modelo, na forma discreta, pode ser escrito na forma contínua de modo que (Almeida, 2003):

$$y = \iint_{A_\rho} \mu(\rho_j) \gamma_\rho[x] d\rho \quad (2.9)$$

no qual, A_ρ é um semiplano definido com a região onde $\rho^- < \rho^+$.

Considerando o sistema de coordenadas (α, β) , o modelo descrito por (2.9) para materiais magnéticos assume a forma

$$M = \iint_{A_{\alpha,\beta}^{T+}} \mu(\alpha, \beta) \gamma_{\alpha,\beta}[H] d\alpha d\beta + \iint_{A_{\alpha,\beta}^{T-}} \mu(\alpha, \beta) \gamma_{\alpha,\beta}[H] d\alpha d\beta \quad (2.10)$$

Sendo,

$$A_{\alpha,\beta}^{T+} \triangleq \{(\alpha, \beta) \in A_{\alpha,\beta}^T | \gamma_{\alpha,\beta}[H] = +1\} \quad (2.11)$$

$$A_{\alpha,\beta}^{T-} \triangleq \{(\alpha, \beta) \in A_{\alpha,\beta}^T | \gamma_{\alpha,\beta}[H] = -1\} \quad (2.12)$$

As regiões descritas pelas equações do conjunto (2.11) e (2.12) são aquelas onde os operadores elementares estão nos estados +1 e -1, respectivamente. Estas regiões compõem o triângulo de Preisach como $A_{\alpha,\beta}^T = A_{\alpha,\beta}^{T+} \cup A_{\alpha,\beta}^{T-}$. Deste modo, utilizando (2.11) e (2.12), o modelo (2.10) se reduz a:

$$M = \iint_{A_{\alpha,\beta}^{T+}} \mu(\alpha, \beta) d\alpha d\beta - \iint_{A_{\alpha,\beta}^{T-}} \mu(\alpha, \beta) d\alpha d\beta \quad (2.13)$$

A função de distribuição $\mu(\alpha, \beta)$ pode ser determinada utilizando-se a definição de Mayergoyz (1991):

$$\mu(\alpha, \beta) = \frac{\partial^2 M}{\partial \alpha \partial \beta} \quad (2.14)$$

Na qual, M é o conjunto de curvas experimentais decrescentes de primeira ordem.

Apesar do modelo de Preisach ser considerado uma ferramenta genérica para descrição da histerese, este se limita a formas de curvas similares a da Figura 2.3 (a). Tal modelo não é capaz de descrever qualquer tipo de histerese assimétrica. Neste caso, a simetria se refere sempre ao centro da curva de histerese (Almeida, 2003).

2.3 Resistência Elétrica nos Filmes Finos de VO_2

Diversos modelos foram propostos para descrever a histerese. Cada um destes foi formulado para caracterizar o fenômeno em um certo conjunto de materiais. O modelo Preisach assim como o de Jiles-Atherton associam propriedades físicas do material, tais como a magnetização total e o campo magnético excitador H , através do uso de equações integro-diferenciais (Almeida et al., 2003b). Os modelos de Jiles-Atherton e Preisach diferem pela memória que cada um possui em relação à excitação. O modelo de Jiles-Atherton inclui informação de H apenas para o estado atual do material enquanto que o de Preisach, inclui os valores extremos passados de H .

A adaptação do modelo de Preisach para descrição da histerese em filmes finos de VO_2 descreve parcialmente a característica $R \times T$ (Almeida et al., 2001). Isto pode ser justificado porque esta modelagem descreve o comportamento experimental observado sem correlacioná-lo, em profundidade, com os mecanismos físicos que se manifestam especificamente no material analisado. Tal tipo de modelo é denominado de fenomenológico ou "caixa preta". Modificações nos modelos "caixa pretas" existentes podem ser feitas visando torná-los mais precisos, através da correlação com a física microscópica do material durante o fenômeno. Baseado nisso, foi proposto um modelo "caixa cinza", na qual parte da estrutura do modelo esta correlacionada com a física do material sensor (Almeida et al., 2003b). Para tal, a resistência R é descrita em termos da teoria EMA, que utiliza a fração volumétrica para obter a condutividade efetiva do filme.

O filme fino de VO_2 é um meio formado por um grande número de microcristais, parte no estado condutor (metálico) e parte no estado semicondutor. As propriedades ópticas e elétricas do filme dependem da proporção ou fração volumétrica de cada um dos dois estados possíveis. No instante em que a fração volumétrica atinge um certo valor crítico para um dos estados

ocorre o fenômeno da percolação. A percolação se caracteriza como o surgimento de pequenos aglomerados microscópicos de cristais no estado metálico no substrato semiconductor. Com o aumento do número de aglomerados metálicos, surgem pontes condutoras entre os aglomerados, o que propicia um primeiro caminho de condução pelo filme. Neste estágio é atingido o chamado limiar de percolação. Este tipo de transição de fase térmica pode ser descrito pela utilização do EMA.

2.4 Histerese no Filme Fino de VO_2

O modelo obtido da aplicação dos conceitos expostos no tópico (2.3) foi denominado de "Limiting Loop Proximity" (L^2P). Este modelo é apresentado em termos de uma expressão algébrica, que possui dependência com apenas quatro parâmetros, e que dispensa a solução de equações diferenciais ou de integração dupla, usuais nos modelos de Preisach e Jiles (Almeida et al., 2003a). Em termos da sequência de memória de H , o modelo L^2P possui uma memória mais curta do que o de Preisach e relativamente mais representativa que o de Jiles-Atherton.

2.4.1 Evolução da fração volumétrica

Na região semicondutora, todos os microcristais de VO_2 comportam-se como um único semiconductor intrínseco (Liddiard, 1986), cuja resistência do filme pode ser dada por:

$$R_s = R_0 \exp\left(\frac{E_a}{kT}\right) \quad (2.15)$$

A partir de (2.15), o coeficiente de variação de resistência com a temperatura (TCR) pode ser escrito como:

$$TCR = \frac{1}{R_s} \frac{dR_s}{dT} = -\frac{E_a}{kT^2} \quad (2.16)$$

Nas equações anteriores, T é a temperatura do filme (Kelvin); R_0 é a resistência para a T tendendo ao infinito como pode ser observado na figura 2.4. Nesta, E_a é a energia de ativação, cujo valor é de 0,22eV (Jerominek et al., 1996) e $k = 8,62 \times 10^{-5}$ eV/K é a constante de Boltzmann. O valor de E_a/k , obtido a partir dos dados experimentais, para um $TCR = -3\% ^\circ C^{-1}$, é condizente com aquele obtido a partir do E_a de Jerominek et al (1996), possuindo um discrepância relativa percentual inferior a 0,1%. O valor experimental, obtido em Almeida (2003) para E_a/k , é de 2553 K.

Microcristais individuais de VO_2 passam para a característica metálica em temperaturas maiores de 68 °C. Deste modo, não é possível medir diretamente o valor de R_0 . Para tal, foi considerado o valor de $T = 20$ °C para definir esta constante a partir da equação (2.15). Para

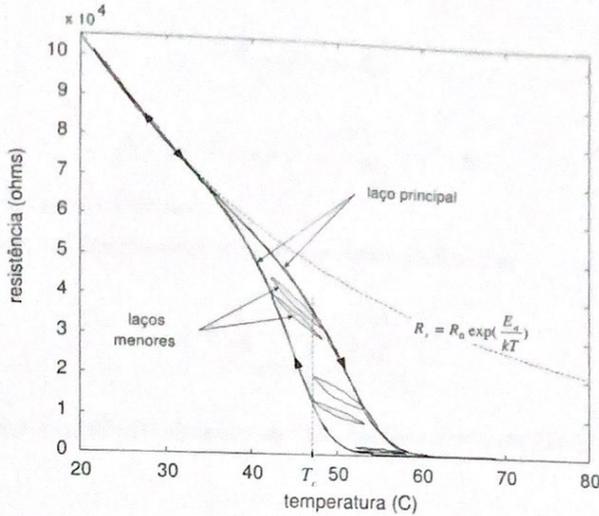


Figura 2.4: Características experimentais $R \times T$ do filme fino de dióxido de vanádio, incluindo laços menores. Fonte: Almeida (2003)

o experimento, $R_0 = 17\Omega$ e nesta temperatura aproximadamente todos os cristais estão na fase semicondutora. A resistência do filme na região semicondutora pode ser escrita, então, como:

$$R_s = 17 \exp\left(\frac{2553}{T}\right) \quad (2.17)$$

Na região metálica ($T > 80 \text{ }^\circ\text{C}$), todos os microcristais se comportam com um metal e a resistência do filme apresenta um valor aproximadamente constante. Para $T = 80 \text{ }^\circ\text{C}$, esta possui um valor $R_m = 140\Omega$.

Em uma dada temperatura, a resistência do filme é devida à constituição de microcristais no estado condutor e semicondutor. Deste modo, a resistividade pode ser escrita como:

$$\rho = g\rho_s + (1 - g)\rho_m \quad (2.18)$$

Na qual g é fração de microcristais na fase semicondutora e ρ_s e ρ_m são, respectivamente, as resistividades nas fases semicondutora e metálica.

A resistência no filme pode ser representada como:

$$R = gR_s + (1 - g)R_m \quad (2.19)$$

Para os filmes finos de VO_2 , a resistência na fase semicondutora é muito maior do que a da fase condutora. Assim substituindo a equação (2.16) em (2.19) e levando em consideração a relação entre a magnitude das resistências nas fases semicondutora e condutora do filme; R pode ser escrita como:

$$R = gR_s + R_m \quad (2.20)$$

$$R(T) = R_0 \exp\left(\frac{2553}{T + 273}\right) \bar{g} + R_m. \quad (2.21)$$

onde, T é dado em graus Celsius.

A fração volumétrica experimental \bar{g} pode ser dada, então, por

$$\bar{g} = \frac{1}{17} \frac{\bar{R} - 140}{\exp\left(\frac{2553}{T + 273}\right)}, \quad (2.22)$$

Na qual, \bar{R} o valor experimentalmente medido da resistência do filme para uma certa temperatura T .

2.5 Modelagem da Histerese na Característica $g \times T$

Durante a variação da temperatura do VO_2 , a temperatura na qual ocorre a transição de fase de um microcristal é significativamente afetada pela tensão interfacial entre os microcristais vizinhos (Potapov et al., 1998), representando a perda de energia ou a componente irreversível da transição. Em transições martensíticas (Otsuka & Wayman, 1998), como as que ocorrem neste material, podem ser identificados três tipos de componentes da energia. Duas reversíveis são a contribuição associada a mudanças cristalográficas ou estruturais/elásticas. A terceira, irreversível, é atribuída ao atrito na interface dos microcristais. Este atrito é a causa primária da ocorrência da histerese e resulta que o filme possua múltiplos valores de resistência para um único valor de temperatura.

Baseado no comportamento das curvas de $g \times T$, em Almeida (2003) foi formulada uma função para descrever a fração volumétrica g :

$$g = F(T) = \frac{1}{2} + \frac{1}{2} \tanh \beta \left(\frac{\delta w}{2} + T_c - (T + T_{pr} P(x)) \right) \quad (2.23)$$

Na qual, T_c é a temperatura crítica ou limiar de percolação, a temperatura na qual $g = 0,5$. No caso de uma curva característica com histerese existem dois valores para T_c : um para a curva ascendente e outro para a descendente principal. É assumido que T_c é o valor médio para $g = 0,5$ no centro da curva de histerese (Figura 5.2). w é a largura da curva de histerese. β é relacionado com $\frac{dg}{dT}$ em T_c . δ é um operador que assume valor +1, para $\frac{dT}{dt} > 0$, ou -1, para $\frac{dT}{dt} < 0$.

No modelo $L^2 P$, as curvas de histerese são divididas em uma componente ascendente e outro descendente. Nestas, δ se mantém 1 ou -1 até que se atinja uma temperatura de reversão T_r , que ocorre quando $\frac{dT}{dt}$ muda de sinal e ocorre a transição de um laço $k - 1$ para outro k . Em (2.23), a função $P(x)$ é descrita por:

$$P(x) \triangleq \frac{1}{2}(1 - \text{sen}(\gamma x))(1 + \tan(\pi^2 - 2\pi x)) \quad (2.24)$$

onde,

$$x = \frac{T - T_R}{T_{pr}} \quad (2.25)$$

T_{pr} é a temperatura de proximidade que se destina a expressar a distância do ponto de reversão (T_r, g_r) , em $k - 1$, até um ponto (T, g) na curva k .

Por sua vez,

$$T_{pr} = \frac{\delta w}{2} + T_c - \frac{1}{\beta} \text{arctanh}(2g_r - 1) - T_r \quad (2.26)$$

Na qual, g_r pode ser obtida com a aplicação de (2.22) para $T = T_r$. Para a identificação dos pontos de reversão se adotou verificar a variação do sinal de $T_{i+1} - T_i$. Deste modo, $T_r = T_i$ na mudança do sinal.

Os valores de δ , T_r e T_{pr} mudam apenas nos pontos de reversão permanecendo constantes até a transição para outro laço.

2.6 Conclusões

O fenômeno de histerese é originário dos mais variados processos. Para sua descrição foram criados modelos como os de Jiles-Atherton e Presaich, dados em função de parâmetros associados ao meio onde a histerese se manifesta. Para a estimação destes parâmetros é necessária a resolução de equações integro-diferenciais, o que exigem um maior esforço computacional e algoritmos complexos para descrição dos modelos. O modelo $L^2 P$ é descrito por relações puramente algébricas, sendo um modelo do tipo "caixa cinza", que incorpora aspectos físicos do fenômeno a uma modelagem empírica, característica dos modelos "caixa preta". Tal modelo foi utilizado por Almeida (2003) para descrever a histerese na característica $R \times T$ de um filme fino de Dióxido de Vanádio. Este modelo apresentou-se mais eficiente na descrição do fenômeno do que a adaptação da modelagem de Presaich.

Capítulo 3

Fundamentos e Métodos de Otimização

3.1 Introdução

O termo otimização pode ser definido como um procedimento numérico para determinar a "melhor" solução de um problema, usualmente relacionado com modelos que descrevem uma certa realidade física. A idéia de melhor solução está usualmente relacionada com uma função $f(\mathbf{x})$, denominada função objetivo. O ponto \mathbf{x} , no espaço n -dimensional \mathfrak{R}^n , é o vetor $[x_1, x_2, x_3, \dots, x_n]^T$ sendo x_n a componente na direção do eixo do sistema de n coordenadas. A maioria dos métodos de otimização são iterativos, gerando uma seqüência de pontos, $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$, representada por $\mathbf{x}^{(k)}$. Para tais métodos espera-se uma convergência para um ponto fixo \mathbf{x}^* , que se constitui na melhor solução do problema. No caso específico da estimação de parâmetros de um modelo matemático, a função $f(\mathbf{x})$ é definida como sendo alguma norma que possa medir a proximidade entre os resultados produzidos pelo modelo e os dados experimentais. Uma norma muito utilizada é a L_2 , definida como:

$$f(\mathbf{x}) = \|\mathbf{e}\|_2 = \sqrt{\mathbf{e}\mathbf{e}^T} \quad (3.1)$$

Sendo \mathbf{e} o vetor de erro entre o modelo e os dados experimentais (Fletcher, 2001).

3.2 Terminologia

- 1 Variáveis: As variáveis possuem valores que são alterados durante o processo de otimização, gerando uma seqüência $\mathbf{x}^{(k)}$.
- 2 Restrições: As restrições são funções de igualdade ou desigualdade que descrevem situações desejáveis ou não do projeto. Mais especificamente estas são denominadas de restrições laterais. Tais restrições limitam os valores que as variáveis podem assumir.

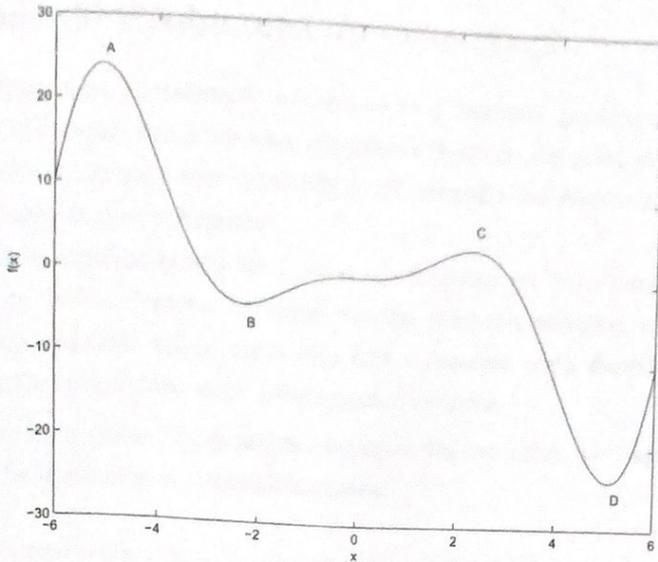


Figura 3.1: Extremos da função $x^2 \text{sen}(x)$ entre -6 e 6 radianos. O ponto C é máximo enquanto B é um mínimo, ambos locais. O máximo global ocorre em A. O mínimo global é D.

- 3 Espaço de Busca ou Região Viável (Factível): Conjunto, espaço ou região que compreende as soluções possíveis do problema a ser otimizado. É definido pelas funções ou conjunto de restrição.
- 4 Função Objetivo ou de Avaliação: Função $f(x)$ que se deseja otimizar. Pode ser constituída de um ou mais variáveis, sendo alvo respectivamente da otimização unidimensional ou multidimensional.
- 5 Ponto Ótimo: É o ponto caracterizado por um vetor x^* , constituído pelas variáveis do projeto, que minimizam a função objetivo e satisfazem as restrições.
- 6 Valor Ótimo: Valor da função objetivo no ponto ótimo ($f(x^*)$).
- 7 Solução Ótima: É o par de soluções formado pelo ponto e valor ótimos $[x^*, f(x^*)]$. Esta solução pode ainda ser:
 - 7.1 Ótimo Local: Quando o valor ótimo é restrito a uma certa vizinhança.
 - 7.2 Ótimo Global: Quando o valor é ótimo em toda região de factibilidade ou viabilidade.
 - 7.3 Infactível ou Inviável: Quando não atende a alguma das restrições impostas pelo problema.

3.3 Classes de Problemas de Otimização

Algoritmos de otimização normalmente relacionam-se a teoremas provando sua convergência (da Silva, 2001). A solução obtida por estes algoritmos depende, em geral, do ponto de partida fornecido, $\mathbf{x}^{(1)}$. Esta limitação é uma característica dos métodos matemáticos clássicos, os quais vem sendo constantemente aperfeiçoados.

Os problemas de otimização não são podem ser divididos em “com restrição” e “sem restrição” (Luenberger, 1989). Para sua resolução, existem inúmeros métodos, associados a classes de problemas, cuja descrição minucionosa esta fora do escopo desta dissertação. Nesta, são apenas apresentadas informações sobre determinados métodos .

