



Pós-Graduação em Ciência da Computação

“Geração de Diversidade na Otimização Dinâmica
Multiobjetivo Evolucionária por Paisagens de
Não-Dominância”

Por

Carlos Renato Belo Azevedo

Dissertação de Mestrado



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE, FEVEREIRO/2011



Universidade Federal de Pernambuco
Centro de Informática
Pós-graduação em Ciência da Computação

Carlos Renato Belo Azevedo

**“Geração de Diversidade na Otimização Dinâmica
Multiobjetivo Evolucionária por Paisagens de
Não-Dominância”**

*Trabalho apresentado ao Programa de Pós-graduação em
Ciência da Computação do Centro de Informática da Univer-
sidade Federal de Pernambuco como requisito parcial para
obtenção do grau de Mestre em Ciência da Computação.*

Orientador: *Aluizio Fausto Ribeiro Araújo*

RECIFE, FEVEREIRO/2011

Catálogo na fonte
Bibliotecária Jane Souto Maior, CRB4-571

Azevedo, Carlos Renato Belo
Geração de diversidade na otimização dinâmica
multiobjetivo evolucionária por paisagens de não-dominância /
Carlos Renato Belo Azevedo - Recife: O Autor, 2011.
xxv, 182 p. : fig., tab.

Orientador: Aluízio Fausto Ribeiro Araújo.
Dissertação (mestrado) Universidade Federal de
Pernambuco. Cln. Ciência da computação, 2011.

Inclui bibliografia e apêndice.

1. Inteligência artificial. 2. Inteligência computacional. 3.
Computação evolucionária. I. Araújo, Aluízio Fausto Ribeiro
(orientador). II. Título.

006.3

CDD (22. ed.)

MEI2011 – 028

À minha amada mãe, Dione Belo, com muito carinho.

Agradecimentos

O climax deste trabalho é, certamente, o momento de destacar as pessoas essenciais sem as quais esta dissertação não teria sido escrita. Falo de minha adorável esposa, Érika, que, nos últimos sete anos, traspassou com frequência as barreiras do possível para que eu superasse todas as adversidades. Sem o seu incentivo, eu não teria tido a coragem de prestar o vestibular para Computação quando eu ainda era um jovem estudante do Centro de Artes e Comunicação da UFPE. Como se não bastasse, ela ainda ajudou na formatação e revisão do texto. À ela, toda a minha imensa gratidão e amor.

Os meus colegas de sala de aula e de laboratório contribuíram muito para tornar os 23 meses em que fui aluno do CIn mais prazerosos e instigantes, em especial, Roberto Pinheiro, com quem desde a graduação tenho tido o prazer de contar nas equipes das disciplinas, Danilo Carvalho, em cuja competência me inspirei ao longo do curso, e Rafael Xavier, com quem criei ao longo dos sete últimos anos profundas raízes de uma duradoura amizade.

Agradeço ao meu orientador, Prof^o Aluizio, que sempre demonstrou receptividade às minhas ideias, fornecendo todas as condições para que a minha criatividade pudesse ser explorada eficientemente e direcionada à obtenção dos resultados que o ofício da pesquisa acadêmica exige. Foi ele quem me introduziu ao estudo da geração de diversidade em algoritmos evolucionários, despertando o meu interesse pelo tema. Também recorro das agradáveis conversas sobre política, cultura e futebol (essas nem sempre tão agradáveis!) que muito me aliviaram o estresse do dia a dia de trabalho.

Gostaria de expressar a minha gratidão aos professores da banca avaliadora, Carmelo Bastos (UPE) e Paulo Adeodato (UFPE), pelo tempo dedicado à leitura e apreciação do meu trabalho e pelas diversas sugestões e críticas que contribuíram para melhorar substancialmente o conteúdo apresentado nesta dissertação.

Agradeço também ao Prof^o Scott Gordon, da California State Univ. Sacramento, que, além de ter contribuído para a minha formação enquanto pesquisador, sempre foi um grande incentivador e amigo. Também gostaria de agradecer aos meus colegas da Singularity University pelo apoio e incentivo, em especial, as pessoas que compuseram o grupo de trabalho *AI Labs*. Não os citarei nominalmente para não cometer a indelicadeza de esquecer alguém.

Não poderia deixar de mencionar os professores que tanto me incentivaram e serviram de referência para o modelo de professor-pesquisador o qual almejo ser: do Prof^o Tiago Ferreira (UFRPE), recordo-me bem de nossas conversas sobre ciência, do seu entusiasmo para novas ideias e da enorme fé que deposita em seus alunos, valorizando-os e lembrando-

lhes o tempo todo de suas capacidades; do Prof^o Francisco Madeiro (UPE/UNICAP), o *Fellow*, admiro o esmero com o qual se dedica à sala de aula. Com Madeiro, aprendi a valorizar ainda mais o trabalho em equipe e a pesquisa multidisciplinar. Foi ele também a quem tanto enfrentei (e venci!) nas gloriosas quadras de tênis do meu querido Sport Clube do Recife.

Por fim, destaco o trabalho de todos os funcionários do Centro de Informática, os quais fazem desse um centro de excelência em pesquisa no Brasil. O meu especial agradecimento à toda a sociedade pernambucana e brasileira, que, com seu labor convertidos em investimentos na educação, permitiu que eu e muitos colegas pudessem se qualificar e sonhar com um Brasil mais igualitário e justo.

Este trabalho contou com o suporte do Conselho Nacional de Pesquisa e Desenvolvimento (CNPq), por meio da concessão de bolsa de mestrado, processo número 132698/2009-9.

*As far as the laws of mathematics refer to reality, they are not certain,
as far as they are certain, they don't refer to reality.*

—ALBERT EINSTEIN (1879–1955)

Resumo

A geração e manutenção de soluções distintas em Algoritmos Evolucionários Multiobjetivo (MOEAs), sobretudo em ambientes dinâmicos nos quais os critérios de avaliação das soluções podem variar ao longo do tempo, é um problema em aberto, em que há poucos estudos sobre a influência das diferentes maneiras de gerar diversidade na qualidade do conjunto de soluções ótimas. A inclusão de geradores de diversidade em MOEAs pode aumentar o custo do processo evolucionário e prejudicar seu desempenho. Surge, então, a necessidade de buscar meios para atenuar a repercussão negativa do aumento dos níveis de dispersão da população de soluções candidatas no caminhamento à superfície onde se situam os pontos ótimos, conhecida como a Frente de Pareto (*PF*).

Em sistemas biológicos, regimes de imigração aumentam as possíveis combinações de trocas genéticas, promovendo diversidade de caminhos evolucionários. Inspirada na modelagem de imigração natural, esta pesquisa investiga a inserção de soluções atípicas (*imigrantes*) em populações de soluções candidatas como forma de gerar diversidade em MOEAs aplicados à otimização dinâmica multiobjetivo. A dissertação também propõe e formaliza as Paisagens de Não-Dominância (NDL) para guiar a inserção dos imigrantes gerados na população. As NDLs proveem os MOEAs das probabilidades dos imigrantes serem não-dominados em uma população a partir da estimação de funções densidade de probabilidade e de estatísticas de ordem multivariadas no espaço de objetivos.

Após caracterizar a influência da diversidade na dinâmica de aproximação da *PF* em MOEAs, incorporaram-se as NDLs a geradores de imigrantes. A validação experimental do Gerador de Diversidade baseado em NDL (NDL-DG) expressa o potencial da abordagem proposta no aumento da qualidade média dos conjuntos de soluções não-dominadas evoluídas. A análise dos resultados da incorporação do NDL-DG ao algoritmo NSGA2 revela a obtenção de soluções de maior qualidade média com significância estatística em 79% dos cenários de otimização dinâmica estudados, em termos do indicador de Hipervolume *offline*, quando comparado com populações evoluídas sem o uso das NDLs. Em seguida, identificaram-se os cenários de otimização em que o NDL-DG se mostra mais promissor. Finalmente, indicaram-se direções de pesquisa para ampliar o alcance da aplicação das NDLs para outros problemas em aberto na otimização multiobjetivo evolucionária.

Palavras-chave: otimização dinâmica multiobjetivo, algoritmos evolucionários, geração de diversidade, modelos de imigração, estatísticas de ordem multivariada, estimação de funções densidade de probabilidade, paisagens de não-dominância.

Abstract

The generation and maintenance of distinct solutions in Multiobjective Evolutionary Algorithms (MOEAs), especially in dynamic environments in which the criteria for evaluating solutions may vary over time, is an open problem, in which there are few studies on the influence of the different ways to generate diversity in the quality of the optimal solutions set. The inclusion of diversity generators in MOEAs can increase the cost of the evolutionary process and impair their performance. Hence, it comes the need to seek ways of mitigating the negative impact of the rising levels of dispersion within the candidate solutions population in the road ahead to the surface where lie the optimal points, known as the Pareto Front (*PF*).

In biological systems, immigration schemes increase the possible combinations of genetic exchanges, promoting diversity of evolutionary paths. Inspired by the natural models of immigration, this research investigates the inclusion of atypical solutions (immigrants) in populations of candidate solutions as a way to generate diversity in MOEAs applied to dynamic multiobjective optimization. This dissertation also proposes and formalizes the Non-Dominance Landscapes (NDL) to guide the insertion of the generated immigrants in the population. The NDLs provide MOEAs with the probabilities of the immigrants being non-dominated in the population, from the estimation of probability density functions and of multivariate order statistics in the objective space.

After characterizing the influence of diversity in the approximation dynamics of the *PF* in MOEAs, the NDLs have been incorporated into the immigrants generators. The experimental validation of the NDL-based Diversity Generator (NDL-DG) expresses the potential of the proposed approach in increasing the average quality of the evolved non-dominated solutions sets. The results analysis of the incorporation of the NDL-DG into the NSGA2 algorithm show that higher average quality solutions are obtained with statistical significance at 79% of the studied dynamic optimization scenarios, in terms of the offline Hypervolume indicator, when compared with populations evolved without the use of NDLs. We then identified the optimization scenarios in which the NDL-DG appears more promising. Finally, we indicated research directions to extend the range of application of NDLs to other open problems in evolutionary multiobjective optimization.

Keywords: dynamic multiobjective optimization, evolutionary algorithms, diversity generation, immigration schemes, multivariate order statistics, estimation of probability density functions, non-dominance landscapes.