De acordo com as características da função objetivo e das restrições, basicamente, classificam-se os problemas de otimização nas seguintes classes:

- 1 Programação Linear: Quando a função objetivo e as restrições são funções lineares das variáveis do problema.
- 2 Programação Não Linear: Quando a função objetivo ou pelo menos uma das restrições é não linear.

Gradativamente, outras classes foram acrescentadas a programação matemática com o objetivo de serem mais adequados aos problemas a serem resolvidos. Assim tem-se, por exemplo:

- 3 Programação Quadrática: Quando a função objetivo é quadrática e as restrições são lineares.

Dependendo do tipo de problema de otimização, linear ou não linear, com ou sem restrições, aplicam-se diferentes métodos de programação matemática para sua resolução.

Os métodos foram desenvolvidos em função de agrupar em classes os problemas em função da estrutura da função objetivo e das suas restrições. Quanto as restrições, os problemas podem ser separados em (da Silva, 2001):

1. Com Restrições de Igualdade
2. Com Restrições de Desigualdade
3. Com Restrições Mistas

3.3.1 Programação Linear

Pode-se definir o problema geral da programação linear da seguinte forma: Dado um conjunto de " m " desigualdades ou equações lineares em " n " variáveis, deseja-se determinar valores não negativos destas variáveis que satisfaçam e minimizem alguma função linear destas variáveis.

O problema da programação linear na forma padrão pode ser descrito pelo conjunto de equações (Luenberger, 1984):

$$\text{Minimizar } \mathbf{c}^T \mathbf{x}$$

$$\text{Sujeito a } \mathbf{Ax} = \mathbf{b}$$

$$\mathbf{x} \geq 0$$

Sendo \mathbf{x} é um vetor coluna e \mathbf{c}^T é um vetor linha, ambos com dimensão n . A segunda equação condiciona a restrição na qual a variável \mathbf{x} esta submetida. O vetor $\mathbf{x} \geq 0$ implica que cada componente de \mathbf{x} é positivo.

O método mais conhecido e largamente usado para a solução de problemas de programação linear é o método Simplex (Dantzig, 1963).

Tal conjuntos de métodos sofreu modificações em relação ao original que gerou sua versão generalizada e revisada (Luenberger, 1984).

3.3.2 Programação Não-Linear

A programação não-linear trata dos problemas no quais as restrições ou a função objetivo possuem dependência não linear com as variáveis do projeto. Os problemas se dividem em restritos ou não-restritos e unidimensionais ou multidimensionais. Os métodos de programação não-linear serão aqui divididos em:

1. Otimização sem restrições
2. Otimização com restrições

3.4 Otimização Sem Restrições

O conceito de linha é de suma importância para algoritmos de otimização, em especial, para a programação linear sem restrições. Uma linha é o conjunto de pontos tal que:

$$\mathbf{x}(\alpha) = \mathbf{x}' + \alpha \mathbf{s} \quad (3.2)$$

Válida para todo α , nos quais \mathbf{x}' é um ponto fixo ao longo da linha e \mathbf{s} é a direção da linha. Considerando inicialmente que $f(\mathbf{x})$ possui derivadas contínuas existe, então, o vetor das primeira derivadas parciais ou vetor gradiente:

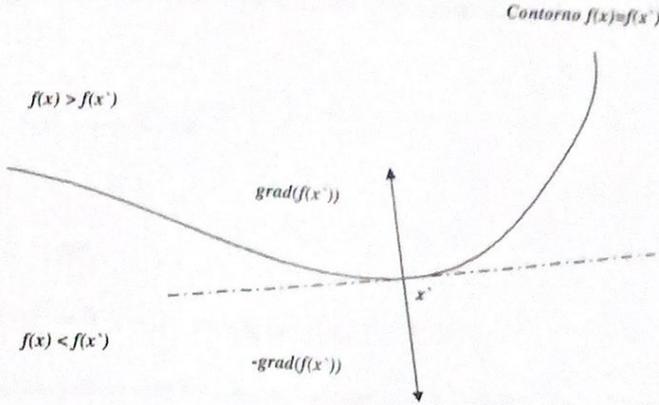


Figura 3.2: Características do vetor gradiente.

$$g(\mathbf{x}) = \nabla f(\mathbf{x}) = [\partial f/\partial x_1, \partial f/\partial x_2, \dots, \partial f/\partial x_n]^T \quad (3.3)$$

Sendo que ∇ denota o operador gradiente $[\partial/\partial x_1, \partial/\partial x_2, \dots, \partial/\partial x_n]^T$. Se $f(\mathbf{x})$ possui derivada segunda existe a matriz de segunda derivadas parciais ou matriz Hessiana, escrita na forma $G(\mathbf{x}) = \nabla^2 f(\mathbf{x})$. Nesta matriz, o i,j -ésimo elemento é representado por $\partial/\partial x_i \partial x_j$.

A inclinação de $f(x(\alpha))$, ao longo da linha definida por (3.2), é dada por:

$$\frac{df}{d\alpha} = s^T \nabla f = \nabla f^T s \quad (3.4)$$

bem como sua curvatura:

$$\frac{d^2 f}{d\alpha^2} = s^T \nabla^2 f s \quad (3.5)$$

Sendo ∇f e $\nabla^2 f$ calculados em $x(\alpha)$.

Na Figura (3.2) são apresentadas as características do vetor gradiente. Nesta figura, O vetor $\nabla f(x')$ representa a direção de maior inclinação. O sentido de crescimento e descimento é indicada pela reta perpendicular a linha tangente ao ponto x' . O sentido de crescimento do vetor gradiente indica o sentido a se seguir para a localização de um ponto de máximo.

3.5 Otimização com Restrição

Para esta situação, alguns tipos de métodos procuram transformar um problema com restrições em um irrestrito. Deste modo, é possível resolver o mesmo com a mesma metodologia empregada no tópico anterior.

Um problema de otimização com restrições pode ser formulado da seguinte forma geral:

Minimizar $f(\mathbf{x})$

Sujeito a $h_i(\mathbf{x}) = 0, \quad i = 1 \dots m$

$g_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, q$

$\mathbf{x} \in \Omega \subset R^n$

$\Gamma = \{\mathbf{x} \in \Omega \mid h_i(\mathbf{x}) = 0, i = 1 \dots m \text{ e } g_j(\mathbf{x}) \leq 0, j = 1 \dots q\}$

No qual, Γ é denominada região factível, sendo que a função $f(\mathbf{x})$ é contínua, possuindo derivadas parciais até segunda ordem. Também $h_i(\mathbf{x})$ e $g_j(\mathbf{x})$ são contínuas possuindo derivadas parciais até a primeira.

A otimização com restrição pode ser separada em três partes:

1. Problemas com Restrições de Igualdade.
2. Problemas com Restrições de Desigualdade.
3. Problemas com Restrições Mistas.

Métodos para resolução de problemas como os acima são os mais variados possíveis tais como o do Lagrangiano, Penalidades e Lagrangiano Aumentado (da Silva, 2001).

3.5.1 Busca ao Longo de uma Linha

Os algoritmos de otimização para problemas de programação não-linear sem restrições possuem, em geral, uma mesma estrutura básica que consiste em aplicar de forma iterativa a equação $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{s}^k$, onde \mathbf{x} é o vetor das variáveis ; α é o tamanho do passo e k representa a iteração corrente.

A diferença entre os diversos algoritmos de solução de problemas de otimização consiste na estratégia empregada para a determinação do vetor \mathbf{s}^k , correspondente às sucessivas direções de busca ao longo das iterações. Tal parâmetro deve minimizar $f(\mathbf{x}^{(k)} + \alpha \mathbf{s}^{(k)})$ com respeito a α . Entre os métodos existentes para resolução de problemas em programação não-linear sem restrições, alguns necessitam apenas como informação o valor da função. Outros, além do valor da função, necessitam da primeira e até da segunda derivada (Hessiana) (Fletcher, 2001). Tais métodos são denominados de ordem zero, um e dois respectivamente.

O problema irrestrito geral, neste caso, pode ser formulado como

Minimizar $f(\mathbf{x})$, $\mathbf{x} \in R^n$

As condições necessárias e suficientes para se ter um mínimo local são (Luenberger, 1984; Fletcher, 2001):

1. Condição necessária de primeira ordem: $\nabla f(\mathbf{x}^*) = 0$, \mathbf{x}^* é um ponto estacionário.
2. Condição necessária de segunda ordem: $\nabla f(\mathbf{x}^*) = 0$; $\mathbf{s}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \mathbf{s} \geq 0 \quad \forall \mathbf{s}$
No qual, \mathbf{s}^T é a transposta do vetor \mathbf{s}

As condições (1) e (2) implicam que no ponto \mathbf{x}^* , no qual passa qualquer linha, a função $f(\mathbf{x}(\alpha))$ possui inclinação zero e curvatura não negativa. Nesta última condição, a matriz $G(\mathbf{x}^*)$ é semi-definida positiva. A partir das condições necessárias pode-se derivar as condições suficientes:

3. Condição suficiente de segunda ordem: $\nabla f(\mathbf{x}^*) = 0$; $\mathbf{s}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \mathbf{s} > 0 \quad \forall \mathbf{s} \neq 0$

A condição (3) implica que no ponto de mínimo local isolado \mathbf{x}^* , $G(\mathbf{x}^*)$ é positiva definida.

Dentro deste contexto, alguns métodos para resolução de problemas de otimização sem restrição serão apresentados nos próximos tópicos:

1. Método de Newton
2. Método Quase-Newton (BFGS)
3. Método Quase-Newton (DFP)
4. Método de Gauss-Newton (GN)
5. Método de Levenberg-Marquardt (LM)

Os dois últimos métodos foram utilizados neste trabalho na otimização do modelo de histerese L^2P para o filme fino de Dióxido de Vanádio. Tais métodos foram concebidos para aproveitar de maneira produtiva a estrutura do problemas mínimos quadrados no processo de otimização. Em tal problema, parâmetros que ajustam um modelo podem ser extraídos de uma seqüência de dados experimentais.

3.6 O Problema de Mínimos Quadrados Não Linear

Para se obter M parâmetros desconhecidos α_k que descrevem um modelo não linear, como a resistência elétrica do filme de VO_2 , pode-se definir uma função objetivo tal como (Press et al., 2001):

$$\chi^2 = \sum_{i=1}^N \left[\frac{y_i - \sum_{k=1}^M a_k X_k(x_i)}{\sigma_i} \right]^2 \quad (3.6)$$

Na qual, y_i é o dado experimental, associado a variável independente x_i .

$X_k(x_i)$ é denominada de função base do problema. O somatório envolvendo a função base pode ser definida como $y(x_i, \mathbf{a})$.

σ_i é o desvio padrão ou avaliado do dado experimental i , se conhecido. Caso contrário, pode-se admitir $\sigma = 1$.

A partir de χ^2 é possível se determinar os parâmetros do modelo por minimização desta função. Com um conjunto de valores iniciais do modelo será efetuado um procedimento de busca visando modificar os valores até a obtenção do mínimo de χ^2 .

Para a função objetivo é esperado, que suficientemente próximo do mínimo, esta seja aproximada na forma quadrática:

$$\chi^2(\mathbf{a}) \approx \gamma - \mathbf{d} \cdot \mathbf{a} + \frac{1}{2} \mathbf{a} \cdot \mathbf{D} \cdot \mathbf{a} \quad (3.7)$$

Na qual \mathbf{d} é vetor de derivadas primeiras de χ^2 com relação a \mathbf{a} , cuja dimensão é M , e \mathbf{D} é a matriz de derivada segunda $M \times M$. Se esta aproximação é razoável, é possível obter uma expressão que correlacione o dado conjunto de parâmetros $\mathbf{a}_{\text{atual}}$ ao valor minimizado $\mathbf{a}_{\text{próximo}}$:

$$\mathbf{a}_{\text{próximo}} = \mathbf{a}_{\text{atual}} + \mathbf{D}^{-1} \cdot [-\nabla \chi^2(\mathbf{a}_{\text{atual}})] \quad (3.8)$$

Por outro lado, (3.7) pode ser um aproximação local ruim dependendo da superfície na qual se pretende minimizar a partir de $\mathbf{a}_{\text{atual}}$. Neste caso, é adequada a aplicação do método gradiente descendente:

$$\mathbf{a}_{\text{próximo}} = \mathbf{a}_{\text{atual}} - \text{constante} [-\nabla \chi^2(\mathbf{a}_{\text{atual}})] \quad (3.9)$$

Na qual, a *constante* em (3.9) é pequena tal que não perturbe a direção de descida do método.

Para a utilização de (3.8) e (3.9), é necessário calcular o gradiente da função χ^2 para um dado conjunto de parâmetros \mathbf{a} . O uso de (3.8) implica ainda no conhecimento da matriz \mathbf{D} , cujos elementos são derivadas segundas de χ^2 para um \mathbf{a} qualquer.

Paraa

3.6.1 Cálculo do gradiente e Hessiana

O modelo a ser ajustado é

$$y = y(x, \mathbf{a}) \tag{3.10}$$

é o modelo dado pela expressão (3.6)

O gradiente de χ^2 com respeito aos parâmetros \mathbf{a} , os quais serão zero quando a função de mérito estiver no mínimo tem componentes

$$\frac{\partial \chi^2}{\partial a_k} = -2 \sum_{i=1}^N \left[\frac{y_i - \sum_{k=1}^M a_k X_k(x_i)}{\sigma_i^2} \right] \frac{\partial y(x_i, \mathbf{a})}{\partial a_k} \tag{3.11}$$

Calculando as derivadas parciais adicionais

$$\frac{\partial^2 \chi^2}{\partial a_k \partial a_l} = 2 \sum_{i=1}^N \frac{1}{\sigma_i^2} \left[\frac{\partial y(x_i, \mathbf{a})}{\partial a_k} \frac{\partial y(x_i, \mathbf{a})}{\partial a_l} - [y_i - y(x_i, \mathbf{a})] \frac{\partial^2 y(x_i, \mathbf{a})}{\partial a_l \partial a_k} \right] \tag{3.12}$$

Se convencionamos definir os fatores de derivada segunda como

$$\beta_k \equiv -\frac{1}{2} \frac{\partial \chi^2}{\partial a_k} \qquad \alpha_{kl} \equiv \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_k \partial a_l} \tag{3.13}$$

Fazendo $[\alpha] = \frac{1}{2} \mathbf{D}$ na equação (3.12) pode ser escrita como um conjunto de equações lineares

$$\sum_{i=1}^N \alpha_{kl} \delta a_l = \beta_k \tag{3.14}$$

Este conjunto é resolvido para os incrementos δa_l os quais, somados a aproximação atual da solução, fornecem a próxima estimativa. Neste contexto dos mínimos quadrados, a matriz $[\alpha]$, igual a metade da matriz Hessiana, é chamada de matriz de curvatura.

A equação (3.9) pode ser reescrita também da seguinte forma

$$\delta a_l = \text{constante} \beta_k \tag{3.15}$$

Pode se notar que os componentes α_{kl} da matriz Hessiana (3.13) dependem ambos das primeiras e segundas derivadas da base de funções com respeito a estes parâmetros. Alguns tratamentos do problema ignoram as segundas derivadas. Tais derivadas ocorrem porque o gradiente (3.11) possui uma dependência com $\partial y / \partial a_k$, de modo que a segunda derivada contém termos envolvendo $\partial y / \partial a_l \partial a_k$.

O termo de segunda derivada pode ser desprezado quando este é zero ou desprezível se comparado com termo envolvendo a primeira derivada. Outra possibilidade para o desprezo do termo seria $[y_i - y(x_i, \mathbf{a})]$ ser pequeno. Na modelagem, este termo mede a variação aleatória do erro em cada ponto. Esta variação pode ser positiva ou negativa e pode, em geral, ser não correlata com o modelo. Entretanto, os termos de segunda derivada tendem a se cancelar quando somados sobre i .

3.7 Métodos de Otimização

3.7.1 Métodos de Otimização Sem Restrição

Existe um grande número de métodos para resolução de problemas de otimização, em especial, os problemas sem restrições. Neste tópico serão apresentados, para fins comparativos, alguns métodos para resolução deste tipo de problema: Newton, os métodos Quase-Newton BFGS (Broyden, Fletcher, Goldfarb, Shanno) e DFP (Davidon, Fletcher, Powell), além dos algoritmos de Levenberg-Marquardt (LM) e Gauss-Newton (GN).

3.7.2 Método de Newton

Considere-se o problema de otimização dado por (da Silva, 2001):

$$\text{Minimizar } f(x) \quad x \in R^n$$

Sendo $f(x)$ contínua com derivadas parciais contínuas até segunda ordem.

Expandindo a função $f(x)$ em série de Taylor em torno do ponto x^k , obtém-se (Fletcher, 2001):

$$f(x^{(k)} + \delta) \approx q^{(k)}(\delta) = f^{(k)} + g^{(k)T}\delta + \frac{1}{2}\delta^T G^{(k)}\delta \quad (3.16)$$

Onde $\delta = x - x^{(k)}$, $G^{(k)}$ é a matriz Hessiana e $q^{(k)}(\delta)$ é o resultado da aproximação ao modelo quadrático da iteração k . Neste modelo $G^{(k)}$ é definida positiva. Para a $x^{(k+1)}$, o método calcula a relação $x^{(k)} + \delta^{(k)}$, onde a correção de $\delta^{(k)}$ minimiza $q^{(k)}(\delta)$. Para tal, é necessário que a função f e sua primeira e segunda derivadas sejam disponíveis para qualquer ponto. Isto implica que os coeficientes $f^{(k)}$, $g^{(k)}$ e $G^{(k)}$, os quais definem $q^{(k)}(\delta)$, também existem. Se $q^{(k)}(\delta)$ tem um minimizador único se $G^{(k)}$ é positiva definida. Nesta condição, o método de Newton alia eficiência e rápida convergência local. $\delta^{(k)}$ é definido na condição que $\nabla q^{(k)}(\delta^{(k)}) = 0$. Uma iteração k qualquer do método de Newton pode ser escrita como

a) Resolva $G^{(k)}\delta = -g^{(k)}$ para $\delta = \delta^{(k)}$

b) Faça $x^{(k+1)} = x^{(k)} + \delta^{(k)}$

O passo (a) envolve a solução de um sistema de $n \times n$ equações lineares.

As principais características deste método são (Fletcher, 2001; da Silva, 2001):

1. Convergência quadrática ($p = 2$), se a otimização se iniciar de um ponto suficientemente próximo da solução.
2. Pelas condições suficientes de segunda ordem, para que um ponto x^* seja um mínimo local do problema, a matriz Hessiana $\nabla^2 f(x^*)$ deve ser positiva definida. Desta forma, em um ponto próximo a x^* , o método apresenta um bom comportamento.
3. Para valores iniciais distantes da solução, o método perde sua eficiência.

3.7.3 Métodos Quase-Newton

A maior desvantagem do método de Newton é a necessidade do cálculo de $G(x)$ (Fletcher, 2001). Entretanto métodos análogos ao newtoniano podem ser obtidos quando somente $g(x)$ é disponível. A aproximação mais simples é o método de Newton da diferenças finitas no qual os incrementos h_i em cada direção coordenada e_i são utilizados para estimar $G(x)$ pela diferença nos vetores gradiente. Isto é \bar{G} é aquela na qual na coluna i é dada por $(g(x^{(k)} + h_i e_i) - g^{(k)})/h_i$. \bar{G} é feita simétrica fazendo-se $\frac{1}{2}(\bar{G} + \bar{G}^T)$ e a matriz resultante substitui $G(x)$ no método de Newton.

Algumas desvantagens do método:

1. $\frac{1}{2}(\bar{G} + \bar{G}^T)$ pode não ser positiva definida, o que pode ocasionar em problema de convergência caso não se façam modificações apropriadas na aproximação de $G(x)$.
2. n cálculos de $g(x)$ são necessários para estimar $G(x)$ e um conjunto de equações lineares deve ser resolvido a cada iteração, o que aumenta o esforço computacional.

Métodos quase-newtonianos com busca linear necessitam de uma aproximação da inversa de matriz Hessiana, $G(x)^{-1}$. Tal aproximação é dada pela matriz simétrica definida positiva $H^{(k)}$. Esta matriz é atualizada em cada iteração. Para uma iteração k , a busca tem a seguinte estrutura:

1. $s^{(k)} = -H^{(k)}g^{(k)}$
2. Busca linear em $s^{(k)}$ dado por $x^{(k+1)} = x^{(k)} + \alpha^{(k)}s^{(k)}$

3. Atualize $H^{(k)} : H^{(k+1)}$

As potencialidades do método quase-newtoniano são:

- (a) Somente as primeiras derivadas são necessárias.
- (b) $H^{(k)}$ positiva definida implica na propriedade descentente mesmo $G^{(k)}$ sendo indefinida.
- (c) São necessárias $O(n^2)$ por iterações no lugar da n^3 do método Newtoniano.

A matriz inicial $H^{(1)}$ pode ser uma matriz positiva definida. Na ausência de uma melhor estimativa pode-se fazer $H^{(1)} = I$. Para atualização de $H^{(k)}$ foram concebidos diversos métodos conforme exemplificados abaixo:

3.7.4 Método Broyden-Fletcher-Goldfarb-Shanno (BFGS)

Seja o problema de otimização dado por (Luenberger,1989; Fletcher, 2001; da Silva, 2001):

$$\text{Minimizar } f(\mathbf{x}) \quad \mathbf{x} \in R^n$$

Sendo $f(\mathbf{x})$ contínua com derivadas parciais contínuas até a segunda ordem.

Definindo-se as diferenças:

$$\delta^{(k)} = x^{(k+1)} - x^{(k)} \text{ e } \gamma^{(k)} = g^{(k+1)} - g^{(k)} \tag{3.17}$$

O gradiente $\nabla f(\mathbf{x}' + \alpha s)$ pode ser expandido em série de Taylor, em torno de \mathbf{x}' , gerando a seguinte equação correlacionada com $\gamma^{(k)}$:

$$\gamma^{(k)} = G^{(k)}\delta^{(k)} + o(\|\delta^{(k)}\|) \tag{3.18}$$

Na qual termos de alta ordem são zero para funções quadráticas. Como $\gamma^{(k)}$ e $\delta^{(k)}$ só podem ser calculados após busca linear, estes dois parâmetros não são geralmente relacionados com o valor de H para a iteração corrente k . De modo que:

$$H^{(k+1)}\gamma^{(k)} = \delta^{(k)} \tag{3.19}$$

A condição acima é denominada de condição quase-Newton ou secante.

A utilização da inversa da matriz Hessiana se dá a partir da geração de uma seqüência de matrizes $H^{(0)}, H^{(1)}, \dots, H^{(k-1)}, H^{(k)}$, onde em cada iteração a matriz $H^{(k+1)}$ é calculada a partir da matriz imediatamente anterior $H^{(k)}$. Se não for possível obter as informações iniciais da matriz Hessiana, adota-se como aproximação inicial $H^{(0)}$ a matriz identidade. Depois de determinar o ponto $x^{(k+1)}$, uma nova aproximação $H^{(k+1)}$ é obtida atualizando $H^{(k)}$, ou seja,

$$H^{(k+1)} = H^{(k)} + U^{(k)} \tag{3.20}$$

onde, $U^{(k)}$ é uma matriz de atualização.

A fórmula para atualização da inversa da matriz Hessiana para o método Quase-Newton BFGS é dada por:

$$H^{(k+1)} = H^{(k)} + \left(1 + \frac{(\gamma^{(k)})^T H^{(k)} \gamma^{(k)}}{(\delta^{(k)})^T \gamma^{(k)}}\right) \frac{\delta^{(k)} (\delta^{(k)})^T}{(\delta^{(k)})^T \gamma^{(k)}} - \frac{(\delta^{(k)} (\gamma^{(k)})^T H^{(k)} + H^{(k)} \gamma^{(k)} (\delta^{(k)})^T)}{(\delta^{(k)})^T \gamma^{(k)}} \quad (3.21)$$

3.7.5 Método Davidon-Fletcher-Powell (DFP)

Outra metodologia para atualização da inversa da matriz Hessiana foi proposta por Davidon, Fletcher e Powell (DFP), sendo dada por (Luenberger, 1989; Fletcher, 2001):

$$H^{(k+1)} = H^{(k)} + \frac{\delta^{(k)} (\delta^{(k)})^T}{(\delta^{(k)})^T \gamma^{(k)}} - \left(\frac{\gamma^{(k)} (\gamma^{(k)})^T H^{(k)}}{(\gamma^{(k)})^T H^{(k)} \gamma^{(k)}} \right) \quad (3.22)$$

Existem algumas relações comuns entre as equações utilizadas pelos métodos BFGS e DFP. Para encontrar tais relações, se define uma matriz B como sendo igual a H^{-1} .

Desta forma a expressão da matriz B , para ambos os métodos, é definida da seguinte forma:

$$B_{BFGS}^{(k+1)} = B^{(k)} + \frac{\gamma^{(k)} (\gamma^{(k)})^T}{(\gamma^{(k)})^T \delta^{(k)}} - \frac{B^{(k)} \delta^{(k)} (\delta^{(k)})^T B^{(k)}}{(\delta^{(k)})^T B^{(k)} \delta^{(k)}} \quad (3.23)$$

e

$$B_{DFP}^{(k+1)} = B^{(k)} + \left(1 + \frac{(\delta^{(k)})^T B^{(k)} \delta^{(k)}}{(\gamma^{(k)})^T \delta^{(k)}}\right) \frac{\gamma^{(k)} (\gamma^{(k)})^T}{(\gamma^{(k)})^T \delta^{(k)}} - \left(\frac{\gamma^{(k)} (\delta^{(k)})^T B^{(k)} + B^{(k)} \delta^{(k)} (\gamma^{(k)})^T}{(\gamma^{(k)})^T \delta^{(k)}} \right) \quad (3.24)$$

Para as quais vale a relação:

$$B_{BFGS}^{k+1} H_{BFGS}^{k+1} = I \quad (3.25)$$

e

$$B_{DFP}^{k+1} H_{DFP}^{k+1} = I \quad (3.26)$$

Outra expressão que une os métodos BFGS e DFP pode ainda ser mencionada:

$$H_{\phi}^{(k+1)} = (1 - \phi) H_{DFP}^{(k+1)} + \phi H_{BFGS}^{(k+1)} \quad (3.27)$$

Esta expressão gera a chamada família de Broyden, incluindo o método DFP ($\phi = 0$) e o BFGS ($\phi = 1$). O valor de ϕ obtido pela fórmula de Hoshino para cada iteração

$$\phi = \frac{1}{\frac{1 \mp (\gamma^{(k)})^T H \gamma^{(k)}}{(\delta^{(k)})^T \gamma^{(k)}}} \quad (3.28)$$

O método DFP trabalha bem, sendo muito mais eficiente que o gradiente descendente e o gradiente conjugado. Entretanto, o método apresenta falhas de convergência em buscas lineares em funções que não sejam quadráticas. Tal metodologia atualmente tem uso menos freqüente em relação a métodos que surgiram posteriormente, como o BFGS. O BFGS é considerado como o "melhor" método quase-Newton, apresentando convergência global em buscas lineares em funções não quadráticas.

Algumas propriedades do método Quase-Newton DFP:

- 1 É possível demonstrar que para funções quadráticas que o método gera direções e gradientes conjugados quando $H^{(1)} = I$.
- 2 Quando se tratam de funções quaisquer, se demonstra que:
 - 2.1 O método preserva a matriz $H^{(k)}$ como positiva definida.
 - 2.2 Necessita $3 \cdot n^2 + O(n)$ multiplicações por iteração.
 - 2.3 Possui convergência numa razão super linear.
 - 2.4 Possui convergência global para funções estritamente convexas.

3.7.6 Método de Gauss-Newton (GN)

Considerando um problema onde a função objetivo é um somatório de m termos quadrados (Fletcher, 2001):

$$\sum_{i=1}^m [r_i(x)]^2 = r^T r \quad (3.29)$$

Na qual, $r = r(x)$. Certos métodos existentes utilizam a estrutura da eq. (3.29) de maneira vantajosa a qual é denominada de problema de mínimos quadrados não linear. Uma alternativa para propor este problema consiste na resolução de sistemas de m equações:

$$r_i(x) = 0, i = 1, 2, \dots, m \quad (3.30)$$

Neste caso, as funções $r_i(x)$ são interpretadas como resíduos das equações. Quando $m > n$, o sistema é denominado de sobre-determinado e, comumente, não é possível a obtenção de uma solução exata. Outra situação ocorre quando $m = n$, sendo o sistema bem-determinado e uma solução pode ser obtida a partir de (3.30). Um classe de problema deste tipo é o ajuste de

dados a um modelo. Entretanto, é mais conveniente para resolução deste tipo de problema a solução da equação (3.29). Para resolução do problema descrito por (3.30) e (3.29) é necessário o conhecimento das derivadas de $r_i(x)$, informação nem sempre disponível para a maioria dos problemas. Os resíduos podem ser definidos como:

$$r_i(x) = \phi(t_i, x) - d_i, i = 1, 2, \dots, m, \quad (3.31)$$

Na qual $\phi(t_i, x)$ é o modelo a ser ajustado aos dados experimentais d_i . Para um problema de mínimos quadrados não linear, as derivadas de $f(x)$ são dadas por:

$$g(x) = 2Ar \quad (3.32)$$

e

$$G(x) = 2AA^T + 2 \sum_{i=1}^m r_i \nabla^2 r_i \quad (3.33)$$

Onde $r_i = r_i(x)$, etc., e $A(x) = [\nabla r_1, \nabla r_2, \dots, \nabla r_m]$ é uma matriz Jacobiana $n \times m$. Na, qual as colunas são os vetores primeiras derivadas ∇r_i dos componentes de r ($A_{ij}(x) = \partial r_j / \partial x_i$). Para a resolução do problema é possível utilizar (3.32) em conjunção com um método quase-newtoniano ou (3.32) e (3.33) com uma modificação do método de Newton. Entretanto, se r_i não possuir expressões para a Hessiana $\nabla^2 r_i$, a resolução do problema pode ser comprometida. No caso dos resíduos r_i serem pequenos, uma boa aproximação de $G(x)$ pode ser dada por

$$G(x) \approx 2AA^T \quad (3.34)$$

Esta aproximação é equivalente a se fazer uma aproximação linear dos resíduos $r_i(x)$. Utilizando a aproximação da Hessiana (eq. 3.34), o método de Newton torna-se o de Gauss-Newton ou método de mínimos quadrados generalizado. Para a k -ésima iteração, os passos do método de Gauss-Newton podem ser escritos como:

a Resolva $A^{(k)} A^{(k)T} \delta = -A^{(k)} r^{(k)}$ para $\delta = \delta^{(k)}$

b Faça $x^{(k+1)} = x^{(k)} + \delta^{(k)}$

Para prevenir falhas na convergência, particularmente em funções com um alto grau de não linearidade, foi introduzida no método de Gauss-Newton uma estratégia de busca linear. Neste caso, a direção de busca s^k é encontra resolvendo o sistema linear:

$$A^{(k)} A^{(k)T} s^{(k)} = -A^{(k)} r^{(k)} \quad (3.35)$$

Um aspecto que parece contribuir para a eficiência do método é a matriz $2AA^T$ ser semi-definida positiva ou usualmente positiva definida de modo que o fato de $G^{(k)}$ não ser explicitamente conhecida não se constitui em um grande problema. Entretanto, em certos problemas

não-lineares, nos quais r^* são relativamente grandes, o termo $\sum r_i^* \nabla^2 r_i^*$ não é desprezível. Isto implica que a taxa de convergência em problemas com grandes resíduos pode ser lenta ou implicar em falhas no processamento.

Uma possível solução para este problema pode ser introduzir uma correção para $\delta^{(k)}$:

$$(A^{(k)}A^{(k)T} + \lambda I)\delta^{(k)} = -A^{(k)}r^{(k)} \quad \lambda \geq 0 \quad (3.36)$$

Que se constitui no método de Levenberg-Marquardt.

3.7.7 Método de Levenberg-Marquardt (LM)

Se $G^{(k)}$ é próxima de ser positiva definida pode ser somente necessário adicionar-se a esta um pequeno múltiplo da matriz identidade I , possibilitando a obtenção de uma boa direção de busca (Fletcher, 2001). Neste contexto, Marquardt (1963) criou um método de otimização, baseado em uma idéia de Levenberg (1944), que está entre os extremos do método da Hessiana inversa e do gradiente descendente (eq. 3.9). Nesta última equação, a "constante" é um fator que relaciona δa_l e β_k (eq. 3.15). Este fator não fornece nenhuma informação relativa ao gradiente, nem da ordem de magnitude do fator de escala do problema, somente o valor da inclinação ou obliquidade. No método de LM, alguma informação a respeito da ordem de magnitude pode ser obtida a partir dos elementos da matriz Hessiana.

Em termos dimensionais, percebe-se pela equação (3.6) que χ^2 é adimensional. Por sua vez, β_k possui dimensão correspondente ao inverso de a_k , o qual pode ter unidades variadas como cm, KWh, etc. A constante de proporcionalidade entre δa_l e β_k , entretanto, possui dimensão de a_k^2 .

Analisando a matriz $[\alpha]$ do tópico (3.6.1), verifica-se que os componentes apresentam uma única dimensão, sendo esta igual a $1/\alpha_{kk}$, a recíproca do elemento da diagonal. Esta quantidade pode ser então um referencial para o fator de escala. Entretanto, este parâmetro poderia ser excessivamente grande. Como solução, é efetuada uma divisão entre este parâmetro e uma constante adimensional λ , com a possibilidade de que $\lambda \gg 1$. A equação (3.15) pode então ser reescrita como (Press et al., 2001):

$$\delta a_l = \frac{1}{\lambda \alpha_{ll}} \beta_l \quad (3.37)$$

ou

$$\lambda \alpha_{ll} \delta a_l = \beta_l \quad (3.38)$$

É necessário que todo α_{ll} seja positivo, mas é garantido por sua própria definição (3.15).

Outra propriedade do LM é que as equações (3.39), (3.40) e (3.14) podem ser combinadas para definirem uma nova matriz α' , com as seguintes prerrogativas

$$\alpha'_{jj} \equiv \alpha_{jj}(1 + \lambda) \quad (3.39)$$

e

$$\alpha'_{jk} \equiv \alpha_{jk} \quad (j \neq k) \quad (3.40)$$

e então as equações (3.39), (3.40) e (3.14) podem ser substituídas por

$$\sum_{l=1}^M \alpha'_{kl} \delta a_l = \beta_k \quad (3.41)$$

Quando λ é muito grande, a matriz α'' é forçada a ser diagonalmente dominante de modo que (3.41) passa a ser idêntica a (3.39) ou (3.40). Se λ se aproxima de zero, a equação (3.41) passa a ser análoga a (3.14).

Para um dado conjunto inicial de parâmetros a , o algoritmo do método de LM é o seguinte:

1. Compute $\chi^2(a)$.
2. Escolha um valor modesto de λ tal como $\lambda = 0,001$.
3. Resolva as equações lineares (5.28) para δa e calcule $\chi^2(a + \delta a)$.
4. Se $\chi^2(a + \delta a) \geq \chi^2(a)$, aumente λ por um fator de 10 e volte ao passo (2).
5. Se $\chi^2(a + \delta a) < \chi^2(a)$, diminua λ por um fator de 10, atualize a solução anterior tal que $a \leftarrow a + \delta a$, e volte ao passo (2).

Também é necessário se estabelecer uma condição para parada tal como um limite estabelecido pelo usuário. Entretanto, é em geral, inútil e desnecessário se ocasionar um número excessivo de iterações devido a qualquer um destes critérios visto que o mínimo é somente uma melhor estimativa estatística dos parâmetros a . Ademais, mudanças em parâmetros que resultem em mudança de $\chi^2(a)$ por uma quantidade $\ll 1$ nunca são estatisticamente úteis.

Além disso, não é incomum encontrar parâmetros a ao largo de um mínimo em um vale plano de um função de topografia complexa. A razão para isto é que o método de LM generaliza o método das equações normais, possuindo o mesmo problema de degenerescência próximo do mínimo. Outro problema que pode surgir é a ocorrência de máximos locais expressivos. Tais máximos podem resultar em uma grande correção, o que resultara em sua rejeição, como resultado do aumento de λ . Para valores suficientemente grandes de λ , a matriz $[\alpha']$ é positiva definida e sob esta circunstância não ocorrem máximos locais consideráveis.

Com as considerações anteriores, na prática, as iterações poderiam cessar quando a quantidade χ^2 for igual a algo como, absolutamente, uma ou duas vezes o valor de 0,01 ou a alguma

quantidade fracional como 10^{-3} . Estas interações não devem cessar após um passo, onde χ^2 aumenta. Esta última observação somente demonstra que λ ainda não está autoajustado de maneira ótima.

Quando um mínimo aceitável é encontrado, isto implica que $\lambda = 0$ e se computa a matriz

$$[C] = [\alpha]^{-1} \quad (3.42)$$

Que se constitui na matriz de covariância estimada dos parâmetros ajustados.