Sumário

Lista de Figuras	xix
Lista de Tabelas	xxiii
Lista de Acrônimos	xxv
1 Introdução	1
1.1 Motivação e Justificativa	2
1.2 Objetivos da Investigação	5
1.2.1 Objetivo Geral	5
1.2.2 Objetivos Específicos	5
1.3 Estrutura da Dissertação	5
2 Fundamentos de Probabilidade	7
2.1 Variáveis Aleatórias	7
2.1.1 Variável Aleatória Uniforme	10
2.1.2 A Densidade Gaussiana Multivariada	11
2.2 Estatísticas de Ordem	11
2.2.1 A Função Densidade de uma Ordem Total na Reta Real	12
2.2.2 Relações de Ordem Multivariadas	12
2.2.3 A Função Densidade de uma Ordem de Pareto em \mathcal{R}^m	13
2.3 Estimação de Funções Densidade de Probabilidade	15
2.3.1 Estimação Paramétrica	16
2.3.2 Estimação Não-Paramétrica	18
2.3.3 Estimação Semiparamétrica	22
2.3.4 Estimação Multivariada Baseada em <i>Kernel</i>	22
2.4 Discussão Final e Síntese	23
3 Fundamentos de Otimização Evolucionária	25
3.1 Otimização Contínua	25
3.2 Otimização Multiobjetivo	28
3.2.1 Indicadores de Proximidade, Espaçamento e Cobertura	30
3.2.2 Teoria do Hipervolume	31
3.3 Otimização Dinâmica	33
3.4 Fundamentos de Algoritmos Evolucionários	34

3.4.1	Operadores de Recombinação	36
3.4.2	Operadores de Mutação	39
3.4.3	Operadores de Seleção	39
3.5	Algoritmos Evolucionários Multiobjetivo (MOEAs)	40
3.5.1	AG com Ordenação por Não-Dominação (NSGA2)	42
3.5.2	MOEAs em Ambientes Dinâmicos	43
3.6	Discussão Final e Síntese	45
4	A Influência da Diversidade Genética	47
4.1	O Balanço Entre Exploração e Exploração	48
4.1.1	Exploração e Exploração em MOEAs	49
4.1.2	Métricas de Diversidade Genética	50
4.2	A Diversidade Genética de Duas Subpopulações	51
4.2.1	O Modelo e Propriedades Estatísticas	51
4.2.2	Exemplo Ilustrativo e Validação	53
4.3	Correlação entre Diversidade e Hipervolume	56
4.3.1	Metodologia e Planejamento dos Experimentos	57
4.3.2	Resultados e Discussão	57
4.4	Geração de Diversidade por Inserção de Imigrantes	64
4.4.1	Imigrantes Não-Correlacionados	66
4.4.2	Imigrantes Correlacionados	68
4.4.3	Estratégias Híbridas	70
4.4.4	Imigrantes na Otimização Multiobjetivo	70
4.5	Discussão Final e Síntese	71
5	Diversidade Dirigida na Otimização Multiobjetivo	73
5.1	Probabilidades de Dominância	74
5.2	Paisagens de Não-Dominância (NDLs)	77
5.2.1	Teorema da Monotonicidade de Elementos Maximais	79
5.3	Geração de Diversidade Dirigida por NDLs	80
5.3.1	Estimação da <i>pdf</i> da População no Espaço de Objetivos	84
5.3.2	Geração dos Imigrantes em MOEAs	85
5.3.3	Avaliação dos Imigrantes	87
5.3.4	Seleção por Probabilidade de Não-Dominância (PND)	87
5.3.5	Estratégia de Inserção dos Imigrantes na População	88
5.4	Validação do Modelo Proposto	89

5.5	Discussão Final e Síntese	93
6	Resultados Experimentais	95
6.1	Metodologia e Arranjo Experimental	95
6.1.1	Incorporação do NDL-DG ao NSGA2	96
6.1.2	Problemas de Otimização Multiobjetivo Seleccionados	97
6.1.3	Métricas de Avaliação	100
6.2	Inserção de Imigrantes na Otimização Multiobjetivo	101
6.2.1	Efeitos do Incremento de α em Ambientes Estáticos	102
6.2.2	Efeitos do Incremento de α em Ambientes Dinâmicos	105
6.3	Seleção de Imigrantes por PNDs	107
6.3.1	Repercussão do Incremento de K_{NDL} no Hipervolume <i>Offline</i>	107
6.3.2	Correlação Entre Diversidade Genética e Hipervolume	125
6.3.3	Correlação Entre Diversidade de Frentes e Hipervolume	126
6.4	Discussão Final e Síntese	129
7	Conclusão e Perspectivas de Investigação	131
7.1	Principais Contribuições	132
7.2	Limitações do Trabalho	133
7.3	Sugestões para a Continuidade do Trabalho	134
7.3.1	Controle de Parâmetros do NDL-DG	134
7.3.2	Modelos Alternativos de Geração de Imigrantes por NDLs	135
7.3.3	Outras Aplicações para NDLs	136
7.4	Publicações Decorrentes da Pesquisa	137
	Apêndices	139
A	Métodos Clássicos de Otimização Multiobjetivo	141
A.1	Articulação <i>a priori</i> de Preferências	142
A.2	Articulação <i>a posteriori</i> de Preferências	143
A.3	Articulação Progressiva de Preferências	145
B	Problemas de Otimização Multiobjetivo	149
B.1	Roteamento <i>Multi-Cast</i>	149
B.2	Agrupamento de Dados	150
B.3	Caixeiro Viajante Multiobjetivo	151
B.4	Problemas ZDT para MOPs Contínuos	152

B.5	Problemas FDA para MOPs Dinâmicos Contínuos	155
C	Outros Algoritmos Evolucionários Multiobjetivo	159
C.1	Algoritmo Genético com Avaliação de Vetores	159
C.2	Algoritmo Genético multiobjetivo	161
C.3	Algoritmos Genéticos com Nicho de Pareto	163
C.4	Algoritmos Evolucionários por Força de Pareto	165
D	Biografia e Publicações	171
D.1	Biografia Resumida	171
D.2	Publicações Seleccionadas	172
	Bibliografia	173

Lista de Figuras

2.1	Estimação da densidade por Kernel Gaussiano.	21
3.1	Representação do espaço de soluções factíveis para o problema da equação (3.3).	27
3.2	Representação do espaço de objetivos. Têm-se as seguintes relações de dominância: $A \prec C, B \prec C$; mas A e B não são comparáveis, i.e., $A \not\prec B$ e $B \not\prec A$	29
3.3	O indicador \mathcal{I}_H corresponde à área delimitada em cinza no gráfico. . .	32
3.4	Módulos comuns em algoritmos evolucionários baseados em população (Adaptado de Zitzler <i>et al.</i> (2004)).	34
3.5	Exemplo de uma operação de cruzamento de 1 ponto em duas cadeias binárias.	36
3.6	Função densidade de probabilidade da razão de espalhamento β no operador SBX.	38
3.7	Representação do maior cubóide que contém uma dada solução no espaço F_m , conforme calculado no NSGA2 (adaptado de Deb <i>et al.</i> (2000)). . .	43
4.1	Duas aproximações da PF com diferentes níveis de diversidade e proximidade (Adaptado de Bosman e Thierens (2003)).	49
4.2	Ilustração de dois modelos de populações com taxas de imigração complementares.	54
4.3	Diversidade genética prevista com variâncias das subpopulações idênticas.	55
4.4	Variação de $I_X(\alpha)$ em função da diferença entre as variâncias das subpopulações.	55
4.5	Correlação entre diversidade acumulada e Hipervolume na última geração.	60
4.6	Evolução do coeficiente de Pearson ($J_{\text{tam}} = 10$) para o conjunto ZDT.	61
4.7	Correlações temporais e autocorrelação para diferentes tamanhos de janela no ZDT1.	63
4.8	Autocorrelação amostral para $J_{\text{tam}} = 10$ nos demais MOPs ZDT.	64
4.9	Representação de um modelo de imigração genérico com taxas assimétricas (Adaptado de Altrock <i>et al.</i> (2010)).	66
4.10	Geração de imigrantes correlacionados com base em memória (Adaptado de Yang (2008)).	68

5.1	Paisagens de Não-Dominância em diferentes gerações (g) do problema ZDT1.	78
5.2	Diagrama de fluxo do mecanismo de geração de diversidade dirigida NDL-DG.	81
5.3	Visualização da pdf da população no espaço de objetivos estimada por Kernel Gaussiano multivariado nas primeiras gerações do NSGA2 no problema ZDT1.	84
5.4	<i>Box plots</i> para análise qualitativa dos imigrantes gerados pelo NDL-DG _R	92
6.1	Resultados para os experimentos com o NSGA(R) no problema ZDT2.	103
6.2	Resultados obtidos com o NSGA(R) no ZDT2: (a) entropia normalizada; (b) diversidade medida após a inserção de imigrantes na população e antes da seleção de sobreviventes.	105
6.3	Resultados obtidos com o NSGA2(R) (\circ) e NSGA2(E) (\bullet) no FDA1 em função de α	106
6.4	Variação no Hipervolume (\mathcal{J}_H) com o incremento de K_{NDL} em MOPs dinâmicos: $n_g = 5, t_g = 5$ (\circ); $n_g = 5, t_g = 10$ (\triangle); $n_g = 10, t_g = 5$ (\bullet) e $n_g = 10, t_g = 10$ (\square).	109
6.5	Evolução temporal do indicador de Hipervolume com e sem o NDL-DG em ambientes dinâmicos ($n_g = 5, t_g = 5$).	114
6.6	Evolução temporal do indicador de Hipervolume com e sem o NDL-DG em ambientes dinâmicos ($n_g = 5, t_g = 10$).	115
6.7	Evolução temporal do indicador de Hipervolume com e sem o NDL-DG em ambientes dinâmicos ($n_g = 10, t_g = 5$).	116
6.8	Evolução temporal do indicador de Hipervolume com e sem o NDL-DG em ambientes dinâmicos ($n_g = 10, t_g = 10$).	117
6.9	<i>Box-plots</i> para a diversidade genética (I_M) com e sem o NDL-DG em ambientes dinâmicos ($n_g = 5, t_g = 5$).	119
6.10	<i>Box-plots</i> para a diversidade genética (I_M) com e sem o NDL-DG em ambientes dinâmicos ($n_g = 5, t_g = 10$).	120
6.11	<i>Box-plots</i> para a diversidade genética (I_M) com e sem o NDL-DG em ambientes dinâmicos ($n_g = 10, t_g = 5$).	121
6.12	<i>Box-plots</i> para a diversidade genética (I_M) com e sem o NDL-DG em ambientes dinâmicos ($n_g = 10, t_g = 10$).	122
6.13	Evolução da Entropia Normalizada (H_N) e da Diversidade Genética (I_M) com e sem o NDL-DG no FDA2 ($n_g = 5, t_g = 5$).	124