3.8 Conclusões

A otimização é uma importante ferramenta para a engenharia e ciências exatas. Neste procedimento, busca-se obter o pontos extremos de uma função. Tais pontos podem ser obtidos a partir da imposição ou não de limitações as funções os quais estes estão correlacionados. Estes limites condicionaram que os problemas de otimização fossem divididos nos "com restrição" e "sem restrição". Uma parte particularmente bem desenvolvida da otimização com restrições é a programação linear. Na programação linear se objetiva o estudo do problema de otimização e a escolha do método mais adequado para sua resolução. Nesta programação, as restrições possuem dependência linear com as variáveis da função objetivo. Além da programação linear existe a não linear, inteira, mista, quadrática, etc. que trabalham com restrições ou funções objetivos diferentes da linear. Para a otimização, existem inúmeros métodos para resolução, cada qual mais adaptado para um tipo de problema. Entre os métodos de otimização sem restrição existem os de tipo newtoniano e quase-newtoniano. No primeiro, utiliza-se a matriz Hessiana $G(x)$, na relação de recursão para obtenção do ponto ótimo. Tal matriz deve ser atualizada a cada interação. Quando $G(x^*)$ é positiva definida, implica que x^* é um minimizador local. Para certas funções, nas quais a derivada segunda é indisponível, no caso de uma busca linear, aplica-se uma aproximação de $H(x) = G(x)^{-1}$, caracterizando assim os chamados métodos quase-newtonianos, tais como os métodos BFGS ou DFP que diferem na metodologia da atualização de $H(x)$. Para a resolução do problema de mínimos quadrados não linear pode ser empregado o método Gauss-Newton ou mínimos quadrados generalizado. Tal método se vale de uma aproximação quadrática da função de mérito na qual em $G(x)$ são ignorados termos associados aos resíduos r_i , admitindo que os mesmo sejam pequenos. Como a aproximação de $G(x)$ é semi-definida positiva ou, em alguns casos, positiva definida se garante uma direção de decréscimo na busca por um ponto de mínimo. Tal condição de convergência pode não ser verdadeira caso os resíduos sejam grandes. Com resíduos grandes, o método pode ter uma convergência lenta ou mesmo falhar. Situação análoga ocorre quando $x^{(1)}$ é distante de x^* . Para tentar contornar este problema, foi sugerido por Levenberg (1944) e Marquardt (1963), um

método no qual se acrescenta um pequeno múltiplo da matriz identidade I a aproximação de $G(x)$ do método GN, visando a obtenção de uma matriz positiva definida. O método, denominado de Levenberg-Marquardt, é considerado robusto para o problema de mínimos quadrados não-linear apresentando, entretanto, de convergência mais lenta que o de Gauss-Newton, em problemas com grandes resíduos.

Capítulo 4

Algoritmos Genéticos

4.1 Introdução

Algoritmo Genético (AG) é um método de otimização e busca que faz uso de conceitos da genética, sendo baseado nos mecanismos de evolução de populações de seres vivos (Castilho, 2003). Esta evolução é governada pelo princípio da seleção natural desenvolvido por Darwin (1859). Neste, em uma população de indivíduos, aqueles com melhores características genéticas apresentam maiores chances de sobrevivência e reprodução. Por sua vez, indivíduos menos aptos tendem a desaparecer durante o processo evolutivo. Com base nestas observações, Holland (1975) estabeleceu os princípios básicos dos AGs. O objetivo desta iniciativa era implementar um algoritmo, que realizando uma simulação da evolução biológica, por meio de uma busca multidirecional no espaço de soluções potenciais do problema, poderia permitir a solução de problemas variados.

O AG geralmente usa regras de transição probabilística para selecionar algumas soluções para reprodução e outras para o descarte. Deste modo, os AGs procuram utilizar uma versão dos mecanismos básicos observados no processo de evolução natural. Este mecanismo fundamenta-se no cromossomo, que é associado a cada indivíduo da população, e corresponde a um ponto no espaço de busca, vinculado a uma possível solução (ou hipótese) do problema. A função dos cromossomos se concentra no processo de seleção dos melhores indivíduos, resultantes da reprodução e mutação dentro de uma dada população (Mitchell, 1994).

Apesar do processo de codificação dos cromossomos biológicos, na época de sua descoberta, não ser totalmente compreendido as características gerais da teoria de Darwin tais como evolução natural, o processo de reprodução e a não necessidade de uma memória para os cromossomos, foram aceitas e compreendidas mais facilmente (Davis, 1991). Com relação as técnicas utilizadas para busca da solução, embora estas apresentem aspectos aleatórios, não implicam que os AGs sejam globalmente aleatórios. Estes algoritmos utilizam um sistema de busca heurística adaptativa para exploração das informações disponíveis para direcionar a busca

para a região que contenha os melhores resultados, dentro do espaço de soluções. Entretanto, não é garantido que os AGs encontrem a solução global do problema mas estes geralmente são capazes de encontrar soluções que são "aceitavelmente boas" de uma maneira "aceitavelmente rápida" (Beasley et al., 1993). Os AGs são capazes de resolver problemas que não são passíveis de serem solucionados por outros métodos.

A seqüência abaixo caracteriza o funcionamento de um AG, segundo Holland (Michalewicz, 1996):

- 1 [Início]: Efetuar a geração aleatória de uma população de n cromossomos.
- 2 [Determinação]: Determinar o grau de aptidão de cada cromossomo x da população.
- 3 [Teste]: Os melhores indivíduos são suficientemente bons? Sim? Vá para o passo 7. Não? Vá para o passo 4.
- 4 [Nova População]: Criar de uma nova população através dos passos 4.1, 4.2 e 4.3.
 - 4.1 [Seleção]: Selecionar cromossomos progenitores na população.
 - 4.2 Cruzamento ("Crossover"): Realizar cruzamentos dos cromossomos selecionados em 4.1 para formar as novos descendentes.
 - 4.3 [Mutação]: Realizar mutações nos novos cromossomos produzidos em 4.2.
- 5 [Eliminação/Substituição]: Eliminar indivíduos da antiga população pela substituição de elementos da população gerada em 4.
- 6 ["Looping"]: Voltar ao passo 2.
- 7 Fim.

O algoritmo de Holland se baseou, inicialmente, na manipulação de algumas poucas cadeias binárias ("strings"), que eram definidas como análogas aos cromossomos. Este algoritmo não possuía conhecimento prévio do tipo de problema que estava sendo solucionado e que resultados haviam dentro dos cromossomos, mas encontrou a solução do problema sem a necessidade de supervisão. As únicas informações fornecidas para o sistema eram os ajustes de cada cromossomo, obtidas a partir de um função específica. Este ajuste representa o grau de adaptação do indivíduo ao problema a ser solucionado. No processamento, notou-se que tanto maior for o ajuste do indivíduo, maiores serão suas chances de se reproduzir, transmitindo suas características para seus descendentes.

Um sistema genético apresenta as seguintes etapas, detalhadas abaixo (Castilho, 2003):

1. Inicializar de forma aleatória a população: corresponde a criação da primeira geração de uma espécie, onde os membros possuem como características o número de cromossomos de cada indivíduo, o número de genes de cada cromossomo e a posição de cada gene no cromossomo.
2. Determinar o ajuste de cada indivíduo da população: Nesta etapa ocorre a avaliação de

cada indivíduo, no momento em que este interage com o meio ambiente. O melhor ajuste se dará para aquele indivíduo que estiver mais próximo da solução procurada.

3. Seleção: São selecionados os indivíduos de maior ajuste para se reproduzirem. Para cada indivíduo é dado uma chance de se reproduzir sendo esta proporcional ao seu ajuste.
4. Reprodução dos Indivíduos Selecionados: Se vale de processos análogos aos naturais que são o cruzamento e a mutação. A fase reprodutiva implica no cruzamento dos cromossomos.

Os AGs, embora muito distantes do modelo darwinista, vem demonstrando possuir um grande potencial para a resolução de problema de otimização, classificação e adaptação. Na resolução de problemas de otimização e adaptação, os AG são caracterizados por não falhar tão facilmente, mesmo que os dados de entrada sejam alterados ou ainda que estes possuam ruído (Goldberg, 1989).

4.2 Terminologia Biológica

Um AG pode ser entendido como uma representação computacional da evolução darwiniana, incorporando conceitos de genética. Isto justifica os AGs possuírem muitos termos originários da Biologia. Os principais termos encontrados na literatura e de relevância para este trabalho são (Castilho, 2003):

1. Gene: é uma informação do cromossomo, representando uma certa característica da solução cromossomo.
2. Indivíduo ou Cromossomo: um elemento da população. Um indivíduo é formado pelo cromossomo e sua função aptidão - Representa uma possível solução do problema.
3. População: conjunto de indivíduos (cromossomos) que representam os atuais pontos que fazem parte do espaço de soluções.
4. Geração: identifica cada uma das várias populações criadas durante o processo evolutivo.
5. Função Aptidão: função que condiciona a adequabilidade de um cromossomo. Geralmente usada no mecanismo de seleção, para identificar quais indivíduos irão sobreviver e recombinar. Está ligada ao valor da função objetivo e pode incluir uma função de penalidade.

4.3 Propriedades dos Algoritmos Genéticos

Existem vantagens na utilização dos AGs. Lemonge (1999) cita várias das quais podem ser destacadas:

1. Otimizam um número grande de variáveis.
2. Realizam buscas simultâneas em várias regiões do espaço de busca.

3. São fáceis de serem implementados em computadores.
4. Fornecem uma lista de parâmetros ótimos e não uma única solução.
5. São flexíveis para trabalhar com restrições e otimizar múltiplas funções com objetivos conflitantes.
6. São facilmente hibridizados com outras técnicas.

Apesar das inúmeras vantagens, Coley (1999) menciona que vários trabalhos que demonstram as prerrogativas favoráveis aos AGs, analisam o comportamento de parâmetros ou metodologias para problemas de otimização de baixa complexidade. Frequentemente, tais trabalhos não deixam claro se os resultados obtidos podem ser considerados em um contexto mais amplo. Na realidade, a única forma de proceder a busca da solução de um problema usando AG é avaliar resultados obtidos em problemas similares e, então, escolher uma abordagem que seja sensível ao problema em questão e factível de implementação.

4.4 Esquemas de Representação dos Cromossomos

A fim de implementar um AG, as soluções candidatas precisam ser representadas e codificadas na forma de cromossomos, sendo a base para o uso de um AG na resolução de problemas do mundo real (Gen & Cheng, 1997). Cada indivíduo é codificado por um vetor finito de componentes ou variáveis relacionados com um tipo de alfabeto. A representação de uma variável pode ser feita usando diferentes alfabetos, destacando-se entre estes, a codificação binária e a real ou, então, representações típicas de um determinado problema e que tentam refletir a estrutura natural dos dados. Geralmente, a representação binária é utilizada para formulação do problema com variáveis discretas e a real para variáveis contínuas.

4.4.1 Representação Binária

No caso do alfabeto binário utilizam-se os números 0 e 1. O tamanho do vetor está ligado diretamente ao limites de valores que estas variáveis podem assumir. Para uma variável contínua, a determinação do número de bits (m) para representar seus possíveis valores depende do limite inferior (b) e superior (c) do intervalo de valores que esta variável possa assumir no domínio do problema, bem como a precisão (k). Tal determinação é feita obtendo-se o m que satisfaça

$$2^{m-1} \leq (c - b)10^k \leq 2^m - 1 \quad (4.1)$$

Para ilustrar a codificação binária com variáveis contínuas, se pode supor que um dado problema envolva três variáveis x_1, x_2 e x_3 , cada qual com valores no intervalo $[-1,0 \ 2,0]$ e condicionado a uma precisão 2. Usando (4.1), o valor de m é igual a 9, portanto implicando que

cada variável sera representada no cromossomo por uma cadeia de 9 bits. Um possível solução gerada randomicamente é

$$s_1 = 100001011101110101000111110$$

onde as variáveis podem ser identificadas como

$$x_1 = 100001011$$

$$x_2 = 101110101$$

$$x_3 = 000111110$$

No fim do processamento genético, a solução estará codificada em binário. A decodificação é possível através do uso da equação

$$x = b + (c - b) \frac{b_{10}}{2^m - 1} \quad (4.2)$$

onde, b_{10} é o número decimal correspondente à cadeia de bits

No caso do problema apresentado, x_1, x_2 e x_3 são respectivamente iguais a 0,57; 1,19 e -0,69 com o uso de (4.2).

Existem limitações na representação de números reais em forma binária. Tais limitações são associadas com a precisão mas podem ser solucionadas com o aumento do número de bits que representam a variável (Coley, 1999).

4.4.2 Representação Real

A representação utilizando números reais exige menos memória e possibilita uma maior versatilidade quando da criação de novos operadores e variação dos existentes (Michalewicz, 1996).

4.5 Operadores Genéticos

Os operadores genéticos tem como finalidade modificar a população em sucessivas gerações. Com esta metodologia busca-se gerar uma população diversificada e que mantenha as características de adaptação adquiridas pelas gerações anteriores (Serrada, 1996).

Um algoritmo genético evolui através de três operadores básicos (Ramos, 1994):

1. Seleção: Realiza a verificação do grau de adaptabilidade dos indivíduos.
2. Cruzamento ("Crossover") : Representa o acasalamento.
3. Mutação: Introduce modificações aleatórias.

4.5.1 Operador de Seleção

Este operador proporciona selecionar o melhor ou melhores indivíduos que irão dar origem a novas gerações através da reprodução. Cada indivíduo, por intermédio deste operador, receberá uma probabilidade de reprodução que depende de seu valor correspondente na função objetivo. As soluções com maior probabilidade terão maiores oportunidades para reprodução, ocasionando que seus descendentes herdem características dos progenitores (Castilho, 2003).

Em um AG, se distinguem três mecanismos de seleção: seleção por elitismo; por criadores e por sobreviventes. Pela definição, a seleção por elitismo se faz por amostragem direta na população. A seleção por criadores é feita de forma aleatória e a de sobreviventes por cruzamento imediato.

Qualquer que seja o mecanismo de seleção, este tem como prerrogativas a insensibilidade a distribuição de aptidões e aos erros da amostra.

Algumas técnicas aleatórias e diretas de seleção são apresentadas a seguir:

- 1 Seleção "rank": os cromossomos são classificados por valor de aptidão. Os melhores cromossomos possuem as melhores posições e, conseqüentemente, maiores chances de reprodução. Segundo Bennett (1997), associa-se a posição 0, para o melhor indivíduo, e $N_{pop} - 1$, para o membro com pior valor da função de aptidão. Nesta definição de posição, N_{pop} é o número de indivíduos da população. Um indivíduo α com posição r_α no "rank" é selecionado com uma probabilidade P_α dado por:

$$P_\alpha = \frac{2r_\alpha}{N_{pop}(N_{pop} - 1)} \quad (4.3)$$

- 2 Seleção por Roleta: O funcionamento da técnica de "roleta" baseia-se em atribuir a cada indivíduo da população um espaço na roleta que é proporcional a sua aptidão (F_i) ao problema estudado. Neste caso, a probabilidade de seleção p_i de um dado cromossomo, com um F_i correspondente, em uma população N_{pop} é dada por:

$$p_i = \frac{F_i}{\sum_{i=1}^{N_{pop}} F_i} \quad (4.4)$$

A partir do qual se calcula a probabilidade acumulada (q_i) de cada cromossomo através de:

$$q_i = \sum_{j=1}^i p_j \quad (4.5)$$

Durante o processo de seleção a roleta é girada N_{pop} vezes, elegendo indivíduos para reprodução. Indivíduos com maiores valores de probabilidade de seleção possuem maiores chances

de escolha. Em certas situações, tal metodologia pode gerar convergência prematura, quando a função de aptidão atribui valores altos a um dado indivíduo. Esta situação leva a uma monopolização da geração por este indivíduo. O processo de girar a roleta é equivalente a gerar aleatoriamente um número entre $r \in [0, 1]$. Se $r \leq q_1$ então, o primeiro cromossomo é selecionado.

O funcionamento desta técnica é exemplificado abaixo:

1 2 3 4 5 6 7 8 Indivíduos

1 5 2 1 3 4 1 2 Aptidão

Espaço na Roleta (primeira linha abaixo

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

1 2 2 2 2 2 3 3 4 5 5 5 6 6 6 6 7 8 8

Sorteio Aleatório (Resultado do Giro da Roleta)

5 12 15 8 17 11 3 1 Posição na Roleta

Casais Selecionados para a Reprodução

2,5 6,3 7,5 2,1

4.5.2 Operador de Cruzamento ("Crossover")

O "crossover" é considerado o operador genético predominante e ocorre com uma probabilidade definida pela taxa de cruzamento p_c ($0,6 \leq p_c \leq 1,0$). Uma taxa alta permite uma exploração maior do espaço de solução e reduz as chances de convergência para um ótimo local. Entretanto, o maior espaço de busca pode ocasionar em um esforço computacional desnecessário, devido a exploração de regiões não promissoras (Castilho, 2003).

O funcionamento deste processo se dá através de uma operação de cruzamento e recombinação. A compreensão deste operador se resume nos seguintes fatores seguindo a seqüência:

1. Escolhe-se dois indivíduos para fazer a reprodução usando o operador de seleção, originando-se um casal de pais.
2. Aleatoriamente escolhe-se um ponto de corte para o cruzamento ao longo do vetor de bits, ou seja, do cromossomo.
3. A partir do ponto de corte, permutam-se partes dos cromossomos dos progenitores (casal de pais). Com esta permuta, os dois novos descendentes são criados sendo inseridos na geração seguinte.

O ponto de corte pode ser escolhido entre várias técnicas, dentre estas, na representação binária:

- 1 Um ponto - Um ponto escolhido nos cromossomos dos pais e a partir deste ponto as informações genéticas são permutadas, gerando os descendentes correspondentes

Exemplo:

Pai - 111P111000000

Mãe - 000P000111111

(Onde P representa o ponto de corte)

Filho 1 - 111 + 000111111 = 111000111111

Filho 2 - 000 + 111000000 = 000111000000

- 2 Cruzamento multi-ponto: Consiste em efetuar o cruzamento entre os indivíduos em dois ou mais pontos dos pares de cromossomos.

Exemplo:

Pai - 111P111100P20000

Mãe -000P100011P21111

(Onde P1 e P2 representam os pontos de corte)

Filho 1 - 111+00011+0000 = 111000110000

Filho 2 - 000+11100+1111 = 000111001111

- 3 Cruzamento segmentado: É considerado como uma versão do cruzamento multi-ponto que permite associar uma probabilidade de segmentação a um número finito de pontos de cruzamento. Esta probabilidade está ligada a geração de um cruzamento quando se seleciona um certo ponto da cadeia.

- 4 Cruzamento uniforme: Não utiliza pontos de cruzamento mas determina através de uma máscara de bits aleatórios, quais genes de cada pai serão herdados por cada um dos filhos. Se o primeiro bit da máscara for 1, o primeiro bit do pai é copiado para o primeiro bit do filho 1; caso contrário, o bit selecionado para ocupar tal posição será o da mãe. O processo se repete para todos os bits. Na geração do segundo filho, o papel dos pais é invertido.

Exemplo

Máscara: 110010010001

Pai - 111111000000

Mãe - 000000111111

Filho 1 - 110011101110

Filho 2 - 001100010001

4.5.3 Operador de Mutação

A finalidade deste operador é garantir a diversidade dos indivíduos na população, impedindo convergências prematuras. Em uma codificação binária dos cromossomos, esta diversidade é o resultado da mudança do valor de um dado gene de 0 para 1, no caso da representação binária. O operador de mutação é aplicado aos indivíduos com uma probabilidade dada pela taxa de mutação p_m ($0,001 \leq p_m \leq 0,1$). Se p_m for muito baixa pode acontecer um comprometimento da diversidade da população. Se p_m for muito alta, aconteceram muitas perturbações aleatórias e os filhos provavelmente começaram a perder suas relações paternas, comprometendo a solução (Castilho, 2003).