6.14	Variação da correlação de Pearson entre I_M e \mathcal{J}_H com o incremento de K_{NDL} em MOPs dinâmicos: $n_g = 5, t_g = 5$ (\circ); $n_g = 5, t_g = 10$ (\triangle); $n_g = 10, t_g = 5$ (\bullet) e $n_g = 10, t_g = 10$ (\square).	127
6.15	Variação da correlação de Pearson entre H_N e \mathcal{J}_H com o incremento de K_{NDL} em MOPs dinâmicos: $n_g = 5, t_g = 5$ (\circ); $n_g = 5, t_g = 10$ (\triangle); $n_g = 10, t_g = 5$ (\bullet) e $n_g = 10, t_g = 10$ (\square).	128
A.1	Operação do método NBI no espaço de objetivos. Os pontos amostrados no envoltório convexo são otimizados na direção do ponto de utopia (ver Definição no capítulo 4). No diagrama, o ECOI é representado pelo seguimento de reta $\bar{P}Q$ (ver Definição A.1).	144
B.1	Representação de uma instância aleatória do TSP.	151
B.2	Frentes de Pareto no conjunto de problemas de otimização ZDT.	153
B.3	Frentes de Pareto variantes no tempo ($PF(t)$) no conjunto FDA.	156
C.1	Exemplo da atribuição de <i>ranks</i> empregada no MOGA (adaptado de Konak <i>et al.</i> (2006)).	161
C.2	Representação do processo de compartilhamento de classes de equivalência no NPGA (adaptado de Horn <i>et al.</i> (1994)).	164
C.3	Exemplo da atribuição de aptidão realizada na população e no conjunto externo de soluções não-dominadas do SPEA (adaptado de Zitzler e Thiele (2002)).	167

Lista de Tabelas

4.1	Resultados da validação por simulação estocástica da diversidade genética em função de α	53
4.2	<i>Setup</i> experimental para o NSGA2(η_{SBX} , η_{mut}).	57
4.3	Diversidade genética, Hipervolume e Correlação para os problemas ZDT.	58
5.1	Quadro comparativo para as diferentes versões possíveis do NDL-DG.	83
5.2	<i>p-values</i> computados para as divergências de Kullback-Leibler em função de K_{NDL}	90
6.1	Quadro comparativo dos algoritmos considerados para os experimentos.	98
6.2	Quadro comparativo dos indicadores considerados para discussão dos resultados.	102
6.3	<i>p-values</i> computados para os testes de significância de $\overline{\mathcal{F}_h}$ entre a melhor configuração com NDLs (K_{NDL} entre parênteses) e sem o uso de NDLs.	110
6.4	Percentual do número de gerações em que as NDLs apresentam ganhos (+) com significância estatística para $K_{\text{NDL}} > 1$ em relação à ausência de NDLs ($n_g = 5, t_g = 5$)	111
6.5	Percentual do número de gerações em que as NDLs apresentam ganhos (+) com significância estatística para $K_{\text{NDL}} > 1$ em relação à ausência de NDLs ($n_g = 5, t_g = 10$).	111
6.6	Percentual do número de gerações em que as NDLs apresentam ganhos (+) com significância estatística para $K_{\text{NDL}} > 1$ em relação à ausência de NDLs ($n_g = 10, t_g = 5$).	111
6.7	Percentual do número de gerações em que as NDLs apresentam ganhos (+) com significância estatística para $K_{\text{NDL}} > 1$ em relação à ausência de NDLs ($n_g = 10, t_g = 10$).	112

Lista de Acrônimos

AE	Algoritmo Evolucionário
AG	Algoritmo Genético
CD	<i>Crowding Distance</i>
<i>cdf</i>	Função de Distribuição Acumulada
EIGA	AG com Imigrantes Baseados na Elite
EMV	Estimador de Máxima Verossimilhança
EOM	Estatística de Ordem Multivariada
GP	Programação Genética
MOEA	Algoritmo Evolucionário Multiobjetivo
MOGA	Algoritmo Genético Multiobjetivo
MOP	Problema de Otimização Multiobjetivo
NDL	Paisagem de Não-Dominância
NDL-DG	Gerador de Diversidade por Paisagem de Não-Dominância
NPGA	Algoritmo Genético por Nicho de Pareto
NSGA	AG com Ordenação por Não-Dominação
<i>pdf</i>	Função Densidade de Probabilidade
<i>PF</i>	Frente de Pareto
PSO	Otimização por Enxame de Partículas
RIGA	AG com Imigrantes Aleatórios
SBX	Cruzamento Binário Simulado
SPEA	AE por Força de Pareto
v.a.	Variável Aleatória

1

Introdução

A honestidade é o primeiro capítulo no livro da sabedoria.

– Thomas Jefferson

Graças ao avanço tecnológico e à ampla disponibilidade de recursos computacionais, diversos problemas de otimização estudados nas décadas anteriores têm sido atualmente reformulados em termos de múltiplos objetivos, muitas vezes conflitantes (Watanabe e Sakakibara, 2007; Handl *et al.*, 2008; Hohm e Zitzler, 2009). A consideração de mais de uma função-objetivo a ser minimizada ou maximizada, além de condizente com a realidade da maioria dos problemas práticos, requer novas definições de otimalidade e novos métodos de busca por soluções.

Algoritmos Evolucionários (AEs) se apresentam como uma alternativa eficiente aos métodos clássicos de otimização multiobjetivo (Shukla e Deb, 2007), os quais operam por meio de otimizações sucessivas de um único objetivo (ver Apêndice A). Por não serem baseados em informações de gradiente, os AEs são aplicáveis em diversos contextos, pois independem da natureza das funções-objetivo e das restrições do problema (Marler e Arora, 2004). Ainda, por serem provadamente otimizadores globais baseados em populações de soluções candidatas (Goldberg, 1989), os AEs têm a capacidade de considerar simultaneamente as informações amostradas de diversos pontos do espaço de soluções, realizando uma busca paralela randomizada por potenciais soluções em diferentes regiões daquele espaço.

Na otimização multiobjetivo por busca evolucionária, a existência de níveis de diversidade adequados em uma população de soluções candidatas não apenas contribui para uma melhor cobertura do espaço de busca (Bosman e Thierens, 2003), mas também proporciona a exploração de diferentes caminhos evolucionários que levam à superfície de compensação (*tradeoff*), na qual não é possível melhorar simultaneamente todos os

valores das funções-objetivo sem acarretar na piora de ao menos um valor dessas funções. Nesse contexto, a superfície de compensação, conhecida como Frente de Pareto (*Pareto Front*, *PF*, do inglês), deve ser aproximada – com o máximo espaçamento de pontos de aproximação possível – por um conjunto finito de soluções. Para tanto, Algoritmos Evolucionários Multiobjetivos (MOEAs) têm se mostrado uma abordagem promissora (Zitzler *et al.*, 2004).

O processo de *geração de diversidade* em MOEAs consiste em produzir soluções candidatas distintas a serem inseridas na população, quer por meio de operadores de variação convencionais, ou por métodos alternativos de amostragem estatística do espaço de busca. A geração de soluções distintas aplicada a ambientes dinâmicos — problemas em que as funções-objetivo e/ou restrições das variáveis mudam com o tempo — mostra-se essencial para a adaptação rápida da população aos ambientes recém-modificados (Gouvêa Jr., 2009). Na otimização dinâmica multiobjetivo, soluções antes encontradas na superfície de compromisso podem não ser mais adequadas em estados temporais posteriores do problema (Jin e Branke, 2005). Nesse cenário, a importância das estratégias de geração e controle da diversidade torna-se ainda maior.

Esta dissertação relata a pesquisa realizada pelo autor sobre o projeto de geradores de diversidade em MOEAs, aplicados a ambientes estáticos e dinâmicos, com foco em uma nova abordagem baseada na caracterização probabilística das relações de dominância (ver definições 3.8 e 3.9, página 28) entre as soluções candidatas em um espaço de objetivos multidimensional. A nova abordagem é utilizada para gerar soluções atípicas (*imigrantes*) que, ao serem inseridas na população, contribuam para o aumento dos níveis de diversidade e que, ao mesmo tempo, possuam maiores chances de serem não-dominadas, i.e., de pertencerem ao subgrupo de soluções na população as quais são efetivamente utilizadas na aproximação da *PF*, contribuindo dessa forma para reverter os efeitos negativos dos níveis excessivos de diversidade gerados ao longo do processo evolucionário, levando a um conjunto de soluções não-dominadas diversificado e de melhor qualidade quando comparado a conjuntos obtidos da inserção de imigrantes sem a consideração das probabilidades de não-dominância.

1.1 Motivação e Justificativa

O sucesso dos recentes avanços no projeto de AEs mono-objetivo – os quais buscam otimizar uma única função – aplicados em ambientes dinâmicos têm motivado o desenvolvimento de mecanismos mais eficientes de geração e controle de diversidade na

população de soluções candidatas (Ursem, 2002; Yang, 2008; Gouvêa Jr. e Araújo, 2010). Embora os operadores de variação, bem como os mecanismos baseados em densidade de soluções e compartilhamento de aptidão (Deb, 2001), sejam tradicionalmente utilizados para promover a dispersão dos indivíduos no espaço de busca em MOEAs, a perda de diversidade ao longo das gerações não pode ser totalmente evitada por estratégias convencionais (ver resultados relatados no Capítulo 6, Seções 6.2.1 e 6.2.2). Uma abordagem alternativa para a geração de diversidade, cujo estudo vem se intensificando nos últimos anos, considera a inserção de soluções atípicas (imigrantes) na população como forma de aumentar as chances de adaptação dos indivíduos em face às mudanças temporais, no contexto dos AEs mono-objetivo (Yang, 2008). No caso da otimização dinâmica multiobjetivo, a aproximação da *PF* requer a manutenção de um nível de diversidade genética adequado ao longo de todo o processo evolucionário, o qual repercute positivamente no espalhamento e cobertura do conjunto de vetores não-dominados obtido no espaço de objetivos. A presente pesquisa, portanto, motiva-se pela possibilidade de estender as estratégias existentes de inserção de imigrantes em AEs mono-objetivo – como originalmente definida em Cobb e Grefenstette (1993) e aplicada, e.g., em Yang (2008) – para o contexto da otimização dinâmica multiobjetivo.