Exemplo:

Cromossomo: 011001011010

Novo Cromossomo: 010001111000

Onde os genes em **negrito** no cromossomo original sofreram a mutação

O operador de mutação possui importância considerada secundária se comparado como o de cruzamento, com o papel de recuperar a perda de material genético, conforme sugerido por Srinivas & Patnaik (1994). Entretanto, verifica-se que este é imprescindível pois um AG pode funcionar sem realizar cruzamentos mas não sem mutações. Sua função específica é redirecionar o local de busca no espaço de soluções possíveis. O operador de cruzamento, por sua vez, tem a capacidade de direcionar a busca na direção de melhores soluções, com a ajuda de uma procura ao acaso. Por esta característica, o operador de mutação gera, com baixa probabilidade, um grande número de novos indivíduos. Alguns tipos de mutações na representação real são apresentadas a seguir:

1. A Mutação randômica, também denominada de aleatória, trata-se da simples substituição de um gene por um número escolhido aleatoriamente no intervalo permitido pelo problema.
2. Mutação "creep" pode ser realizada multiplicando-se o gene por um número aleatório próximo de 1. A idéia é que perturbando um cromossomo, que esteja próximo do ponto ótimo, poderia deslocar o algoritmo mais rápido para a solução. A taxa de mutação "creep" pode ser relativamente alta, uma vez que esse operador é usado apenas para explorar localmente o espaço de busca.

4.6 Elitismo

Introduzido por Kenneth De Jong, em 1975 (Goldberg, 1989), o mecanismo de elitismo faz parte do operador de seleção. Esta técnica força o algoritmo a reter os melhores indivíduos

em cada geração. Basea-se no fato de que na natureza o indivíduo mais apto, além de se reproduzir mais, possui maior longevidade passando de uma geração para outra (Davis, 1991). O procedimento consiste em realizar a seleção em duas etapas: Primeiro, verifica-se a elite de n componentes entre os melhores da população inicial e incorporando-os diretamente na população final. Na segunda etapa, é efetuada a seleção na população restante. A vantagem do uso deste procedimento é a convergência mais rápida do algoritmo, proporcionando um ótimo global. Entretanto, se a função de evolução possuir soluções sub-ótimas, a convergência não será satisfatória (Karlec, 1998).

4.7 Codificação Real \times Binária

A simbologia utilizada para a representação dos genes dos cromossomos pode ser o fator determinante do tipo de codificação. O alfabeto binário (0,1) tem sido comumente mais utilizado, permitindo se ter um maior número de esquemas por bit de informação. Sua popularidade é atribuída a facilidade de análise teórica e por este permitir a construção de operadores genéticos mais sofisticados.

A codificação real foi inicialmente utilizada para solucionar problemas de quimiometria. Posteriormente, foram aprofundados os estudos sobre esta codificação sendo propostos diversos esquemas de cruzamento e mutação. Estes estudos se basearam no fato de ser mais conveniente o uso de variáveis de domínio contínuo em problemas como os de otimização de parâmetros.

A utilização de parâmetros reais para variáveis a serem otimizadas permite o uso de grandes domínios, mesmo que estes sejam desconhecidos. Neste caso, o uso da codificação binária seria dificultada pois para incrementar o domínio é necessário a parte não inteira de um dado número, comprometendo a precisão da determinação da solução. Outra característica é que as bases das variáveis não precisam ter necessariamente potência de dois. A utilização da codificação real permite ainda um projeto menos complexo das ferramentas que manejam restrições não triviais e que a representação das soluções fiquem muito próximas à formulação natural de muitos problemas (Karlec, 1998; Charbonneau & Knapp, 1995).

4.8 PIKAIA - Descrição da Estrutura Utilizada

PIKAIA é uma sub-rotina de otimização, escrita em ANSI FORTRAN-77, baseada em algoritmos genéticos, desenvolvida no High Altitude Observatory (EUA) e disponível em domínio público (www.hao.ucar.edu/public/research/si/pikaia/pikaia.html). Esta sub-rotina maximiza globalmente uma função especificada pelo usuário através de sua invocação no corpo do programa principal. Esta sub-rotina utiliza codificação decimal. Esta escolha foi motivada por

necessidades de portabilidade, pois as operações com binários são usualmente efetuadas funções externas à plataforma FORTRAN. Isto tornaria mais complexa a utilização do código fonte em sistemas operacionais variados, como os comuns a PCs ou "workstations". A portabilidade, facilidade de uso e robustez tem permitido que a sub-rotina seja utilizada em um grande número de problemas variados que requerem otimização global (Charbonneau et al., 1998; Harries & Howarth, 2000). A descrição da estrutura da sub-rotina a seguir se destina a dar uma visão global dos operadores e estratégias utilizadas. Para um estudo em maior profundidade de funções diferentes das utilizadas neste estudo é sugerida a leitura do "A User's Guide to PIKAIA 1.0" (Charbonneau & Knapp, 1995), disponível na página da internet da sub-rotina. Neste, existe ainda uma lista com os usuários registrados e suas aplicações da sub-rotina.

A tarefa de maximização da função do usuário se inicia com a definição de um vetor ponto flutuante com dimensão n , representando o conjunto de parâmetros do modelo a ser otimizado. Como saída, é fornecida uma medida da qualidade do ajuste, S , baseada no valor final da função objetivo. O código evolui uma população com n soluções iniciais no espaço de dimensão n , guardadas em um vetor $P_old(1 : n, 1 : n)$, até um número limite de gerações. A população é usualmente inicializada com soluções escolhidas aleatoriamente em um intervalo definido pelo usuário no espaço dos parâmetros a serem explorados. A cada geração, pares de "progenitores" são extraídos da população através da técnica da "roleta" (Davis, 1991). Nesta, cada indivíduo possui um espaço na roleta que é proporcional ao seu valor de S relativo (s), definido como:

$$s = \frac{n_p - r + 1}{n_p} \quad (4.6)$$

Sendo r a ordem de importância na população ("ranking") em termos do ajuste S . O parâmetro r pode assumir valor igual a 1, para o membro da população com melhor qualidade e $r = n_p$, para o ajuste de pior qualidade. O uso do "ranking" destina-se a forçar a manutenção de características adequadas para a obtenção da solução do problema, tornando a medida de probabilidade de seleção independente da distribuição populacional atual de S .

Desta forma, aqueles indivíduos que estão mais aptos terão maiores chances de reprodução. Para o cruzamento é utilizado o cruzamento de um ponto e a mutação uniforme de um ponto com taxa variável de acordo com o valor de s .

O efeito da mutação pode ser prejudicial ou benéfico, destruindo cromossomos potencialmente superiores, mas de suma importância para preservar a variabilidade da população e evitar convergências prematuras. Para tal, a sub-rotina avalia o nível de convergência na população, possibilitando um ajuste da taxa de mutação. Este ajuste é baseado no valor de s do melhor indivíduo e do mediano da população, através da relação:

$$\Delta s = \frac{s(r = 1) - s(r = n_p/2)}{s(r = 1) + s(r = n_p/2)} \quad (4.7)$$

Na qual, a taxa de mutação pode ser aumentada (diminuída) quando é menor (maior) que um nível pré-determinado, sendo Δs a medida do grau de convergência da população.

Um plano reprodutivo é utilizado pelo AG como meio de introduzir novos indivíduos gerados pelo "crossover" e mutação na população. Neste trabalho, foi utilizada a metodologia "Steady-state-replace-worst", onde o membro da população, com pior valor de S , é substituído por um novo cromossomo. O elitismo foi aplicado nas otimizações realizadas.

Valendo-se da equação (5.3), como função objetivo, foi executado o AG na estimação de parâmetros do filme fino de VO_2 . Foram utilizados valores "default" da sub-rotina como o número máximo de gerações igual a 500 e número de indivíduos por geração (n) de 100. A probabilidade de cruzamento assumiu valor igual a 0,85. A taxa de mutação possui limites inferiores e superior, respectivamente, iguais a 0,0005 e 0,25 sendo estes os limites da equação (4.7). Ao se alterar os valores do número de gerações e as probabilidades de cruzamento ou o intervalo da taxa de mutação não se gerou resultados com um valor de e menor do que o obtido com uso dos valores "default". Isto indica que a escolha destes parâmetros esta intimamente ligada ao nível de complexidade do problema, sendo difícil estipular valores ideais como os sugeridos no tópico (4.5) deste capítulo.

4.9 Conclusões

Algoritmo Genético (AG) é um método de otimização e busca que faz uso de conceitos de genética, sendo baseada no princípio de seleção natural desenvolvido por Charles Darwin. Um AG geralmente usa regras de transição probabilística para selecionar algumas soluções candidatas à resolução do problema de otimização e outras, para o descarte. Os AG,s utilizam uma versão dos mecanismos básicos observados no processo de evolução natural. Este mecanismo basea-se no cromossomo, que corresponde a cada solução candidata e a um ponto no espaço de busca. Apesar das técnicas empregadas para a busca das soluções apresentem aspectos aleatórios, não implicam que os AG,s sejam globalmente aleatórios. Esta metodologia emprega um sistema de busca heurística adaptativa para a exploração da informações disponíveis para direcionar a busca para a região que contenha os melhores resultados, dentro do espaço de busca. Para a busca da solução são empregados os chamados operadores genéticos. Tais operadores buscam gerar uma população diversificada e que mantenha as características de adaptação adquiridas em gerações anteriores. Existem três operadores: Seleção, Cruzamento ("crossover") e Mutação. O operador de Seleção proporciona selecionar o melhor ou melhores indivíduos que irão dar origem a novas gerações através da reprodução. Através da aplicação

deste operador, cada indivíduo recebe uma probabilidade de reprodução que depende do seu valor correspondente na função objetivo. O operador de cruzamento representa o acasalamento, ocorrendo com uma probabilidade definida pela taxa de cruzamento p_c ($0,6 < p_c < 1,0$). Um valor grande desta taxa permite uma maior exploração do espaço de solução e reduz as chances de convergência para um extremo local. Entretanto, o maior espaço de busca pode ocasionar um esforço computacional desnecessário, devido a exploração de regiões com baixo potencial. O operador de mutação, que é responsável pela introdução de modificações aleatórias nos cromossomos com o objetivo de recuperar a perda de material genético, é imprescindível pois um AG pode funcionar sem realizar cruzamentos mas não sem mutações. Sua função específica é redirecionar o local de busca no espaço de soluções possíveis, uma vez que o operador de cruzamento tem apenas a capacidade de direcionar a busca na direção das melhores soluções. A aleatoriedade introduzida pela mutação favorece a busca ao longo da direção selecionada pelo cruzamento. Além destes operadores, estratégias reprodutivas são vitais para a evolução da população de soluções. Uma destas técnicas é o elitismo, parte integrante do operador de seleção, que faz com o que o algoritmo retenha os melhores indivíduos de cada geração. Esta metodologia baseia-se no fato de que na natureza o indivíduo mais apto, além de se reproduzir mais, possui maior longevidade passando de uma geração para outra.

Capítulo 5

Processamento e Análise

5.1 O Experimento

A histerese na curva característica $R \times T$ do VO_2 é bem conhecida e reportada na literatura (Choi et al., 1996) quanto ao laço principal. Entretanto, quase nenhuma informação está disponível no que se refere a laços menores e sobre a influência da história térmica na trajetória de histerese. Para elaborar um modelo matemático que descreva a relação histerética $R \times T$ no VO_2 é importante a escolha das condições experimentais adequadas para coleta de dados. Também é relevante, o conhecimento mais detalhado destas condições experimentais de forma que seja possível distinguir entre o fenômeno de interesse e as possíveis perturbações introduzidas pela plataforma experimental.

Para estudar as características $R \times T$ no VO_2 , Almeida (2003) desenvolveu um sistema capaz de variar a temperatura do filme segundo funções de ondas arbitrárias, com os seguintes requisitos:

1. Variação de temperatura do filme entre $20^{\circ}C$ e $80^{\circ}C$, faixa assumida como satisfatória para permitir modelar a transição de fase,
2. Capacidade de variar a taxa $\frac{dT}{dt}$ entre 10 e $0,01^{\circ}Cmin^{-1}$ para referências arbitrárias de temperatura,
3. Temperatura ambiente variando entre $10^{\circ}C$ a $40^{\circ}C$
4. Medição de resistência do filme na faixa de 10^2 e $1,2 \times 10^5 \Omega$, faixa considerada satisfatória para os filmes disponíveis no laboratório,
5. Erro de regime de $\pm 0,02^{\circ}C$, necessário para evitar a influência de oscilações na história térmica do filme.

Considerando que a temperatura do ambiente T_a pode ser superior à temperatura desejada do filme, foi necessária a utilização de um módulo termoeletrtrcic (Thermoelectrctric Module - TEM) para permitir o resfriamento do sensor. Este dispositivo consiste de vários semiconduto-

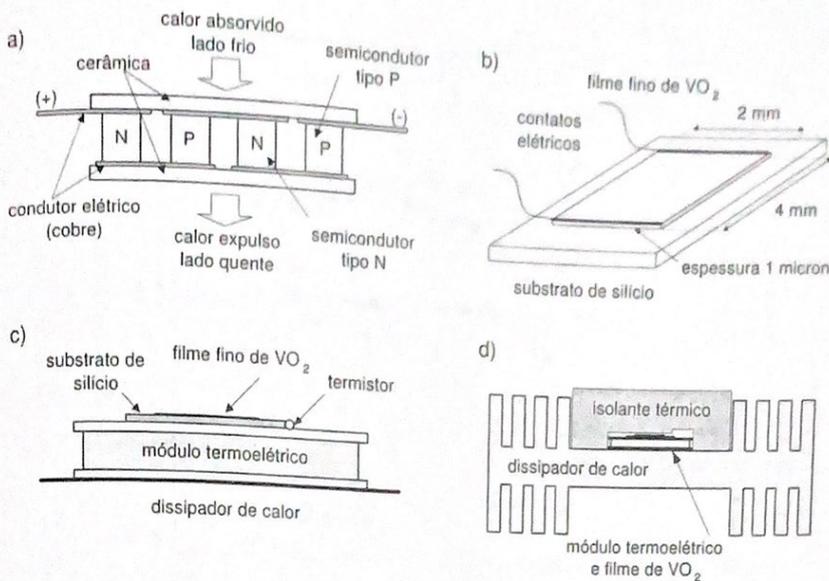


Figura 5.1: Dispositivos empregados e detalhes de montagem da câmara térmica: (a) módulo termoelétrico; (b) dimensões dos filmes utilizados nos experimentos; (c) montagem do filme e termistor sobre o módulo termoelétrico; e d) conjunto da câmara térmica. Fonte: Almeida (2003)

res conectados em série e prensados entre duas placas de cerâmica (vide figura 5.1(a)). Quando conectados a uma fonte de alimentação contínua, a corrente passa pelos elementos produzindo o bombeamento de calor de um lado para o outro. Como conseqüência, se cria um lado quente e outro frio. Se a corrente é revertida, o sentido do bombeamento também muda.

As amostras de filmes finos de VO_2 utilizadas nos experimentos, realizados por Almeida (2003), foram fornecidas pelo professor Igor A. Khrebtov do S.I. Vavilov State Optical Institute, St. Petersburg, Rússia. Estas amostras têm uma espessura de 1000Å e foram depositadas em substratos quase isolantes de Silício, cujas dimensões são apresentadas na figura 5.1(b). A amostra foi montada sobre o TEM, dentro de uma estrutura termicamente isolada, conforme as figura 5.1(c) e (d). A temperatura do filme foi medida utilizando-se um termistor com cerca de $0,5\text{mm}$ de diâmetro, disposto em contato com o substrato do filme e a superfície do dispositivo termoelétrico, através de uma pasta térmica.

O TEM utilizado é de fabricação da Melcor Corporation, modelo CP 1.0-31-06L, com corrente e tensão máxima de $3,0\text{A}$ e $3,75\text{V}$. A capacidade máxima de bombeamento de calor é de $6,3\text{ Watts}$. Este modelo pode alcançar, no vácuo, uma diferença de temperatura máxima de 63°C entre as faces fria e quente. Em condições experimentais de campo, esta diferença é reduzida. No experimento efetuado por Almeida (2003) foi obtida uma diferença máxima de 40°C entre as faces do dispositivo e coletadas mais de 10.000 medidas de resistência e temperatura para o filme de VO_2 . Tal amostragem serviu de base para as otimizações do modelo Algebrico.

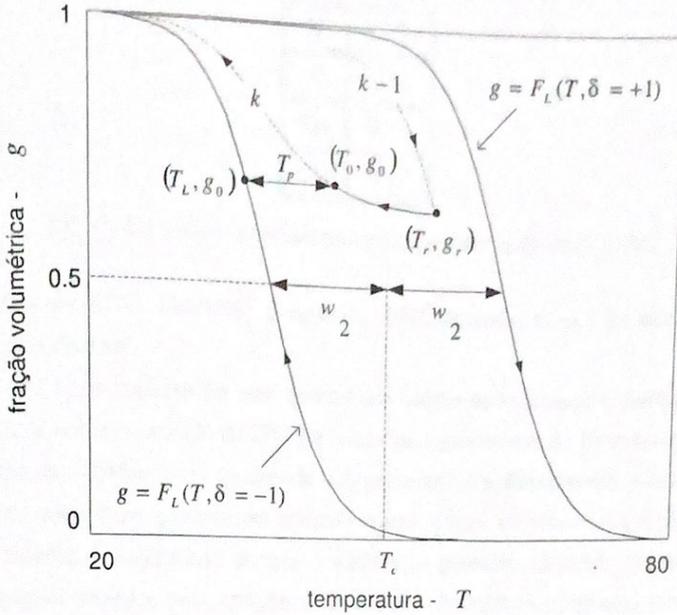


Figura 5.2: Representação esquemática da histerese na evolução da fração volumétrica com a temperatura. Esta construção geométrica de uma curva reversa ilustra o conceito de proximidade de um ponto nesta curva ao laço principal. Fonte Almeida (2003)

Igualmente a construção prévia deste modelo se baseou nesta mesma amostragem.

5.2 Problema de Otimização

De acordo com o modelo L^2P , descrito no capítulo 2, combinando as equações (2.21), (2.23), (2.24) e (2.26) a função $R(T)$ pode ser escrita em função dos parâmetros T_c , β , γ e w . Para a construção da função objetivo do problema de otimização que determina tais variáveis do modelo L^2P , as seguintes relações foram utilizadas:

$$e = \sqrt{\frac{\sum_{i=1}^N (R(i) - R_{esp}(i))^2}{\sum_{i=1}^N R^2(i)}} \quad (5.1)$$

$$\chi^2 = \sum_{i=1}^N \left[\frac{R(i) - R_{esp}(i)}{\sigma_i} \right]^2 \quad (5.2)$$

$$F = \frac{1}{e} \quad (5.3)$$

Em (5.1) e (5.2), $R(i)$ é o valor experimental da resistência e $R_{esp}(i)$ é o seu valor esperado,

N	e (%)
75	1,84
393	2,10
684	2,12

Tabela 5.1: Erro quadrático médio normalizado para o AG

dado pelo modelo de $R(T)$. Em (5.2), o teste da Chi-Quadrada, $\sigma_1 = 1$ foi adotado para todas as medidas experimentais.