No Teorema *No Free Lunch* para Busca e Otimização, Wolpert e Macready (1997) afirmam que não existe um único algoritmo que seja preeminente em todos os tipos de problemas. Essa afirmação decorre principalmente do viés induzido pelo processo de amostragem do espaço de busca (Woodward, 2010). Tal viés é usualmente introduzido pelo projetista do algoritmo ao, por exemplo, definir os operadores de busca, e não pode ser facilmente eliminado, mesmo que o algoritmo projetado possua mecanismos de auto-adaptação (Woodward, 2010). Assim, se determinadas distribuições de probabilidade forem consideradas para a amostragem de novas soluções, a menos que a distribuição seja uniforme (e, portanto, todas as soluções possuam a mesma probabilidade de serem produzidas a cada instante de tempo), qualquer algoritmo privilegiará regiões particulares do espaço de soluções em detrimento de outras (i.e., introduzirá viés). Assim, supondo que, a cada iteração, um algoritmo arbitrário produza como saída um valor de aptidão que indica a qualidade da melhor solução armazenada na memória para a resolução do problema, então, quando todas as instâncias possíveis do problema de otimização são consideradas, não importa *como* a busca é executada: se as instâncias são igualmente prováveis, a probabilidade de se observar uma determinada ordem de valores de aptidão é estatisticamente independente do algoritmo (Wolpert e Macready, 1997) e, portanto, todo algoritmo de busca possui o mesmo desempenho esperado quando executado para todas

as instâncias possíveis. Entretanto, para uma instância fixa, diferentes parametrizações da estratégia de busca podem acarretar diferenças significativas, em termos da qualidade das soluções obtidas e do tempo necessário para encontrá-las (ver, por exemplo, Seção de resultados do Capítulo 4, página 57).

Na presente pesquisa, portanto, busca-se conceber estratégias de adaptação às mudanças temporais em AEs de forma a atenuar o viés introduzido pelas estratégias de busca estudadas, uma vez que, na otimização dinâmica, os fatores os quais levam a um bom desempenho da estratégia de busca em consideração podem mudar com o tempo, comprometendo o desempenho do algoritmo nos novos ambientes. Um exemplo de um desses fatores críticos é justamente o nível de diversidade genética medido em uma população de soluções candidatas, o qual se constitui em uma mensuração do grau de dispersão (ou espalhamento) das soluções no espaço de busca: se, por um lado, um AE em geral admite perder diversidade ao longo do processo de otimização para promover a convergência das soluções às regiões ótimas, por outro, em face às mudanças no ambiente, baixos níveis de diversidade tendem a retardar o tempo de adaptação das soluções no novo contexto temporal do problema.

Logo, a geração de soluções distintas em torno de regiões promissoras com a abordagem proposta pode se mostrar importante para promover diversidade em populações de soluções candidatas em MOEAs à medida que as soluções são conduzidas em direção à *PF* no processo de otimização dinâmico, antecipando eventuais mudanças temporais as quais exigem rápida adaptação da população de soluções aos novos objetivos e/ou restrições do problema. Dessa forma, a atenuação do viés da busca evolucionária pode contribuir para aumentar a robustez dos AEs estudados com respeito à imprevisibilidade das diferentes dinâmicas temporais dos problemas de otimização para os quais os AEs são projetados.

Com a solução proposta, baseada em geradores de imigrantes selecionados e inseridos na população de soluções candidatas com base em probabilidades de não-dominância, espera-se que diversos problemas do mundo real – dinâmicos e multiobjetivo – sejam resolvidos de forma mais eficiente, como o projeto de controladores para veículos autônomos (Pereira, 2001) ou a otimização do roteamento em redes de comunicação interdispositivos *ad hoc* (Bilgin e Khan, 2010).

1.2 Objetivos da Investigação

1.2.1 Objetivo Geral

A pesquisa relatada nesta dissertação objetiva compreender o papel da diversidade genética em populações de soluções candidatas em MOEAs. A partir desse entendimento, busca-se melhorar o desempenho de MOEAs que empreguem geração de diversidade por inserção de imigrantes na população, em concomitância ao caminharmento à *PF*.

1.2.2 Objetivos Específicos

Objetiva-se a investigação de um modelo algorítmico de geração de diversidade na otimização multiobjetivo evolucionária por meio da estimação de estatísticas de ordem multivariada, as quais representem probabilisticamente as relações de dominância entre as soluções candidatas. Especificamente, o modelo proposto é integrado às estratégias de inserção de soluções atípicas (imigrantes) estudadas em AEs mono-objetivo. A literatura da inserção de imigrantes em AEs é revisada no intuito de compreender a repercussão da diversidade gerada por tais técnicas no grau de adaptabilidade das populações de soluções ante mudanças nos ambientes.

Ademais, busca-se avaliar experimentalmente a dinâmica da diversidade em um MOEA do estado da arte, com o propósito de identificar a influência da diversidade genética na aproximação da *PF* em problemas de otimização com ambientes estáticos. Finalmente, objetiva-se determinar a contribuição e o mérito do mecanismo de geração de diversidade proposto para reverter a repercussão negativa dos geradores de diversidade na aptidão das populações em MOEAs, por meio da comparação estatística com resultados experimentais obtidos da aplicação de mecanismos de geração de diversidade convencionais, generalizados e estendidos nesse trabalho para o caso da otimização multiobjetivo.

1.3 Estrutura da Dissertação

O restante do texto organiza-se da seguinte forma: o Capítulo 2 apresenta as ferramentas estatísticas utilizadas ao longo da dissertação, especialmente as estatísticas de ordem multivariada e os métodos de estimação de funções densidade de probabilidade. O Capítulo 3 trata dos problemas de otimização matemática – contínua, dinâmica e multiobjetivo – os quais são utilizados nos estudos experimentais desse trabalho, além de discutir como os

AEs podem ser utilizados para resolvê-los. O Capítulo 4 trata do papel da diversidade no balanço entre exploração e exploração em AEs, discute e formula modelos simples de regimes de imigração e apresenta resultados iniciais acerca da influência da diversidade genética na variação dos indicadores de Hipervolume de populações evoluídas com o algoritmo NSGA2 (Deb *et al.*, 2000) sob diferentes parametrizações. O Capítulo 5 estende as estratégias de inserção de imigrantes em AEs mono-objetivo para o caso multiobjetivo, aplica as estatísticas de ordem multivariadas para a computação de probabilidades de dominância na otimização multiobjetivo, concebe e formaliza o conceito das NDLs, propõe o uso das NDLs em critérios de seleção na inserção de imigrantes em MOEAs e valida experimentalmente o gerador de diversidade baseado em NDLs proposto (NDL-DG). Finalmente, os Capítulos 6 e 7 apresentam, discutem e sintetizam resultados experimentais e conclusões da incorporação do NDL-DG no algoritmo NSGA2 para geração de diversidade dirigida em MOEAs em ambientes estáticos e dinâmicos.

A dissertação conta ainda com quatro apêndices: o Apêndice A apresenta alguns métodos clássicos (não-evolucionários) para a otimização multiobjetivo, dividindo-os quanto ao tratamento de preferências entre os objetivos. O Apêndice B apresenta classes de problemas de otimização multiobjetivo, tanto de natureza discreta quanto contínua, alguns dos quais são utilizados nos experimentos relatados ao longo da dissertação. O Apêndice C apresenta outros MOEAs de importância histórica e/ou os quais constituem o estado da arte na otimização multiobjetivo evolucionária. Finalmente o Apêndice D lista as principais publicações e biografia resumida do autor da dissertação.

2

Fundamentos de Probabilidade

A Natureza usa o mínimo possível de qualquer coisa.

– Johannes Kepler

Este Capítulo apresenta os modelos e ferramentas da teoria de probabilidades utilizadas direta ou indiretamente nesse trabalho, com foco em espaços amostrais contínuos.

2.1 Variáveis Aleatórias

Uma variável aleatória (v.a.) contínua é completamente caracterizada pela função densidade de probabilidade (*pdf* – *probability density function*, do inglês) a ela associada. Formalmente, uma v.a. X é contínua se existir uma função $p(x)$ não-negativa definida nos reais com a propriedade de que

$$P\{X \in A\} = \int_A p(x)dx, \quad (2.1)$$

em que $p(\cdot)$ é a *pdf* associada à X , $A \subset \mathcal{R}$ e

$$P(a) = P\{X \in (-\infty, a)\} = P\{X < a\} = \int_{-\infty}^a p(x)dx \quad (2.2)$$

é conhecida como função de distribuição acumulada de X (*cdf*, do inglês *cummulative distribution function*) (Ross, 2001). A *pdf* é uma função que representa a probabilidade da ocorrência de um evento quando X encontra-se em um intervalo ao redor de um valor de interesse, i.e., $p(x_0) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} P\{x_0 \leq x < x_0 + \varepsilon\}$. Ademais, deve-se notar que

$$P\{X = a\} = \int_a^a p(x)dx = 0, \quad (2.3)$$

isto é, existe uma probabilidade nula de uma v.a. contínua X assumir um valor específico e, uma vez que X deve assumir algum valor na reta real, $p(x)$ deve satisfazer

$$P\{X \in (-\infty, +\infty)\} = \int_{-\infty}^{+\infty} p(x)dx = 1. \quad (2.4)$$

Portanto, a *pdf* $p(x)$ não fornece a probabilidade $P\{X = x\}$, mas pode ser interpretada como uma medida do quão provável é X estar em um intervalo suficientemente grande em torno de x .