Na equação (5.1), e trata-se no erro quadrático médio normalizado e destina-se a quantificar a discrepância entre o modelo de $R(T)$ e os dados experimentais. Esta relação foi utilizada como parâmetro de referência da qualidade dos parâmetros estimados em Almeida (2003). Na otimização com algoritmo genético se utilizou como função objetivo a equação (5.3). A utilização desta relação é justificável porque o algoritmo genético utilizado efetua maximização global e, conseqüentemente, pela estrutura da função objetivo, o programa deverá minimizar os parâmetros procurados para maximizar a função. Na determinação de parâmetros com os métodos de Levenberg-Marquardt (LM) e Gauss-Newton (GN), foi utilizada a função objetivo (5.2), sendo esta relação "default" das rotinas adotadas no MATLAB, sendo intrínica a estrutura dos métodos. O uso desta última equação, com o algoritmo genético, forneceu um valor do erro e superior aquele encontrado com o uso da combinação de (5.1) e (5.3).

Para analisar a qualidade da determinação dos parâmetros otimizados (PO,s) foram utilizados dois critérios: a discrepância relativa entre as saídas dos programas de otimização, para várias seqüências de dados, os parâmetros de referência (PR,s) e o valor de e . Os valores de referência de T_c e w foram extraídos diretamente das curvas características de $g \times T$, como pode ser verificado na figura 5.2. Com T_c e w , o valor de β pode ser obtido do ajuste das funções $F_1(T, \delta = +1)$ e $F_1(T, \delta = -1)$ nas curvas experimentais ascendentes e descendentes principais respectivamente. Obtidos os valores de w, T_c e β , γ podem ser obtido algebricamente valendo-se de (2.24), em pontos colhidos em uma curva crescente de primeira ordem de $g \times T$. O valor dos parâmetros obtidos através deste procedimento são: $w = 6,5^{\circ}C, T_c = 47,6^{\circ}C, \beta = 0,2^{\circ}C^{-1}$ e $\gamma = 0,85$.

5.3 Processamento de Dados Experimentais

Na otimização com o AG foram obtidos parâmetros para três seqüências de dados com $N = 75, 393$ e 684 pares de resistência elétrica e temperatura. Em uma inspeção nos dados, buscando as temperaturas de reversão, foi verificado que estes correspondiam ao laço principal e este acrescido de uma ($N=393$) e duas ($N=684$) curvas descendentes de primeira ordem, segundo

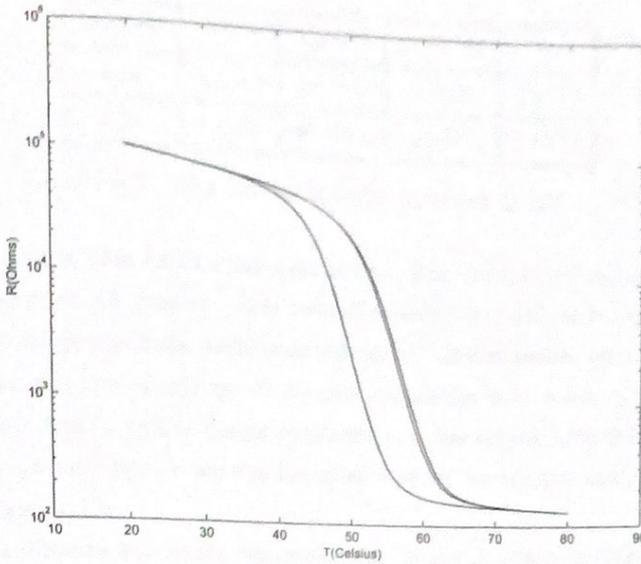


Figura 5.3: Curva de histerese do filme fino de Dióxido de Vanádio para uma amostragem de 684 pares de resistência e temperatura experimentais. Observa-se que o laço principal e as duas curvas de primeira ordem estão praticamente superpostas.

N	w(°C)	T _c (°C)	β(°C ⁻¹)	γ
75	6,35330	46,9140	0,215000	0,900001
393	6,70380	46,8480	0,193700	0,909440
684	6,54780	46,8940	0,194610	0,901900

Tabela 5.2: PO,s fornecidos pelo AG

definição de Mayergoz & Friedman (1988). A escolha destas seqüências visou evitar o uso de uma amostragem limitada ao laço principal, tal como nas otimizações feitas por Schmidt & Guldner (1996) e Lederer et al (1999).

Os limites escolhidos dos parâmetros para otimização genética foram $w = 0, 0$ e $6, 5^{\circ}C$, $T_c = 0$ e $100^{\circ}C$, $\beta = 0, 0$ e $1, 0^{\circ}C^{-1}$ e $\gamma = 0, 0$ e $1, 0$.

Os PO,s para o AG são mostrados na tabela 5.2. Nesta tabela, convencionou-se que os parâmetros são apresentados com o mesmo número de significativos dos valores de temperatura coletados. A temperatura possui o menor número de significativos da amostra (seis). Por uma regra de física experimental, quando um dado é manipulado em operações diferentes das quatro básicas, este deve ter o mesmo número de significativos da medida original (Piacentini et al., 1998). Isto foi empregado para se evitar uma precisão irrealística nos valores dos parâmetros obtidos.

Na tabela 5.3 são apresentados os PO,s obtidos com os métodos determinísticos LM e GN

Método	$w(^{\circ}C)$	$T_c(^{\circ}C)$	$\beta(^{\circ}C^{-1})$	γ
LM	6,1319	46,7391	0,1921	0,8773
GN	6,1319	46,7391	0,1921	0,8773

Tabela 5.3: PO,s fornecidos pelos Métodos de GN e LM

utilizando os PR,s como ponto inicial para otimização. Tais resultados foram obtidos utilizando-se uma amostragem de 75 pontos. Esta restrição pode ser justificada pela falha da rotina "lsqnonlin" para as outras duas seqüências de dados. Inicialmente, esta rotina apresentou mensagens informando que o número de funções calculadas bem como, o de iterações havia sido ultrapassado. Após o ajuste destes parâmetros, a mensagem apresentada foi "divide by zero", indicando que em alguma fase das iterações ocorreu divisão por zero, comprometendo o sucesso da otimização.

Os resultados deste tópico foram divulgados em Betzler & Almeida. (2003) e Betzler et al. (2004).

5.3.1 Discussão e Conclusões

De acordo com os resultados apresentados nas tabelas 5.2 e 5.3, pode-se verificar que a os PO,s, obtidos para vários conjuntos de dados, são próximos aos PR,s. Tal avaliação pode ser feita pela análise da discrepância relativa percentual dos quatro parâmetros:

$$\Delta = \frac{|X' - X''|100\%}{X''} \quad (5.4)$$

Na qual, X' representa o valor dos PO,s e X'' um valor de referência (PR,s) (Argollo et al., 1998). A discrepância teve uma tendência de aumento e posterior redução para três parâmetros nos dados processados com o AG, com um Δ máximo de aproximadamente 7 %. Esta baixa discrepância, em relação ao ponto inicial, pode sugerir que o ponto fornecido para otimização esta próximo do ótimo identificado pelo AG. De fato, isto pode ser verificado no contorno da superfície de $e(\beta, \gamma)$, com $w = 6,5^{\circ}C$ e $T_c = 47,6^{\circ}C$, para a amostragem de 75 pares de resistência-temperatura (figura 5.4). Nesta figura, verifica-se que o ponto para o qual o AG convergiu encontra-se nas cercanias do ponto inicial. Tal tendência é também observada nas outras amostragens processadas com o AG e para os métodos determinísticos. Nos métodos de LM e GN, os PO,s tiveram um Δ inferior a 6%.

Na tabela 5.1, verifica-se que os valores de e para os POs foram inferiores ao e obtido com os PR's. Tal diferença pode ser justificada pela amostragem utilizada para o cálculo do $e(2,5\%)$. Na estimativa dos PRs foi utilizada um conjunto de dados que privilegiava o número de laços das curva de primeira ordem em detrimento do número de pontos o qual, embora grande, não

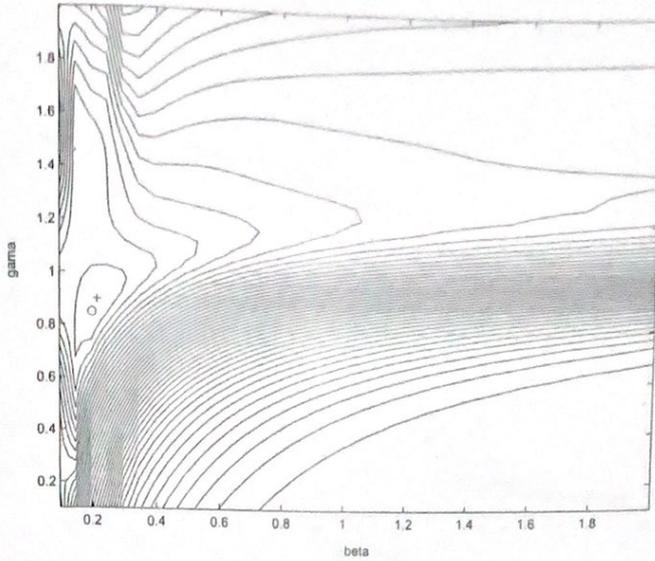


Figura 5.4: Mapa de contorno da superfície $e(\beta, \gamma)$ para 75 pares resistência-temperatura. O círculo (o) o ponto inicial para a otimização com o AG enquanto a cruz (+) o ponto ótimo para onde houve convergência. Nota-se que ambos encontram-se no interior de um curva de nível de um possível mínimo local.

Método	$e(\%)$	Iterações
LM	1,43	9
GN	1,43	14

Tabela 5.4: e e Iterações para os Métodos de GN e LM

era o total ao longo de cada laço. Tal característica pode ser verificada pelo valor de e obtido para uma amostragem de 75 pontos com os PR's (3,6%).

A qualidade do ajuste com os PO,s e o conjunto de referência pode ser verificada nas Fig. (5.5), (5.6) e (5.7). Comparando as Figuras (5.7) e (5.6) com (5.5), apesar da pequena diferença apresentada, pode-se sugerir que T_c e, possivelmente, w não foram adequadamente estimados pelo processo algébrico ou pela otimização com os métodos determinísticos. Para a otimização com os métodos de LM e GN tal situação pode ser decorrente de se efetuar uma busca em uma direção condicionada pelo gradiente local a cada iteração k . Para esta metodologia de busca, existe a possibilidade de parada do método em pontos de mínimos local. Os estudo do espaço de busca e sua relação com falhas de convergência será feita no tópico (5.4), com os dados gerados pelo modelo L^2P .

Na Tabela 5.4, é mostrado o e obtido com os métodos determinísticos utilizando como ponto inicial a solução fornecida pelo AG. Nesta tentativa de combinação dos métodos de otimização,

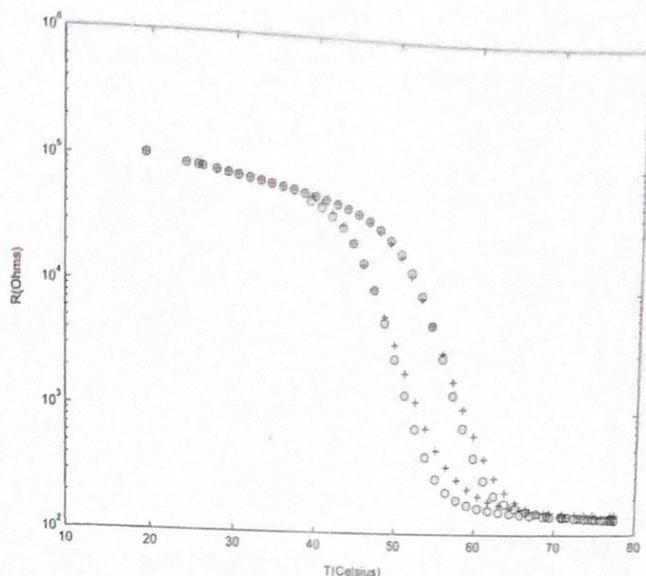


Figura 5.5: Curva $R \times T$ do filme fino de Dióxido de Vanádio. As cruces (+) representam o modelo com parâmetros otimizados pelo AG enquanto os círculos (o), os dados experimentais, para uma amostra de 75 pontos.

se percebe uma pequena redução no valor de e se comparado com o equivalente da tabela 5.1. Entretanto, isto não caracteriza que o uso destes PO,s seja o responsável principal pela sutil melhoria do e . Com o uso dos PR,s, o método LM convergiu para o mesmo conjunto expresso na tabela 5.3.

A falha no processamento com os métodos LM e GN nas amostragens $N=396$ e 696 , pode sugerir que o resíduo não é desprezível. Nesta circunstância, a aproximação de $G(x)$ (eq. 3.34) não é válida e a matriz Hessiana pode não ser definida ou semi-definida positiva. Isto pode implicar em um aumento excessivo do número de iterações ou mesmo em uma falha na convergência. De fato, o maior número de iterações para o método de GN, se comparado com o de LM, implica que o resíduo (eq. 3.31) não foi pequeno, pois nesta condição o método possui uma rápida convergência (Fletcher, 2001).

Outra possibilidade para a falha é a escolha de um ponto inicial inadequado para a realização do processo de otimização. Tal ponto inicial distante de uma possível solução global também pode gerar problema com aproximação da Hessiana e levar a problemas de convergência como os anteriormente mencionados.

A influência do resíduo e do ponto inicial de busca serão investigados em detalhes no tópico (5.4), valendo-se de dados gerados pelo modelo L^2P .

Na tabela 5.4, verifica-se que o e é igual para os métodos de GN e LM, entretanto, na amostragem considerada, o método LM exigiu um menor número de iterações.

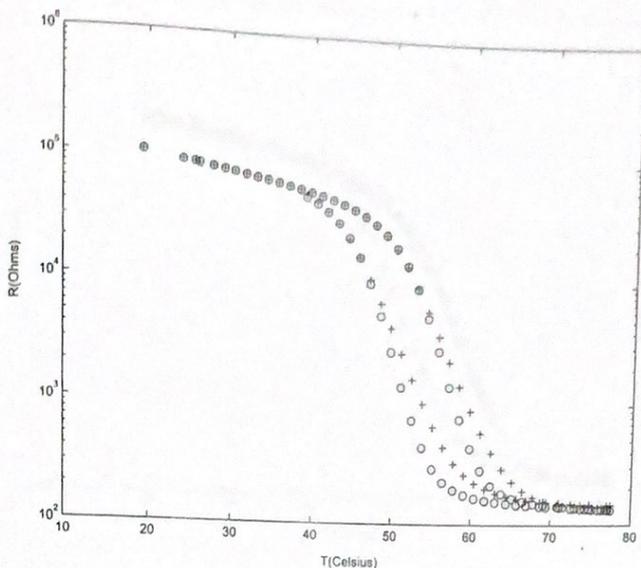


Figura 5.6: Curva $R \times T$ do filme fino de Dióxido de Vanádio. As cruces (+) representam o modelo com parâmetros otimizados pelos métodos de GN e LM enquanto os círculos (o), os dados experimentais, para uma amostra de 75 pontos.

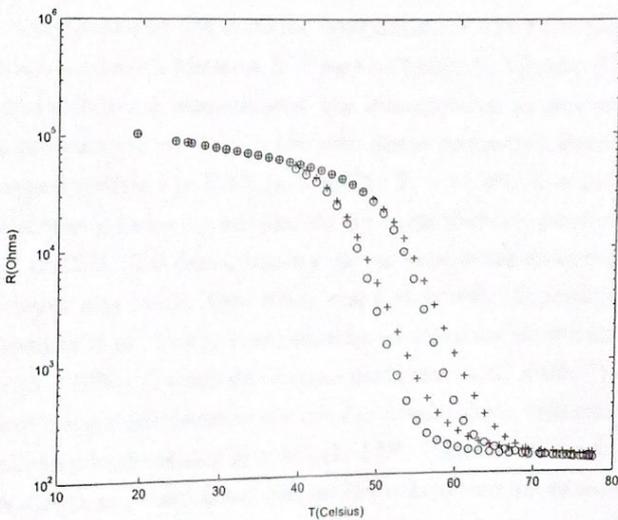


Figura 5.7: Curva $R \times T$ do filme fino de Dióxido de Vanádio. As cruces (+) representam o modelo com os Parâmetros de Referência enquanto os círculos (o), os dados experimentais, para uma amostra de 75 pontos.

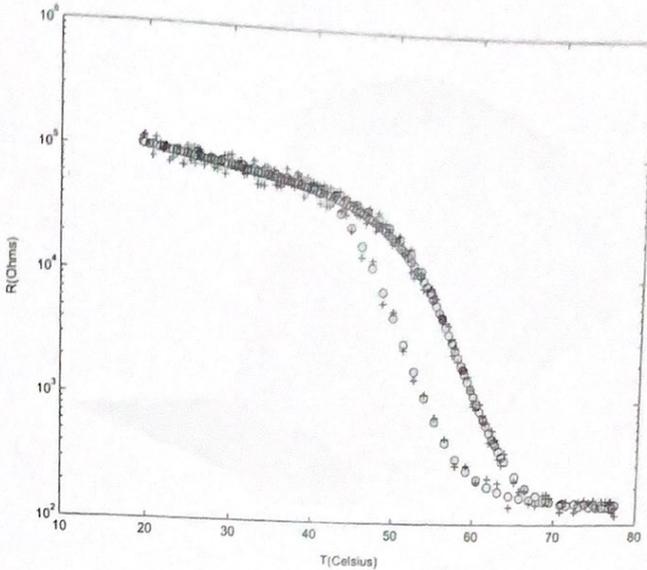


Figura 5.8: Curvas de histerese do filme fino de Dióxido de Vanádio com 198 pares resistência-temperatura. Os círculos (o) e as cruzes (+) representam, respectivamente, as resistências elétricas geradas pelo modelo L^2P sem e com um desvio relativo percentual de 20%.

5.4 Processamento de Dados Gerados pelo Modelo L^2P

Afim de analisar a convergência dos métodos determinísticos LM e GN foram gerados dados artificiais utilizando o modelo de histerese L^2P para o Dióxido de Vanádio. Os dados foram 198 pares de resistências elétricas e temperaturas, que correspondem ao laço principal e parte da primeira curva de primeira ordem. Para construção destas curvas de histerese foram utilizadas as temperaturas experimentais e os PR'S ($w = 6,5^{\circ}C$, $T_c = 47,6^{\circ}C$, $\beta = 0,2^{\circ}C^{-1}$ e $\gamma = 0,85$). Nas resistências elétricas geradas foi introduzido um ruído aleatório que ocasionou um desvio relativo percentual de 20%. Tal desvio implica que as resistências estão compreendidas entre 80%-120% de seu valor sem ruído. Este ruído, que é decorrente de perturbações estatísticas imprevisíveis (Piacentini et al., 1998), é bem superior ao dos dados experimentais coletados por Almeida (2003) ($\sim 5 - 10\%$). O ruído foi criado a partir do uso da rotina "rand" do MATLAB 6.5, que gera um número aleatoriamente entre 0 e 1. A saída desta rotina foi multiplicada por 0.2 e pelo valor da resistência elétrica gerada pelo L^2P . Uma análise condicional ($\text{rand}(1) < 0,5$?) foi utilizada definir se o valor anteriormente definido deveria ser somado ou subtraído da resistência gerada pelo modelo. O objetivo da introdução deste ruído é aumentar o resíduo r_i ; a ocorrência de pontos críticos e verificar se estes são os motivos para convergência inadequada ou a falha dos métodos determinísticos.