A *cdf* $P(x)$ de uma v.a. contínua relaciona-se com a *pdf* $p(x)$ de tal forma que

$$\frac{d}{dx}P(x) = p(x). \quad (2.5)$$

Note-se que: (i) $P(x)$ é não-decrescente; (ii) $\lim_{x \rightarrow \infty} P(x) = 1$; e (iii) $\lim_{x \rightarrow -\infty} P(x) = 0$. Ademais,

$$\begin{aligned} P\{a \leq X \leq b\} &= \int_a^b p(x)dx \\ &= \int_{-\infty}^b p(x)dx - \int_{-\infty}^a p(x)dx = P(b) - P(a). \end{aligned} \quad (2.6)$$

O valor esperado (ou esperança) de uma v.a. contínua X , expresso por

$$E[X] = \int_{-\infty}^{+\infty} xp(x)dx, \quad (2.7)$$

é uma média ponderada do domínio de X , em que os pesos são os valores mapeados pela *pdf* associada. Se $g(X)$ é uma função da v.a. X , tem-se

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)p(x)dx. \quad (2.8)$$

Por exemplo, se a, b são constantes reais e $g(X) = aX + b$,

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{+\infty} (ax + b)p(x)dx \\ &= a \int_{-\infty}^{+\infty} xp(x)dx + b \int_{-\infty}^{+\infty} p(x)dx \\ &= aE[X] + b. \end{aligned} \quad (2.9)$$

A esperança de uma v.a. X é apenas uma das medidas que caracterizam a função de

distribuição de X e a *pdf* associada. Para descrever o grau de dispersão dos valores assumidos por X , utiliza-se o conceito de variância (Ross, 2001):

$$\text{Var}(X) = E[(X - \mu)^2], \quad (2.10)$$

em que $\mu = E[X]$. Vê-se, então, que $\text{Var}(X)$ é o valor esperado das diferenças quadráticas entre os valores assumidos por X e a sua média, dada por μ . Uma propriedade da variância é dada a seguir:

$$\begin{aligned} \text{Var}(aX + b) &= E[(aX + b - a\mu - b)^2] = E[(a^2(X - \mu)^2)] \\ &= a^2 E[(X - \mu)^2] = a^2 \text{Var}(X). \end{aligned} \quad (2.11)$$

Um sistema probabilístico composto por duas v.a.'s, digamos X e Y , é caracterizado unicamente pela distribuição conjunta

$$P_{XY}(a, b) = P\{X \leq a, Y \leq b\}. \quad (2.12)$$

As *cdf*'s de X e Y são obtidas a partir de P_{XY} da seguinte forma:

$$\begin{aligned} P_X(a) &= P\{X \leq a, Y < \infty\} = P(a, \infty) \\ P_Y(b) &= P\{X < \infty, Y \leq b\} = P(\infty, b). \end{aligned} \quad (2.13)$$

Dados dois conjuntos reais, A e B , a *pdf* conjunta de X e Y é uma função $p(x, y)$ tal que

$$P\{X \in A, Y \in B\} = \int_B \int_A p(x, y) dx dy. \quad (2.14)$$

Nesse caso, as *cdf*'s $P_X(x)$ e $P_Y(y)$ podem ser obtidas a partir de $p(x, y)$:

$$\begin{aligned} P_X(x) &= P\{X \in A\} = \int_A \int_{-\infty}^{\infty} p(x, y) dy dx \\ P_Y(y) &= P\{Y \in B\} = \int_B \int_{-\infty}^{\infty} p(x, y) dx dy, \end{aligned} \quad (2.15)$$

em que

$$\begin{aligned} p_X(x) &= \int_{-\infty}^{\infty} p(x, y) dy \\ p_Y(y) &= \int_{-\infty}^{\infty} p(x, y) dx. \end{aligned} \quad (2.16)$$

Os conceitos de *cdf* conjunta e *pdf* conjunta são naturalmente generalizados para vetores aleatórios em \mathcal{R}^m .

Se duas v.a.'s X e Y são estatisticamente independentes, então

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}. \quad (2.17)$$

Nesse caso, a *pdf* conjunta $p(x, y)$ é dada por

$$\begin{aligned} p(x, y) &= \frac{d^2}{dxdy} P_{XY}(x, y) \\ &= \frac{d^2}{dxdy} P_X(x)P_Y(y) = p(x)p(y). \end{aligned} \quad (2.18)$$

Finalmente, ao se caracterizar a relação entre as v.a.'s X e Y , utiliza-se o conceito de covariância:

$$\begin{aligned} Cov(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY] - \mu_X\mu_Y, \end{aligned} \quad (2.19)$$

em que $\mu_X = E[X]$ e $\mu_Y = E[Y]$. Logo, uma vez que o produto entre duas v.a.'s contínuas e independentes é o produto das esperanças de cada v.a., tem-se que $Cov(X, Y) = 0$, se X e Y são independentes. Note-se que a covariância é uma generalização da variância para mais de uma v.a., em que $Cov(X, X) = Var(X)$.

2.1.1 Variável Aleatória Uniforme

Uma v.a. X é uniformemente distribuída em um intervalo (a, b) na reta real se a *pdf* $p(x)$ associada tiver a seguinte forma:

$$p(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{caso contrário} \end{cases}. \quad (2.20)$$

Nesse caso, utiliza-se a notação $X \sim \text{Uniforme}(a, b)$. Note-se que, se (a, b) é um intervalo unitário (i.e., $b - a = 1$), então para qualquer $a < x_1 < x_2 < b$,

$$P\{x_1 \leq X \leq x_2\} = P(x_2) - P(x_1) = \int_{-\infty}^{x_2} p(x)dx - \int_{-\infty}^{x_1} p(x)dx = x_2 - x_1. \quad (2.21)$$

Ademais,

$$\begin{aligned} E[X] &= \frac{a+b}{2}, \\ \text{Var}(X) &= \frac{(b-a)^2}{2}. \end{aligned} \quad (2.22)$$

2.1.2 A Densidade Gaussiana Multivariada

Uma função densidade bastante utilizada, em especial em conjunto com os métodos de estimação paramétricos, é a Gaussiana, ou distribuição Normal, a qual, no caso multivariado, é expressa como

$$p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^m |\boldsymbol{\Sigma}|}} e^{(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))}, \quad (2.23)$$

em que $\boldsymbol{\mu} = [\mu_1 \cdots \mu_m]^T$ é o vetor de médias dos componentes de \mathbf{x} , e $\boldsymbol{\Sigma}$ é a matriz de covariância de primeira ordem, definida como $\Sigma_{ij} = \text{Cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)]$. Observa-se que o coeficiente $\frac{1}{\sqrt{(2\pi)^m |\boldsymbol{\Sigma}|}}$ é um fator de normalização para assegurar a propriedade definida na equação (2.4), ou seja,

$$\frac{1}{\sqrt{(2\pi)^m |\boldsymbol{\Sigma}|}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))} dx_1 dx_2 \cdots dx_m = 1. \quad (2.24)$$

Finalmente, escreve-se $X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ para designar que a *pdf* de X é uma função Gaussiana centrada em $\boldsymbol{\mu}$ e com matriz de covariância $\boldsymbol{\Sigma}$.

2.2 Estatísticas de Ordem

Descreve-se nessa seção uma generalização para estatísticas de ordem multivariada, dada por Corley (1984). Esse é um conceito chave, uma vez que os modelos propostos no Capítulo 5 utilizam essa generalização para construir um tipo de gradiente estocástico.

Dado um conjunto de n v.a.'s contínuas e identicamente distribuídas, $D = \{X_i\}_{i=1}^n$, os elementos da sequência $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ são conhecidos como as *estatísticas de ordem* de D (Ross, 2001). Uma estatística de ordem de um conjunto de v.a.'s contínuas é também uma v.a. contínua a qual descreve a distribuição dos valores classificados em uma determinada posição da ordem subjacente à D . Por exemplo, se há um número ímpar de v.a.'s em D , denota-se a estatística que divide o conjunto de dados ordenados

exatamente na metade por $X_{([n+1]/2)}$. Essa estatística é conhecida como *mediana*. A *pdf* conjunta das estatísticas de ordem é dada por (Ross, 2001):

$$p_{X_{(1)}X_{(2)}\dots X_{(n)}}(x_1, x_2, \dots, x_n) = n!p(x_1)\dots p(x_n), \quad x_1 < x_2 < \dots < x_n. \quad (2.25)$$

2.2.1 A Função Densidade de uma Ordem Total na Reta Real

Para uma observação x , a probabilidade de que haja exatamente $i - 1$ valores menores que x e $n - i$ valores maiores que x é dada pelo produto $P\{X < x\}^{i-1}P\{X > x\}^{n-i}$. Uma vez que o número de combinações em que isso é possível é o coeficiente multinomial $\binom{n}{i-1, n-i, 1}$, logo, a *pdf* da estatística de i -ésima ordem é (Ross, 2001)

$$p_{X_{(i)}}(x) = \frac{n!}{(n-i)!(i-1)!}P(x)^{i-1}[1-P(x)]^{n-i}p(x). \quad (2.26)$$

Da mesma forma, $P\{i \text{ ou mais elementos do conjunto } D \text{ 's são } \leq x\}$, é dada por

$$P_{X_{(i)}}(x) = P\{X_{(i)} < x\} = \sum_{j=i}^n \binom{n}{j} P(x)^j [1-P(x)]^{n-j}. \quad (2.27)$$

2.2.2 Relações de Ordem Multivariadas

A noção de ordem em um espaço multidimensional não é trivial como para o caso da reta real. Isso porque não existe uma ordem total universal para classificar n vetores em \mathcal{R}^m . Uma forma simples de lidar com essa dificuldade consiste em reduzir a dimensionalidade dos vetores para uma única variável por meio de uma função de distância a um vetor específico, por exemplo, um vetor mediano \mathbf{x}_{med} definido como o elemento do conjunto de amostras com a seguinte propriedade (Pitas *et al.*, 1996):

$$\sum_{i=1}^n \|\mathbf{x}_i - \mathbf{x}_{\text{med}}\| \leq \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{x}_j\|, \quad \forall j \neq i, \quad (2.28)$$

em que $\|\mathbf{x} - \mathbf{y}\|$ representa a norma Euclidiana para dois vetores \mathbf{x} e \mathbf{y} . Essa técnica é um caso especial de uma ordem de p -norma, com $p = 2$, que é uma pré-ordem total (ver exemplos a seguir). Na otimização multiobjetivo, no entanto, procura-se extrair o máximo de informações sobre as relações de ordem entre pares de vetores no espaço de objetivos, de modo que uma ordem parcial, na inexistência de uma ordem total em \mathcal{R}^m , é amplamente adotada. Antes de apresentar a generalização de estatísticas de ordem

para o caso multivariado, definem-se e exemplificam-se a seguir algumas relações de ordem possíveis em \mathcal{R}^m , com destaque para a ordem de Pareto, a qual é de fundamental importância na otimização multiobjetivo evolucionária (Corley, 1984).

Definição 2.1 (Quasi-ordem). Uma relação binária \preceq em um subconjunto A de \mathcal{R}^m é uma quasi-ordem se

$$(\mathbf{x} \preceq \mathbf{x}) \forall \mathbf{x} \in A \quad (\text{reflexividade}) \quad (2.29)$$

Definição 2.2 (Pré-ordem). Uma quasi-ordem \preceq em A é uma pré-ordem se

$$(\mathbf{x} \preceq \mathbf{y}, \mathbf{y} \preceq \mathbf{z} \rightarrow \mathbf{x} \preceq \mathbf{z}) \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in A \quad (\text{transitividade}) \quad (2.30)$$

Definição 2.3 (Ordem parcial). Uma pré-ordem \preceq em A é uma ordem parcial se

$$(\mathbf{x} \preceq \mathbf{y}, \mathbf{y} \preceq \mathbf{x} \rightarrow \mathbf{x} = \mathbf{y}) \forall \mathbf{x}, \mathbf{y} \in A \quad (\text{antissimetria}) \quad (2.31)$$

Definição 2.4 (Ordem total). Uma ordem parcial \preceq em A é uma ordem total se

$$(\mathbf{x} \preceq \mathbf{y} \vee \mathbf{y} \preceq \mathbf{x}) \forall \mathbf{x}, \mathbf{y} \in A \quad (\text{totalidade}) \quad (2.32)$$

Por exemplo, a ordem usual na reta real é uma ordem total. A seguir, algumas ordens em \mathcal{R}^m são apresentadas:

- A *ordem majoritária*, definida como $\mathbf{x} \preceq \mathbf{y}$ se $x_i \leq y_i$ em ao menos $n/2$ componentes é uma quasi-ordem;
- A *ordem de p-norma*, definida como $\mathbf{x} \preceq \mathbf{y}$ se $\sum_i |x_i|^p \leq \sum_i |y_i|^p$, $p > 0$ é uma pré-ordem total;
- A *ordem de norma suprema*, definida como $\mathbf{x} \preceq \mathbf{y}$ se $\max_i (|x_1|, \dots, |x_n|)^T \leq \max_i (|y_1|, \dots, |y_n|)^T$ é uma pré-ordem total.

Definição 2.5 (Ordem de Pareto). Uma ordem parcial \preceq em \mathcal{R}^m é uma *ordem de Pareto* se $\mathbf{x} \preceq \mathbf{y} \rightarrow x_i \leq y_i, i = 1, \dots, m$.

2.2.3 A Função Densidade de uma Ordem de Pareto em \mathcal{R}^m

Corley (1984) generalizou a equação (2.26) para um conjunto $D = \{\mathbf{X}_i\}_{i=1}^n$ de n vetores

aleatórios $\mathbf{X}_i = (X_1 \cdots X_m)^T$ de m componentes:

$$p_{\mathbf{x}_{(i)}}(\mathbf{x}) = \frac{n!}{(n-i)!(i-1)!} P\{\mathbf{X}_{(j)} \preceq \mathbf{x}, j = 1, \dots, i-1; \mathbf{X}_{(j)} \not\preceq \mathbf{x}, j = i, \dots, n-1\} p(\mathbf{x}). \quad (2.33)$$

A partir da equação (2.33), é possível derivar as *pdfs* de estatísticas de ordem multivariadas, dentre elas, a que fora apresentada na equação (2.26) para uma ordem total em \mathcal{R}^1 . No contexto da otimização multiobjetivo evolucionária, é interessante calcular a probabilidade das estatísticas de i -ésima ordem em um espaço de objetivos m -dimensional. A seguir, apresenta-se o resultado da aplicação da equação (2.33) à ordem de Pareto em \mathcal{R}^m para os casos de independência e dependência entre os componentes de \mathbf{X}_i .

Caso 1: independência. Pela Definição 2.5, uma vez que $X_{(j)k} \leq x_k, k = 1, \dots, m$, tem-se:

$$P\{\mathbf{X}_{(j)} \preceq \mathbf{x}, j = 1, \dots, i-1\} = \left[\prod_{k=1}^m P_k(x_k) \right]^{i-1} e$$

$$P\{\mathbf{X}_{(j)} \not\preceq \mathbf{x}, j = i, \dots, n-1\} = \left[1 - \prod_{k=1}^m P_k(x_k) \right]^{n-i}, \quad (2.34)$$

em que $P_k(x_k)$ denota a *cdf* da k -ésima componente de $\mathbf{X}_{(j)}$, i.e., $P\{X_{(j)k} < x_k\}$. Dessa forma, obtém-se a *pdf* da estatística de i -ésima ordem de Pareto em \mathcal{R}^m para m v.a.'s independentes como:

$$p_{\mathbf{x}_{(i)}}(\mathbf{x}) = \frac{n!}{(n-i)!(i-1)!} \left[\prod_{k=1}^m P_k(x_k) \right]^{i-1} \left[1 - \prod_{k=1}^m P_k(x_k) \right]^{n-i} p(\mathbf{x}). \quad (2.35)$$

Caso 2: dependência. Quando as componentes dos vetores aleatórios possuem dependência estatística, basta utilizar a *cdf* conjunta

$$P(\mathbf{x}) = P\{X_1 < x_1, \dots, X_m < x_m\} = \int_{-\infty}^{x_m} \int_{-\infty}^{x_{m-1}} \cdots \int_{-\infty}^{x_1} p(\mathbf{x}) dx_1 \cdots dx_{m-1} dx_m, \quad (2.36)$$

obtendo:

$$p_{\mathbf{x}_{(i)}}(\mathbf{x}) = \frac{n!}{(n-i)!(i-1)!} [P(\mathbf{x})]^{i-1} [1 - P(\mathbf{x})]^{n-i} p(\mathbf{x}). \quad (2.37)$$

2.3 Estimação de Funções Densidade de Probabilidade

A estimação da função densidade de probabilidade constitui-se em uma etapa fundamental para vários processos decisórios, os quais objetivam encontrar uma política ótima para determinar a seqüência de ações que maximiza a recompensa esperada. A sinergia entre metodologias estatísticas e processos decisórios foi inicialmente estabelecida no trabalho de Wald (1939), em que testes de hipóteses, bem como a teoria da estimação estatística, foram identificados como casos especiais de uma teoria decisória mais geral. Como exemplo, regras de decisão podem ser estabelecidas a partir da análise de um conjunto de observações mediante uma abordagem Bayesiana, expressa pela regra de *maximum a posteriori*,

$$\text{Decisão} = \begin{cases} \omega_1, & \text{se } P(\omega_1|\mathbf{x}) > P(\omega_2|\mathbf{x}) \\ \omega_2, & \text{caso contrário} \end{cases} \quad (2.38)$$

em que ω_1, ω_2 expressam rótulos de classe, $\mathbf{x} \in \mathcal{R}^m$ é um vetor de atributos m -dimensional e $P(\omega_j|\mathbf{x})$ denota as probabilidades *a posteriori* da ocorrência da classe ω_j ao observar-se \mathbf{x} . Pode-se expressar, por sua vez, $P(\omega_j|\mathbf{x})$ pelo teorema de Bayes:

$$P(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{p(\mathbf{x})}. \quad (2.39)$$

O sucesso da construção de um classificador Bayesiano reside, portanto, na capacidade do projetista em aproximar a *pdf* dos vetores de observação, $p(\mathbf{x})$, a partir do conjunto de observações $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. A estimação de *pdfs* é um problema desafiador no sentido de que o número de observações necessárias para obter uma estimativa $\bar{p}(\mathbf{x})$ com precisão arbitrária cresce exponencialmente com a dimensionalidade do espaço de características (Freund *et al.*, 2009).

Essa seção tem por objetivo introduzir as principais técnicas de estimação de *pdfs*, paramétricas, não-paramétricas e semiparamétricas, com especial atenção à estimação baseada em *kernel*. Considera-se a análise de amostras m -dimensionais a partir de uma *pdf* $p(\mathbf{x})$ desconhecida, em que $\mathbf{x} \in \mathcal{R}^m$.

A finalidade do estudo das técnicas de estimação de *pdfs* para o presente trabalho reside na necessidade exploratória das amostras com as quais se pretende trabalhar. Espera-se que a estimação de densidade forneça informações ricas acerca de algumas características de interesse do espaço de objetivos em um problema de otimização multiobjetivo, tais como a existência de multimodalidade, assimetria, dependências entre variáveis, dentre outros fatores que influenciarão as ações de controle a serem empregadas para a geração de diversidade em algoritmos evolucionários.

Dessa forma, busca-se a identificação de uma ou mais técnicas com as seguintes características: (i) robustez, i.e., baixa sensibilidade em relação à quantidade de amostras existentes; (ii) relativa baixa complexidade computacional; e (iii) técnicas incrementais que possam atualizar a estimativa com relativa facilidade.

2.3.1 Estimação Paramétrica

As técnicas paramétricas supõem que $p(\mathbf{x})$ é desconhecida e que, todavia, possui uma forma paramétrica conhecida, $p(\mathbf{x}; \theta)$, em que θ é o vetor de parâmetros m -dimensional $[\theta_1 \cdots \theta_m]^T$. A maior vantagem dessas técnicas é o fato de que, caso o fenômeno estudado possua uma *pdf* conhecida, o processo de estimação se limita ao ajuste dos valores que parametrizam o modelo. A técnica mais conhecida é o Estimador de Máxima Verossimilhança (EMV).

Seja $X = [X_1 \cdots X_m]^T$ uma v.a. multivalorada. Na abordagem EMV, supõe-se uma família de distribuições conhecida, por exemplo, a densidade Gaussiana $X \sim \mathcal{N}(\mu, \Sigma)$ e objetiva-se estimar os parâmetros μ e Σ que melhor ajustem a *pdf* ao conjunto de treinamento, $\mathcal{D} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$ em que μ é o vetor de médias e Σ é a matriz de covariância que especifica o modelo de dependências lineares entre os componentes de X . Supondo que as observações $\mathbf{x}_k \in \mathcal{D}$ são independentes, pode-se escrever a verossimilhança de $\theta = [\mu \ \Sigma]^T$ em relação ao conjunto \mathcal{D} como

$$p(\mathcal{D}|\theta) = \prod_{k=1}^N p(\mathbf{x}_k|\theta), \quad (2.40)$$

de forma que a estimativa

$$\hat{\theta} = \arg \max_{\theta} P(\mathcal{D}|\theta)$$

é a de máxima verossimilhança (Duda *et al.*, 2000). Para a distribuição Normal, obtêm-se analiticamente o EMV para θ a partir do logaritmo da verossimilhança (log-verossimilhança, $l(\theta)$), i.e.,

$$\begin{aligned} l(\theta) &= \ln p(\mathcal{D}|\theta) = \ln \prod_{k=1}^N p(\mathcal{D}|\theta) \\ &= \sum_{k=1}^N \ln p(\mathcal{D}|\theta). \end{aligned} \quad (2.41)$$

Igualando o gradiente da log-verossimilhança a zero, ou seja, $\nabla_{\theta} l(\theta) = 0$, obtêm-se

$$\begin{aligned}\hat{\theta} &= [\hat{\mu} \quad \hat{\Sigma}] \\ &= \left[\frac{1}{n} \sum_{k=1}^N \mathbf{x}_k \quad \frac{1}{n} \sum_{k=1}^N (\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^T \right]^T, \end{aligned} \quad (2.42)$$

em que $\hat{\mu}$ é a média amostral (o vetor médio no caso multivariado) e $\hat{\Sigma}$ é a matriz de covariância amostral, calculada a partir de $\hat{\mu}$. Note-se que o EMV é viesado. Observa-se também que o EMV possui ao menos duas limitações práticas: (i) custo proibitivo para o cálculo da matriz de covariância amostral para problemas de alta dimensionalidade, de complexidade $O(m^2N)$; e (ii) o método não é robusto no sentido de que, caso a suposição inicial do modelo seja pobre, o erro induzido pelo EMV para um parâmetro desconhecido pode ser grande, além de não haver garantias que o EMV sobre o parâmetro de interesse fornece a melhor estimativa possível em tal situação (Duda *et al.*, 2000).