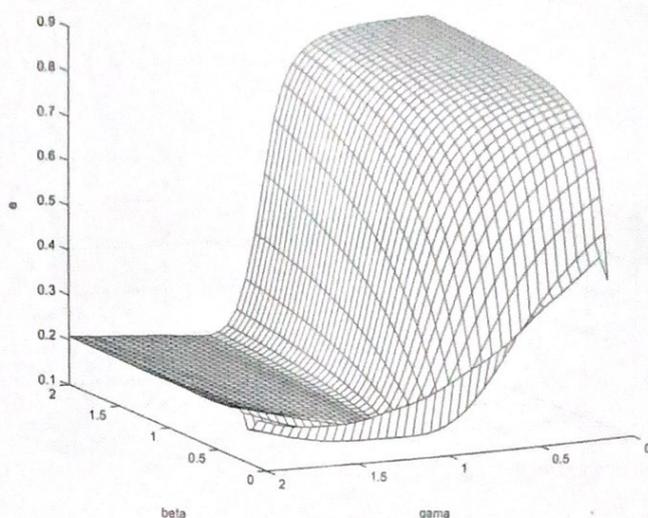


Figura 5.9: Superfície de erro $e(\beta, \gamma)$ para os dados artificiais. Tal superfície é análoga a sua equivalente gerada a partir dos dados experimentais com uma amostragem de 75 pontos.

5.4.1 Primeiro Processamento

Os dados ruidosos gerados com o modelo L^2P foram otimizados com os métodos LM e GN, utilizando-se como ponto inicial os PR,s. Para os métodos LM e GN houve convergência para o mesmo ponto $w = 6,2885^{\circ}C$, $T_c = 47,4799^{\circ}C$, $\beta = 0,2361^{\circ}C^{-1}$ e $\gamma = 1,1411$. Em ambos, o e registrado foi de 12,99%.

5.4.2 Segundo Processamento

A convergência dos métodos de LM e GN esta criticamente ligada a escolha adequada de um ponto para início da otimização. Em função disto, foram criadas as superfícies de erro $e(\beta, \gamma)$ e $e(w, T_c)$ para a busca de regiões onde máximos e mínimos locais que possam ocasionar falhas de convergência nos métodos determinísticos. A superfície $e(w, \gamma)$ (fig.5.11) e $e(\beta, T_c)$ não apresentam a complexidade observada em $e(\beta, \gamma)$ e $e(w, T_c)$, implicando que este experimento e o seguinte se restringiram as duas últimas superfícies de erro. A superfície $e(\beta, \gamma)$ foi construída valendo-se dos parâmetros otimizados $w = 6,2885^{\circ}C$, $T_c = 47,4799^{\circ}C$, variando-se β e γ , em 40 passos regulares, entre 0 e $2^{\circ}C$.

Comparando-se as figuras (5.9) e (5.10) pode-se concluir que no plano $\beta \times \gamma$ existe um ponto de máximo nas cercanias de $\beta = 1,8^{\circ}C^{-1}$ e $\gamma = 0,2$ e um mínimo próximo de $\beta = 0,3^{\circ}C^{-1}$ e $\gamma = 1,95$, ambos locais. A partir destes pontos foi realizada a otimização, valendo-se dos métodos de LM e GN, com os valores ótimos $w = 6,2885^{\circ}C$, $T_c = 47,4799^{\circ}C$. Com a utilização

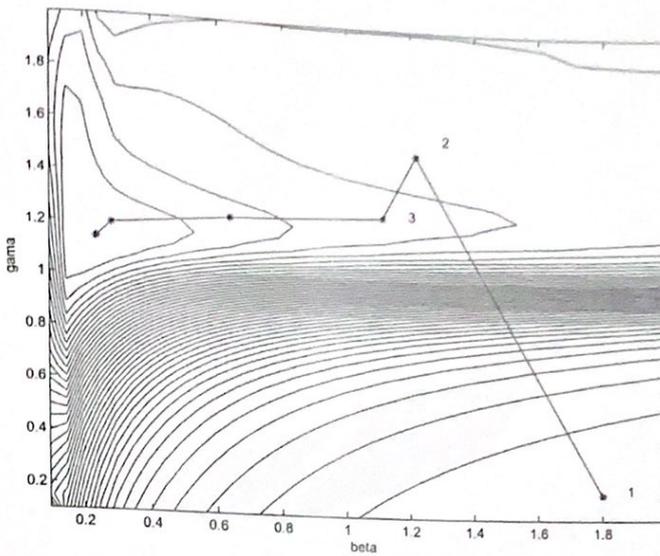


Figura 5.10: Mapa de contorno da superfície de $e(\beta, \gamma)$ mostrando as iterações do método LM a partir de $\beta = 1,8^{\circ}C^{-1}$ e $\gamma = 0,2$

dos valores ótimos de w e T_c , a otimização será feita em duas dimensões evitando que uma busca através de uma hipersuperfície previna uma possível falha de convergência no domínio de β e γ . Para o ponto $\beta = 1,8^{\circ}C^{-1}$ e $\gamma = 0,2$, os métodos de GN e LM não apresentaram convergência adequada.

No processamento com o método GN, após a iteração 1, a rotina "lsqnonlin" modificou o algoritmo para o LM com a apresentação da mensagem "Conditioning of Gradient Poor - Switching To LM method". Na iteração 2, o método apresentou a falha quando o número máximo de funções permitidas foi alcançado. O ajuste deste último parâmetro para um valor maior não ocasionou convergência e novamente a mensagem de erro foi apresentada. Com o método de LM cessou a otimização, após 13 iterações, com o aparecimento da mensagem "No improvement in search direction: Terminating", no ponto $\beta = 0,2361$ e $\gamma = 1,1411$ (figura 5.10). O mesmo foi observado para $\beta = 0,3^{\circ}C^{-1}$ e $\gamma = 1,95$ (figura 5.12). Para o método LM, após 16 iterações, o número máximo de funções calculadas foi novamente atingido. O valores obtidos na iteração 16 foram de $\beta = 0,1350^{\circ}C^{-1}$ e $\gamma = 1,3982$. O método de GN apresentou falha equivalente após a iteração 16. Neste mesmo processo, após a iteração 9, a rotina do MATLAB passou a utilizar o método de LM. Neste processamento foram obtidos $\beta = 0,1350^{\circ}C^{-1}$ e $\gamma = 1,3985$.

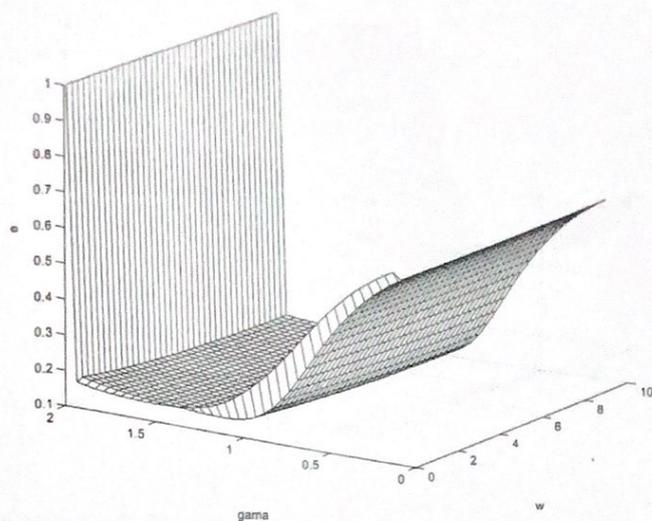


Figura 5.11: Superfície de erro $e(w, \gamma)$ para os dados artificiais.

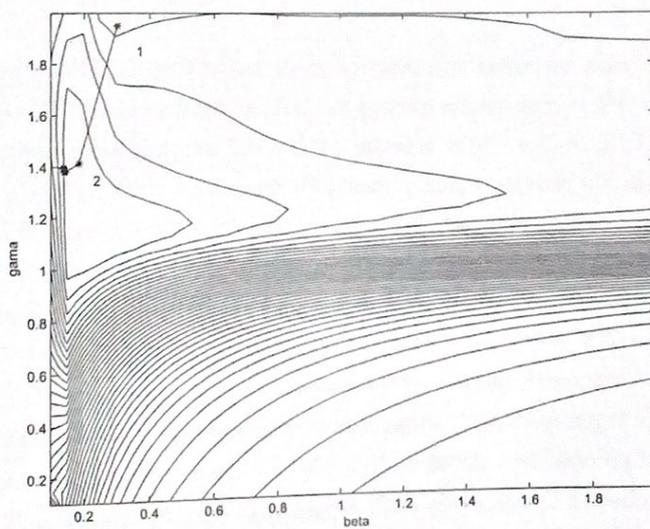


Figura 5.12: Mapa de contorno da superfície de $e(\beta, \gamma)$ mostrando as iterações do método LM a partir de $\beta = 0,30 C^{-1}$ e $\gamma = 1,95$

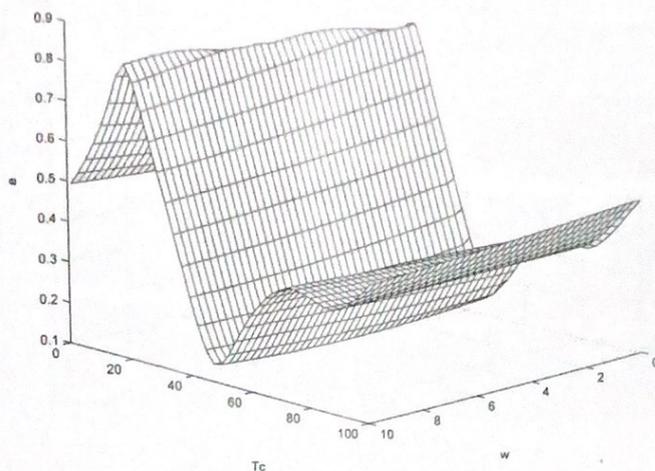


Figura 5.13: Superfície de erro $e(w, T_c)$ para os dados artificiais.

5.4.3 Terceiro Processamento

A superfície $e(w, T_c)$ (figura 5.13) foi construída valendo-se dos parâmetros otimizados $\beta = 0,2361^{\circ}C^{-1}$, $\gamma = 1,1411$, variando-se w entre 0 e $10^{\circ}C$ e T_c entre 0 e $100^{\circ}C$, ambos em 40 passos regulares.

Através das figuras (5.13) e (5.14), pode-se verificar que existe um máximo global próximo de $w = 5^{\circ}C$ e $T_c = 20^{\circ}C$ e um mínimo global nas proximidades de $w = 6^{\circ}C$ e $T_c = 50^{\circ}C$. Para verificar a convergência dos métodos LM e GN o ponto $w = 5^{\circ}C$ e $T_c = 10^{\circ}C$, mantendo-se fixo $\beta = 0,2361^{\circ}C^{-1}$ e $\gamma = 1,1411$ foi utilizado. Para este ponto, o método LM apresentou falhas de convergência, tendo ultrapassado o número permitido de funções calculadas, além da mensagem "divide by zero" a partir da sétima iteração. O processamento foi interrompido no ponto $w = -0,7261^{\circ}C$ e $T_c = -1,0365^{\circ}C$, em 8 iterações, com um $e = 45,39\%$. Para o mesmo ponto inicial, o método de GN convergiu para o mínimo local $w = -4340,5^{\circ}C$ e $T_c = -2168,8^{\circ}C$, com $e = 57,49\%$. Entretanto, para este processamento, a rotina "lsqnonlin" mudou do método GN para o LM, na iteração 6, apresentando a mensagem "Conditioning of Gradient Poor".

Utilizando como estimativa inicial $w = 6^{\circ}C$ e $T_c = 50^{\circ}C$, o método de LM convergiu para $w = 5,4002^{\circ}C$ e $T_c = 46,6727^{\circ}C$ com 11 iterações (fig. 5.14 e 5.15). O método de GN apresentou falhas de convergências análogas as observadas com o ponto $w = 5^{\circ}C$ e $T_c = 10^{\circ}C$, tendo sido modificado o método a partir da iteração 7. Não houve uma convergência adequada, pois o número máximo de funções calculadas foi novamente ultrapassado. O LM, neste caso, forneceu o valor $w = 5,3993^{\circ}C$ e $T_c = 46,6732^{\circ}C$ após a iteração 15, quando o número máximo de

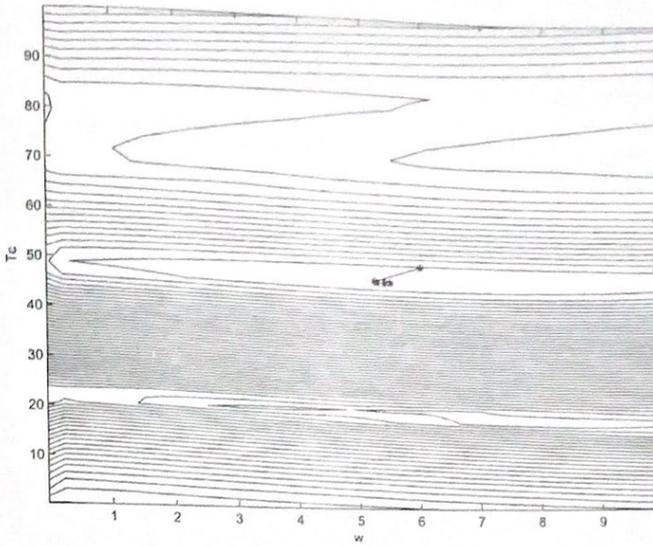


Figura 5.14: Contorno da superfície de erro $e(w, T_c)$ para os dados artificiais mostrando as iterações do método LM a partir de $w = 6^0 C$ e $T_c = 50^0 C$.

funções calculadas foi atingido.

5.4.4 Conclusões

Pelos resultados obtidos no primeiro experimento, o ruído introduzido ocasionou em um aumento de e , se comparado com os obtidos pelo AG em amostragens com um maior número de elementos, com os dados experimentais (tabela 5.2). A introdução do ruído incrementa o resíduo, ocasionando que a convergência do método LM seja mais lenta que a do GN. Verifica-se com o segundo e terceiro experimento que a escolha de um ponto adequado para o início da otimização é crítico para o sucesso do processamento com os métodos determinísticos aplicados. Em especial, quando $x^{(1)}$ é próximo do ponto ótimo x^* e matriz $\sum_{i=1}^m r_i^* \nabla^2 r_i^*$ é pequena, sendo válida a aproximação de $G(x)$. Este motivo está associado a apresentação da mensagem de erro "número máximo de funções calculadas atingido". Esta mensagem implica que o algoritmo de otimização ficou confinado nas proximidades de um ponto subcrítico. O número excessivo de funções calculadas está associado a tentativa do método em sair deste extremo local, variando o passo $\alpha^{(k)}$ e a direção de busca $d^{(k)}$.

A escolha do ponto inicial passa por uma inspeção nas superfícies de erro. No caso, as superfícies $e(\beta, \gamma)$ e $e(w, T_c)$ por apresentarem uma maior complexidade, com um maior número de regiões com máximos e mínimos locais. O resíduo não se caracteriza como a razão pela falha dos métodos uma vez que com amostragens envolvendo dezenas ou centenas de dados, o valor nominal dos r_i apresentou a mesma ordem de grandeza. Outra mensagem apresentada

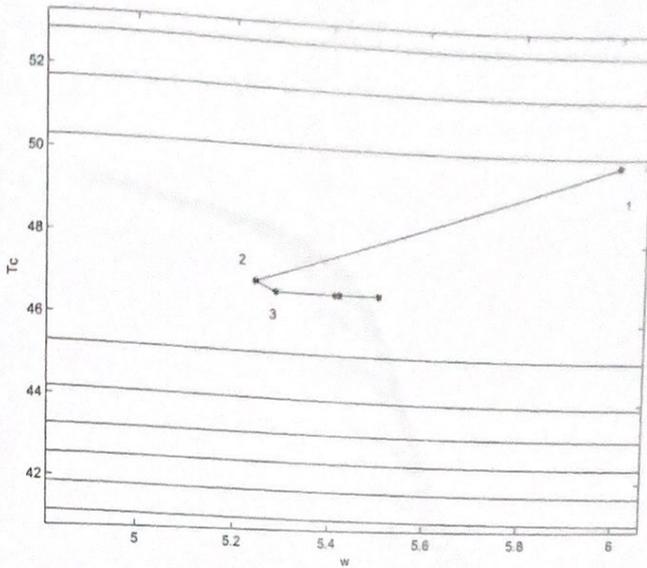


Figura 5.15: Detalhe da superfície $e(w, T_c)$ mostrando as iterações do método LM a partir de $w = 6^{\circ}C$ e $T_c = 50^{\circ}C$.

"Conditioning of Gradient Poor - Switching To LM method" condiciona que a matriz Jacobiana possui colunas linearmente dependentes. Este fato foi de fato observado com a inspeção do valor da Jacobiana na otimização da amostra com $w = 5^{\circ}C$ e $T_c = 10^{\circ}C$, antes da mudança para o LM. Como a direção de busca para o método GN utiliza a Jacobiana e sua transposta, a rotina adiciona um múltiplo da matriz identidade I a este produto, prevenindo a geração de uma matriz singular. Esta adição torna a convergência do método mais lenta. A singularidade do produto da Jacobiana e sua transposta pode ser originária de uma escolha de um ponto inicial inadequado ou a uma sobreparametrização do problema. A mensagem "divide by zero", ocasionada por uma divisão por zero, provavelmente no cálculo da direção de busca, é uma consequência da singularidade do produto da Jacobiana e sua transposta.