Diferentemente do EMV, a estimação Bayesiana trata o vetor de parâmetros θ como uma variável aleatória. Ambas as técnicas, entretanto, estimam valores quase idênticos para os parâmetros de interesse. Supõe-se inicialmente que o vetor de parâmetros θ possui densidade $p(\theta)$ conhecida. A observação dos exemplos do conjunto de treino \mathcal{D} , porém, torna possível o uso da densidade *a posteriori* $p(\theta|\mathcal{D})$ para a estimação de $p(\mathbf{x}|\mathcal{D})$, que consiste na melhor aproximação para $p(\mathbf{x})$ possível (Duda *et al.*, 2000). Para tanto, considera-se todo o espaço de hipóteses e integra-se a densidade conjunta $p(\mathbf{x}, \theta|\mathcal{D})$ em relação a θ :

$$p(\mathbf{x}|\mathcal{D}) = \int p(\mathbf{x}, \theta|\mathcal{D}) d\theta. \quad (2.43)$$

Outrossim, Duda *et al.* (2000) observam que $p(\mathbf{x}, \theta|\mathcal{D})$ pode ser expressa como o produto $p(\mathbf{x}|\theta, \mathcal{D})p(\theta|\mathcal{D})$ e, uma vez que a escolha de \mathbf{x} independe de \mathcal{D} , a estimação de $p(\mathbf{x})$ reduz-se à avaliação, através de simulação Monte Carlo, da integral

$$p(\mathbf{x}|\mathcal{D}) = \int p(\mathbf{x}|\theta)p(\theta|\mathcal{D}) d\theta, \quad (2.44)$$

em que, pelo teorema de Bayes,

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int p(\mathcal{D}|\theta)p(\theta) d\theta}. \quad (2.45)$$

Supondo independência entre as observações em \mathcal{D} , têm-se:

$$p(\mathcal{D}|\theta) = \prod_{k=1}^N p(\mathbf{x}_k|\theta). \quad (2.46)$$

A equação (2.46) pode ser reescrita recursivamente, para $N > 1$, como

$$p(\mathcal{D}^N|\theta) = p(\mathbf{x}_N|\theta)p(\mathcal{D}^{N-1}|\theta), \quad (2.47)$$

em que a notação \mathcal{D}^N expressa um conjunto de exemplos de cardinalidade $|\mathcal{D}| = N$. Note-se que o caso base da recursão é $p(\mathcal{D}^0|\theta) = p(\theta)$. Substituindo-se a equação (2.45) na equação (2.47), obtêm-se

$$p(\theta|\mathcal{D}^N) = \frac{p(\mathbf{x}_N|\theta)p(\theta|\mathcal{D}^{N-1})}{\int p(\mathbf{x}_N|\theta)p(\theta|\mathcal{D}^{N-1})d\theta}. \quad (2.48)$$

A equação (2.47) produz a sequência $p(\theta), p(\theta|\mathbf{x}_1), \dots, p(\theta|\mathbf{x}_1 \cdots \mathbf{x}_N)$. Desta forma, o método de aprendizado Bayesiano recursivo pode ser visto como um método incremental. Na prática, todo o conjunto de exemplos é mantido em memória para o cálculo de $p(\mathcal{D}^N|\theta)$, como observam Duda *et al.* (2000).

A amostragem de Gibbs (Geman e Geman, 1984) é um método baseado em simulação Monte Carlo em Cadeia de Markov utilizado para avaliar numericamente a integral resultante do aprendizado Bayesiano recursivo. A amostragem de Gibbs gera novas amostras de forma incremental, a partir das amostras existentes, até que o processo convirja para uma distribuição estacionária, a qual independe do estado inicial. Após algumas iterações, todas as amostras geradas seguirão a densidade alvo a qual se deseja estimar. Nessa abordagem, as probabilidades de transição entre as amostras dependem apenas da observação mais recente.

2.3.2 Estimação Não-Paramétrica

As técnicas de estimação não-paramétricas dispensam a especificação de uma forma funcional para a densidade, ao custo de uma menor eficiência na estimação caso a forma correta possa ser analiticamente estabelecida *a priori*, o que é vantajoso uma vez que é desnecessário especificar-se uma forma paramétrica a qual pode não capturar a complexidade da *pdf* subjacente. Em particular, os métodos clássicos de estimação de densidade paramétricos supõem unimodalidade, o que introduz uma limitação quanto à

2.3. ESTIMAÇÃO DE FUNÇÕES DENSIDADE DE PROBABILIDADE

aplicação em problemas envolvendo densidades multimodais. A seguir apresentam-se os principais métodos não-paramétricos.

Quando aplicados a problemas contínuos, modelos baseados em histograma provêm uma discretização do espaço de atributos a partir da qual é possível estimar probabilidades em regiões específicas. Em comparação com a densidade Gaussiana unimodal, os histogramas possibilitam a representação de vários picos de diferentes alturas. No caso univariado, dado o conjunto de observações $\mathcal{D} = \{x_n\}_{n=1}^N$, em que N é o tamanho da amostra, a *pdf* de interesse é aproximada através da divisão da reta real em M intervalos de largura $2h$. Dada uma origem x_0 e a largura $2h$, o intervalo construído para cada barra B_j será $[x_0 + jh, x_0 + (j+1)h]$, com $j = 1, \dots, M$. A altura de cada barra é proporcional ao número de amostras observadas nos respectivos intervalos. Para que a integral da estimativa resulte some 1, introduz-se um fator de normalização de modo que

$$\hat{p}(x) = \frac{1}{2Nh} \sum_{n=1}^N \sum_{j=1}^M I(x_n, B_j) I(x, B_j), \quad (2.49)$$

em que $I(\cdot)$ é uma função indicadora definida como

$$I(a, A) = \begin{cases} 1, & \text{se } a \in A \\ 0, & \text{caso contrário} \end{cases} \quad (2.50)$$

Note-se que a escolha da origem e largura controla a suavidade da estimativa. O modelo também pode ser generalizado para larguras variáveis em cada barra, provendo maior precisão na estimativa.

Para problemas multivariados, uma possível abordagem consiste na aproximação da *pdf* de interesse mediante o produto das densidades marginais de cada variável, construídas a partir de histogramas univariados (Ding e Zhou, 2008). Obviamente, essa técnica não leva em consideração o modelo de dependências entre os atributos e pode resultar em uma aproximação pobre para a *pdf* de interesse. Histogramas multivariados podem ser utilizados para considerar o modelo de dependências. Entretanto, além da complexidade computacional exponencial em função do número de atributos, a quantidade de amostras necessárias para uma boa aproximação também cresce rapidamente. A razão reside no fato de que, em espaços de alta dimensionalidade, a probabilidade da não observação de amostras em certas regiões é bastante elevada, o que resulta em descontinuidades indesejadas em torno das regiões esparsas de baixa densidade.

A estimação por *kernel*, também conhecida como estimação por janelas de Parzen (Parzen, 1962), advém da observação de que, no cálculo da probabilidade de uma amostra

\mathbf{x} ser encontrada em uma região R (equação (2.1)) suficientemente pequena, de tal forma que $p(\mathbf{x})$ seja praticamente constante em R , têm-se a seguinte aproximação:

$$P\{\mathbf{x} \in R\} = \int_R p(\mathbf{x}) d\mathbf{x} \approx p(\mathbf{x}) \int_R d\mathbf{x} = p(\mathbf{x}) V_R, \quad (2.51)$$

em que V_R é o volume de R , da qual se obtêm a estimativa

$$p(\mathbf{x}) \approx \frac{P_R(\mathbf{x})}{V_R}. \quad (2.52)$$

Aproximando-se $P_R(\mathbf{x})$ pela frequência relativa entre o número k de amostras que caem em R a partir das N observações contidas em \mathcal{D} , obtêm-se

$$p(\mathbf{x}) \approx \frac{k/N}{V_R}. \quad (2.53)$$

A estimação da densidade por *kernel* considera, então, um volume V_R fixo, porém suficientemente pequeno, de tal forma que $p(\mathbf{x})$ possa ser aproximada satisfatoriamente. Por exemplo, suponha que R é um hipercubo m -dimensional centrado em \mathbf{x} , cujo comprimento da aresta é h e volume $V_R = h^m$. Então, a função

$$\phi\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) = \begin{cases} 1, & \text{se } \frac{|x_{ij} - x_j|}{h} \leq 1/2, j = 1, \dots, m \\ 0, & \text{caso contrário} \end{cases}, \quad (2.54)$$

que indica se a observação \mathbf{x}_i encontra-se no hipercubo R , pode ser incorporada na equação (2.53), expressando o número total de amostras k que caem no hipercubo:

$$k = \sum_{i=1}^N \phi\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right). \quad (2.55)$$

Assim, obtêm-se a seguinte estimativa para $p(\mathbf{x})$ (Duda *et al.*, 2000):

$$\hat{p}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^N \frac{1}{h^m} \phi\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right), \quad (2.56)$$

em que $\phi(\cdot)$ é conhecida como *função janela* ou *função de kernel* e h é um parâmetro de suavização. Outras funções *kernel* podem ser utilizadas na estimativa da *pdf*, como a função Gaussiana da equação (2.23), em que, ao se escolher $h = 1$ e $\sigma^2 = 1$ no caso unidimensional, obtêm-se a estimativa para a densidade como simplesmente uma média