5.5 Otimização de Outras Curvas de Primeira Ordem

O modelo de Preisach permite o ajuste de curvas de transição de primeira ordem, ligadas a parte ascendente do laço de histerese (figura 1.1). Para esta descrição, as funções $\mu(\alpha, \beta)$ são obtidas experimentalmente a partir das curvas de primeira ordem. Entretanto, o filme fino de VO_2 apresenta uma histerese assimétrica na qual a adaptação do modelo de Preisach é imprecisa, principalmente na transição de fase (Almeida, 2003). Nos dados experimentais disponíveis, as curvas descendentes de primeira ordem estão presentes em uma amostragem de 1807 pares de resistência e temperatura, em intervalos definidos por Almeida (2003). Tais intervalos geram o

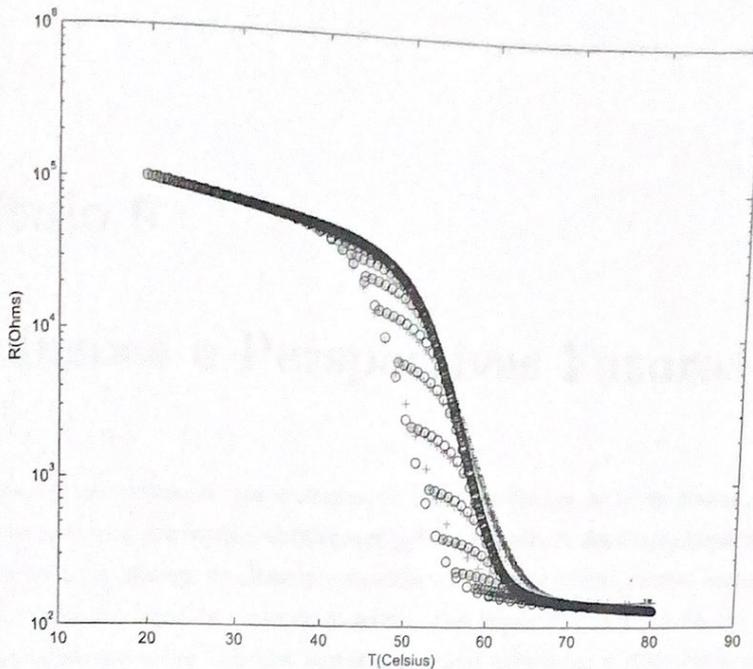


Figura 5.16: Curvas de primeira ordem para o filme fino de VO_2 . Os círculos (o) representam os dados experimentais e as cruzes (+) o modelo com parâmetros estimados pelo AG.

ramo ascendente e descendente da curva de histerese bem como as curvas de primeira ordem. Utilizando esta amostra, os métodos determinísticos GN e LM e os PR,s como ponto inicial forneceu, após uma iteração, em uma otimização bem sucedida, novamente os PR,s. Isto sugere que o ponto escolhido por Almeida (2003) é extremo local, fornecendo um ajuste que atende ao critério de parada dos métodos determinísticos. Para o AG, utilizando uma população de dez soluções candidatas, foi obtido $w = 6,43250^{\circ}C$, $T_c = 46,7000^{\circ}C$, $\gamma = 0,831880$ e $\beta = 0,202111^{\circ}C^{-1}$, com um $e = 2,61\%$. Tal valor é um pouco maior que o obtido em Almeida (2003)(2,5%), o que sugere que o AG convergiu para um possível mínimo local nas proximidades do ponto ótimo indentificado por Almeida (2003). Verifica-se que a solução encontrada pelo AG fornece um ajuste satisfatório das curvas de primeira ordem (figura 5.16).

Capítulo 6

Conclusões e Perspectivas Futuras

Pelos resultados, demonstrou-se que a otimização com os métodos de busca linear geram PO,s com um e pouco menor que aqueles obtidos com o AG. Os métodos determinísticos falharam em amostragens com um número de elementos superior a do laço principal, tais as correspondentes aos segundo e terceiro laços da curva de histerese. Tais falhas de convergência são reportadas na literatura existente sobre métodos numéricos, sendo associadas a algoritmos com ótimas propriedades de convergência e simplicidade de implementação, como o método de Newton. Isto implica que métodos iterativos tradicionais podem não garantir a convergência para a raiz procurada, devido um cancelamento que gere divisão por zero, principalmente em funções mal comportadas.

O sucesso do AG, em uma grande amostragem e trabalhando na modelagem de função complexa como a equação (5.3), confirmam a característica desta metodologia de dificilmente falhar. Entretanto, os AG,s apresentam um esforço computacional grande e taxa de convergência muito lenta.

Os valores de e encontrados colaboram com uma análise equivalente efetuada em Almeida (2003), que concluiu que o modelo L^2P é adequado para a descrição da histerese em filme finos de Dióxido de Vanádio. Neste trabalho, tal conclusão foi obtida devido ao baixo valor de e encontrado para uma amostragem selecionada, que apresenta curvas decrescentes de primeira ordem.

Uma possível extensão a este trabalho é a caracterização de uma amostra de filme de V_0_2 , através de otimização, que se destinará à construção de um bolômetro, e utilizá-la para comparar a resposta prevista com a experimental.

Capítulo 7

Bibliografia

- Almeida, L.A.L., Deep, G.S., Lima, A.M.N. and Neff, H. (2000). Modeling a magnetostrictive transducer using genetic algorithm, ICM 2000 - International Conference on Magnetism.
- Almeida, L.A.L., Deep, G.S., Lima, A.M.N. and Neff, H. (2001). Modeling a magnetostrictive transducer using genetic algorithm, *Journal of Magnetism and Materials* 226: 1262-1264.
- Almeida, L.A.L., Deep, G.S., Lima, A.M.N. and Neff, H. (2002). Modeling of the hysteresis Metal-Insulator Transition in a vanadium dioxide infrared detector. *Optical Engineering* 41 (10): 2582-2588
- Almeida, L. A. L., Deep, G.S., Lima, A. M. N. and Neff, H. (2003a). Um Modelo Algébrico para Histerese Magnética, *Revista de Controle & Automação*, Vol. 14, pp. 58-68.
- Almeida, L. A. L., Deep, G.S., Lima, A. M. N. and Neff, H. (2003b). Limiting Loop Proximity (L^2P) Hysteresis Model, *IEEE Transactions on Magnetics*, vol. 39, no.1, pp. 1-6.
- Almeida, L. A. L. (2003). Modelo de Histerese para a Transição Semicondutor-Metal em Filmes Finos de V_0_2 . Tese de Doutorado, Departamento de Engenharia Elétrica, UFPA, Campina Grande.
- Argollo, R. M., Ferreira, C. and Sakai, T. (1998). Física Geral e Experimental - Apostila de Teoria dos Erros, Departamento de Geofísica Nuclear, Instituto de Física, UFPA, Salvador.
- Basso, V. and Bertotti, G. (1996). Hysteresis models for the description of domain wall motion, *IEEE Transactions on Magnetics* 32(5): 4210-4212.

- Basso, V., Bertotti, G., Serpico, C. and Visone, C. (1998). Application of an exactly invertible hysteresis model to magnetic field computations. *Jornal de Physique IV*, 8(2): 639-642.
- Bertotti, G. (1999). *Hysteresis in Magnetism*. Academic Press.
- Beasley, D., Ralph, R.M. and David, R.B. (1993). An overview of genetic algorithms: Part 2, Research Topics. *University Computing*, 15, n.4, p.170-181.
- Bennett, A.P. (1997). Finite population effects for ranking and tournament selection. *Complex System*, v.11, p. 1-24.
- Brokate, M. and Sprekels, J. (1996). *Hysteresis and Phase Transitions*, Springer-Verlag.
- Betzler, A.S. and Almeida, L.A.L. (2003). Utilização de algoritmo genético na caracterização de filmes finos de VO_2 . VI Simpósio Brasileiro de Automação Inteligente, Bauru.
- Betzler, A.S., Almeida, L.A.L., Santos, D. E., Esperidião, A. S., Celedon, J.Q. and Martinez, L. (2004). Caracterização de filmes finos de VO_2 através de algoritmos genéticos e métodos tradicionais de otimização. XV Congresso Brasileiro de Automação, Gramado.
- Brandão, G.B., Almeida, L.A.L., Deep, G.S., Lima, A.M. and Neff, H. (2001a). Instability and dynamics of thermal runaway of thermoresistive microsensors. *IEEE Instrumentation and Measurement Technology Conference 3*: 1467-1472.
- Brandão, G.B., Almeida, L.A.L., Deep, G.S., Lima, A.M. and Neff, H. (2001b). Stability conditions, nonlinear dynamics, and thermal runaway in microbolometers. *Journal of Applied Physics* 90 (4): 1999-2008.
- Castilho, V.C. (2003). Otimização de Componentes de Concreto Pré-Moldado Protendidos Através de Algoritmos Genéticos. Tese de Doutorado. Escola de Engenharia de São Carlos, USP. São Carlos.
- Charbonneau, P. and Knapp, B. (1995). *A User's Guide to PIKAIA 1.0*. NCAR/TN-418+IA, National Center for Atmospheric Research, Boulder, CO.
- Charbonneau, P., Tomczyk, S., Schou, J. and Thompson, M. (1998). The Rotation of Solar Core Inferred by Genetic Forward Modeling, *Astrophys. J.* 496, 1015.
- Choi, H., Ahm, J., Jung, J., Noh, T. and Kirn, D. (1996). Mid-infrared properties of a VO_2 near the metal-insulator transition, *Physical Review B*, 54 (7): 4621-4628.

- Coley, D.A. (1999). *An introduction to genetic algorithms for scientists and engineers*. World Scientific. Cingapore.
- Dantzig, G. (1963). *Linear Programming and Extensions*, Princeton University Press, Princeton.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*, John Murray (ed), London.
- da Silva, E.E. (2001). *Otimização de Estruturas de Concreto Armado Utilizando Algoritmos Genéticos*, Dissertação de Mestrado, Departamento de Engenharia de Estruturas e Fundações, Escola Politécnica, USP, São Paulo, SP.
- Davis, L. (1991). *Handbook of Genetic Algorithms*. Van Nostrand Reinhold (ed). New York, NY.
- Duhem, P. (1897). Die dauernden änderungen und die thermodynamik i, *Z. Phys.Chemie* 22: 543-589.
- Dupré, L. R.; van Keer, R. and Melkebeek, J. A. A. (1999). Identification of the relation between the material parameters in the Preisach model and in the Jiles-Atherton hysteresis model, *Journal of Applied Physics*, Vol 85, Issue 8, pp.4376-4378.
- Fabian, K. and Von Dobeneck, T. (1997). Isothermal magnetization of samples with stable Preisach function: A survey of hysteresis, remanence, and rock magnetic parameters. *Journal of Geophysical Research*, Vol102, Issue B8, pp.17659-17678.
- Fletcher, R. (2001). *Practical Methods of Optimization*. Wiley-Interscience, Chichester.
- Fontana, M. (2001). *Caracterização e modelagem de propriedades óticas de sensores de dióxido de vanádio*. Dissertação de Mestrado. Departamento de Engenharia Elétrica. UFPB. Campina Grande. PB.
- Fontana, M., Almeida, L.A.L., Lima, A.M.N. and Deep, G.S. (2001). *Caracterização e modelagem das propriedades óticas de sensores de dióxido de vanádio*. II Seminário Nacional de Controle e Automação. Salvador.
- Galinaitis, W. S. and Poling, J. W. (2003). Modeling the hysteresis of a simple SMA wire actuator with a load-dependent Preisach-type operator. *Smart Structures and Materials 2003: Active Materials: Behavior and Mechanics*. Edited by Lagoudas, Dimitris C. Proceedings of the SPIE, Vol. 5053, pp. 305-316.

- Gen, M. and Cheng, R. (1997). *Genetic algorithms and engineering-design*. New York, John Wiley.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, MA.
- Harries, T. and Howarth, I. (2000). Spectropolarimetric Orbits of Symbiotic Stars: Five S-type Systems, *Astron. & Astrophys.* 361, 139
- Hernandez, E.D.M., Muranaka, C.S. and Cardoso, J.R. (2000). Identification of the Jiles-Atherton model parameters using random and deterministic searches. *Physica B: Condensed Matter*, Vol. 275, Issues 1-3, pp.212-215.
- Hodgdon, M.L. (1988). Mathematical theory and calculation of magnetic hysteresis curves, *IEEE Transactions on Magnetics* 24(6): 3120-3122.
- Holland, J.H. (1975). *Adaptation in natural and artificial system*. Ann Arbor, University of Michigan Press.
- Jerominek, H., Renaud, M., Swart, N. R., Picard, F., Pope, T. D., Levesque, M., Lehoux, M., Biladeau, G., Pelletier, M., Audet, D. and Lambert, P. (1996). Micromachined v02-based Uncooled IR Bolometric Detector Arrays with Integrated CMOS Readout Electronics, *Proc. SPIE*, Vol. 2882, pp. 111-121.
- Jiles, D. and Atherton, D (1983). Ferromagnetic Hysteresis, *IEEE Transactions on Magnetics*, 19 (5), pp. 2183-2185.
- Jiles, D.C. and Atherton, D.L. (1984). Theory of ferromagnetic hysteresis. *J. Appl. Phys.* vol 55. no.6, pp. 2115-2120
- Jiles, D.C. and Atherton, D.L. (1986). Theory of ferromagnetic hysteresis. *J. Magn. Magn. Mater*, vol 61. pp. 48-60
- Jiles, D., Thoelke, J.B. and Devine, M.K. (1992). Numerical determination of hysteresis parameters for the modeling of magnetic properties using the theory of ferromagnetic hysteresis. *IEEE Transactions on Magnetics*, 28 (1): 27-35
- Karlec, V. (1998). *Aplicação de Algoritmos Genéticos na Determinação de Raízes de Equações Não Lineares*, Monografia de Final de Curso de Graduação em Informática, Departamento de Engenharia e Ciência da Computação, Universidade Regional Integrada do Alto Uruguai e das Missões. Frederico Westphalen.

- Krasnoselskii, M.A, Pokrovskii, A. and Niezgodka, M. (1989). *Systems with Hysteresis*, Springer-Verlag.
- Lederer, D., Igarashi, H., Kost, A. and Honma, T. (1999). On the parameter identification and application of the Jiles-Atherton hysteresis model for numerical modeling of measured characteristics. *IEEE Trans. Magn.*, vol 35, pp. 1211-1214, may 1999.
- Lemonge, A.C.C. (1999). *Aplicação de algoritmos genéticos em otimização estrutural*. Tese de Doutorado, COPPE, UFRJ, Rio de Janeiro.
- Levenberg, K. (1944). A method for the solution of certain problems in least squares, *Quart. Appl. Math.* vol. 2, pp 164-168,
- Liddiard, K. C (1986). Thin-Film Resistance Bolometer IR Detectors - II, *Infrared Physics*, 26 (1), pp. 43-49.
- Luenberger, D.G. (1984). *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA.
- Malyarov, V. G., Khrebtov, I.A., Kulikov, Y. V., Shaganov, I. I.; Zerov, V. Y. and Feoktistov, N.A (1999). Comparative investigations of the bolometric properties of thin film structures based on vanadium dioxide and amorphous hydrated silicon *Proc. SPIE, International Conference on Photoelectronics and Night Vision Devices*, Anatoly M. Filachev; Ed. Vol. 3819, pp. 136-142
- Marquardt, D.W. (1963). An Algorithm for Least-squares Estimation of Nonlinear Parameter. *J. Soc. Indust. Appl. Math.*, vol. 11, n° 2, pp. 431-441.
- Mayergoyz, I.D. and Friedman, G. (1988). Generalized Preisach Model of Hysteresis (Invited). *IEEE Trans. on Magnetics*, Vol. 24, pp. 212-217.
- Mayergoyz, I. (1991). *Mathematical Models of Hysteresis*, Springer-Verlag.
- Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs*. 3 ed. Springer-Verlag, Berlin.
- Mitchell, M. (1994). *An introduction to genetic algorithms*. MIT Press, New York.
- Noh, T.W. and Song, P.H. (1991). Self-consistency conditions for the Effective-Medium Approximation in composite materials. *Physical Review B* 44 (11): 5459-5464.

- Ohta, T., Enomoto, Y. and Kato, R. (1991). Domain growth with time-dependent front velocity in one dimension, *Physical Review B*, 43 (16): 262-268
- Otsuka, K. and Wayman, C. M. (1998). *Shape Memory Materials*, Cambridge University Press.
- Piacentini, J.J., Grandi, B. C. S., Hofmann, M.P., de Lima, F. R. R. and Zimmermann, E. (1998). *Introdução ao Laboratório de Física*. Florianópolis, Editora da UFSC, Florianópolis.
- Potapov, P., Gotthardt, R. and Bataillardil, L. (1998). The Transient Internal Friction Peak Associated with The Martensitic Transformation in Ni-Mn-Ti, *Physica Status Solidi, A - Applied Research*, 165 (2), pp. 401-408.
- Preisach, F. (1935). Über die magnetische nachwirkung. *Zeitschrift für Physik* 94: 277-302
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P.(2001). *Numerical Recipes in Fortran 77 - The Art of Scientific Computing*, Second Edition, Volume 1 of Fortran Numerical Recipes, Cambridge University Press.
- Progozy, S. (1993). PSPICE computer modeling of hysteresis effects. *IEEE Transactions on Education*, vol. 36, pp. 2-5.
- Ramos, A. R. (1994). *Aplicação de Algoritmos Genéticos em Problemas de Otimização*, CPGCC, UFRGS.Porto Alegre.RS.
- Salvini, A. and Fulginei, F.R. (2002). Genetic algorithms and neural networks generalizing the Jiles-Atherton model of static hysteresis for dynamic loops. *IEEE Trans. Magn.*, vol. 38, no.2, pp. 873-876.
- Sena, J.A.S., Almeida, L.A.L., Deep, G.S., Freire, R.C.S. and Lima, A.M.N. (2000). *Modelagem da histerese magnética utilizando algoritmo genético*. XIII Congresso Brasileiro de Eletromagnetismo, Natal.
- Serrada, A. P. (1996). *Uma Introducción a la Computación Evolutiva*. ETS Ing. Industrial/Universidad de Valladolid, Valladolid.
- Schmidt, N. and Guldner, H. (1996). Simple method to determine dynamic hysteresis loops of soft magnetic materials, *IEEE Transactions on Magnetics*, vol 32, pp. 489-496.

- Srinivas, M. and Patnaik, L. (1994). Genetic algorithms: a survey. *IEEE Computers*, v.27, n.6, p. 17-26.
- Stauffer, D. and Aharony, A. (1994). *Introduction to Percolation Theory*. Taylor and Francis.
- Stroud, D (1998). The Effective Medium Approximations: Some Recent Developments, *Superlattices and Microstructures*, 23 (3-4), pp. 567-573.
- Tipler, P.A (2000). *Física para Cientistas e Engenheiros, Quarta Edição, Vol.2 - Eletricidade e Magnetismo, Ótica*, LTC- Livros Técnicos e Científicos, Rio de Janeiro, RJ.
- Torres, D. (1991). Existence of magnetization-dependent Preisach models, *IEEE Transactions on Magnetics* 27(4): 3697-3699.
- Torres, E.D. (1994). A Preisach model for accommodation, *IEEE Transactions on Magnetics* 30(5): 2701-2707.
- Visintin, A. (1994). *Differential Models of Hysteresis*, Springer-Verlag.
- Wilson, P. R. Ross, J.N. and Brown, A.D. (2001). Optimizing the Jiles-Atherton model of hysteresis by a genetic algorithm. *IEEE Transactions on Magnetics*, vol. 37, no.2, pp. 989-993.
- Wilson, P. R. Ross, J.N. and Brown, A.D. (2002). Magnetic material model characterization and optimization software. *IEEE Transactions on Magnetics*, vol. 38, no.2, pp. 1049-1052.