2.3. ESTIMAÇÃO DE FUNÇÕES DENSIDADE DE PROBABILIDADE

de Gaussianas centradas em cada amostra de \mathcal{D} :

$$\hat{p}(x) = \frac{1}{n} \sum_{i=1}^N \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{(x_i-x)^2}{2}\right)}. \quad (2.57)$$

Por exemplo, a *pdf* estimada por *kernel* Gaussiano da equação (2.57) para o conjunto de amostras $\mathcal{D} = \{5/2, 2, 1, 3, 6\}$, $d = 1$, pode ser visualizada na Figura 2.1. Observe-se que a estimativa é capaz de capturar uma aparente multimodalidade na distribuição dos dados descritos pelo conjunto \mathcal{D} , apesar do número de amostras ser pequeno, com uma boa granularidade. A qualidade da estimativa, entretanto, depende da escolha do parâmetro de suavização h . Como observado em Webb (2002), à medida que h aproxima-se de zero, a *pdf* estimada tende a apresentar picos abruptos em torno de cada amostra. Um procedimento utilizado para a determinação de h na estimação por *kernel* no caso unidimensional é, fixado o vetor \mathbf{x} , escolher h proporcional à distância média para a K -ésima amostra mais próxima em \mathcal{D} (Webb, 2002). De fato, quanto maior o valor escolhido para h , menor será a variância da estimativa e vice-versa. Ademais, quão menor for o valor de h , melhor será a qualidade da estimativa para o conjunto de dados apresentado. Essas propriedades decorrem dos seguintes limites:

$$\begin{aligned} \lim_{n \rightarrow \infty} nh = \infty &\Rightarrow \lim_{n \rightarrow \infty} \text{Var}[\hat{p}(\mathbf{x})] = 0, \forall \mathbf{x} \in \mathcal{R} \text{ e} \\ \lim_{n \rightarrow \infty} h = 0 &\Rightarrow \lim_{n \rightarrow \infty} E[\hat{p}(\mathbf{x})] = p(\mathbf{x}), \forall \mathbf{x} \in \mathcal{R}. \end{aligned} \quad (2.58)$$

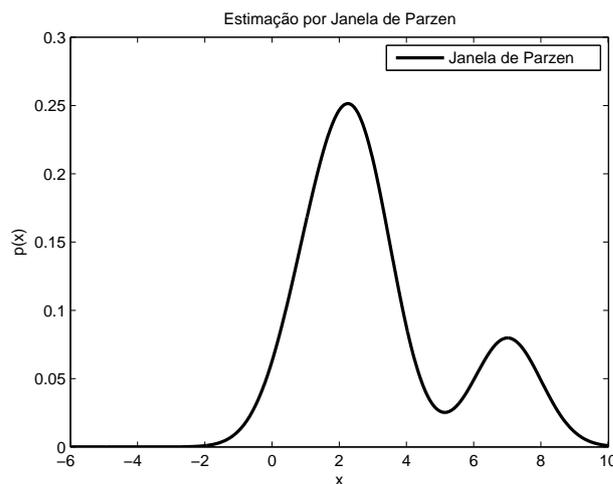


Figura 2.1 Estimação da densidade por Kernel Gaussiano.

2.3.3 Estimação Semiparamétrica

Finalmente, as técnicas semiparamétricas, apesar de, assim como as técnicas não-paramétricas, não suporem uma família de distribuições conhecida para a estimação da *pdf*, fixam previamente a complexidade do modelo para lidar com o principal problema encontrado nos modelos não-paramétricos – a *maldição da dimensionalidade* – e também para evitar sobreajuste (Assenza *et al.*, 2008). Na abordagem de mistura finita de Gaussianas, a densidade alvo é aproximada por uma combinação linear de g densidades Gaussianas, isto é,

$$p(\mathbf{x}) = \sum_{j=1}^g \pi_j p(\mathbf{x}; \theta_j), g \ll N, \quad (2.59)$$

em que $\sum_j \pi_j = 1$, $\pi_j \geq 0$ e $p(\mathbf{x}; \theta_j)$ são densidades Gaussianas parametrizadas por θ_j . Desta forma, são três os conjuntos de parâmetros a serem estimados, a saber: (i) os valores dos pesos π_j ; (ii) os componentes de θ_j ; e (iii) o número de misturas g .

Webb (2002) observa que, em geral, não é possível resolver a equação $\frac{\partial L}{\partial \Psi} = 0$ e, que portanto, métodos iterativos como o algoritmo *Expectation, maximization* (EM), são preferíveis para estimar o conjunto de parâmetros $\Psi = \{\pi_1, \dots, \pi_g, \theta_1, \dots, \theta_g\}$ através da maximização da função de verossimilhança

$$L(\Psi) = \prod_{i=1}^N \sum_{j=1}^g \pi_j p(\mathbf{x}_i; \theta_j). \quad (2.60)$$

2.3.4 Estimação Multivariada Baseada em *Kernel*

Considere um vetor aleatório m -dimensional $\mathbf{X} = (X_1 \cdots X_m)^T$ e um conjunto de amostras $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. *Kernel* multivariados da forma $\varphi(\mathbf{x}) = \varphi(x_1, \dots, x_m)$ podem ser usados para estimar a *pdf* conjunta $p(\mathbf{x}) = p(x_1, \dots, x_m)$ (Müller *et al.*, 2004). A abordagem mais simples consiste em escolher h constante para todas as dimensões. Se as v.a.'s forem independentes e identicamente distribuídas, essa pode ser uma escolha adequada. Caso as v.a.'s sejam independentes, porém não identicamente distribuídas, uma estratégia mais refinada é escolher, para cada dimensão, um conjunto de m valores apropriados para h . Nesse caso, a estimação é descrita por

$$\hat{p}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h_1 \cdots h_m} \varphi\left(\frac{x_1 - x_{i1}}{h_1}, \dots, \frac{x_m - x_{im}}{h_m}\right). \quad (2.61)$$

Ao fazer uso do *kernel* multiplicativo $\varphi(\mathbf{x}) = \phi(x_1)\phi(x_2)\cdots\phi(x_m)$, têm-se

$$\hat{p}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \left\{ \prod_{j=1}^m \frac{1}{h_j} \phi\left(\frac{x_j - X_{ij}}{h_j}\right) \right\}. \quad (2.62)$$

Uma forma mais geral para $\hat{p}(\mathbf{x})$ considera uma matriz não-singular $m \times m$ (i.e., uma matriz quadrada invertível) \mathbf{H} para codificar os parâmetros de suavização h_1, \dots, h_m na forma vetorial $\mathbf{h} = (h_1, \dots, h_m)^T$, representados em uma forma diagonal de \mathbf{H} . Nesse caso, usando a notação $\varphi_{\mathbf{H}}(\cdot) = \frac{1}{\det(\mathbf{H})} \phi(\mathbf{H}^{-1}\cdot)$, a estimação é realizada por

$$\hat{p}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \varphi_{\mathbf{H}}(\mathbf{x} - \mathbf{x}_i). \quad (2.63)$$

Note-se que, no caso de um único parâmetro de suavização para todas as dimensões, $\mathbf{H} = h\mathbf{I}_m$, em que \mathbf{I}_m é a matriz identidade $m \times m$. Ademais, para capturar dependências lineares entre as v.a.'s, é conveniente escolher \mathbf{H} proporcional à raiz quadrada da matriz de covariância amostral dos dados (Müller *et al.*, 2004), i.e., $\mathbf{H} \propto \hat{\Sigma}^{1/2}$. Uma regra prática (Scott, 1992) consiste em tomar

$$\mathbf{H} = N^{\frac{-1}{m+4}} \hat{\Sigma}^{1/2}. \quad (2.64)$$

O método descrito pela equação (2.63) é utilizado no modelo de otimização evolucionária multiobjetivo formulado no Capítulo 4, por fornecer as características e a flexibilidade desejada para os propósitos do presente trabalho.

2.4 Discussão Final e Síntese

Apresentaram-se nesse Capítulo os conceitos estatísticos utilizados ao longo da dissertação. Especificamente, definiram-se os conceitos e objetos matemáticos incorporados no desenvolvimento dos modelos algorítmicos estudados nos Capítulos 4, 5 e 6.

Identificou-se o método de estimação de *pdf* baseado em *kernels* multivariado como o mais apropriado para a aplicação compreendida nessa dissertação. A escolha justifica-se, primeiro, por ser um método não-paramétrico em que a suposição de um modelo de *pdf* prévio não se faz necessária. Esse é um aspecto desejável na aplicação desenvolvida, pois, no caso de vetores distribuídos em espaços de funções-objetivo, há dependências não-lineares evidentes entre as variáveis de modo que, na forma paramétrica, haveria a necessidade de supor modelos iniciais complexos. Segundo, o método é robusto o sufici-

ente para representar com acurácia características comumente observadas na distribuição de pontos sobre a Frente de Pareto (ver próximo Capítulo, página 29): multimodalidade, assimetria e dependências não-lineares entre as variáveis. Por fim, considerou-se a precisão da estimação como um critério importante para as investigações iniciais do modelo proposto no Capítulo 5 como forma de reduzir a influência do erro de estimação nas ações empregadas para a geração de diversidade nos Algoritmos Evolucionários (AEs) estudados.

No que concerne às Estatísticas de Ordem Multivariada (EOM) (Seção 2.2), ressalta-se a utilização crescente desse conceito em diversas aplicações. Por exemplo, na inteligência computacional, EOMs são incorporadas em um algoritmo de classificação de padrões baseado em mapas auto-organizáveis para tratar a presença de valores atípicos (*outliers*) no conjunto de treinamento. O algoritmo é aplicado à organização de documentos (Georgakis *et al.*, 2004). Na mesma linha de trabalho, Pitas *et al.* (1996) obtêm uma generalização para o algoritmo *Learning Vector Quantization* por meio de EOM para melhorar a robustez de métodos de quantização vetorial de imagens coloridas na presença de *outliers*. Na computação evolucionária (ver Capítulo 3), Miller e Goldberg (1996) utilizam estatísticas de ordem univariadas para computar a “*intensidade de seleção*” de um operador de torneio em AEs. Finalmente, Barakat (2009) sugere a aplicação de EOM no escalonamento de tarefas em metodologias de projeto baseadas em caminho crítico.

O conteúdo desse Capítulo contribui da seguinte forma para o desenvolvimento da dissertação: (i) as relações de ordem multivariada apresentadas na Seção 2.2.2 contribuem para o entendimento das definições de otimalidade e relações de dominância na otimização multiobjetivo, discutidas no Capítulo 3; (ii) no Capítulo 4, a variância de um modelo composto por duas v.a.'s contínuas é interpretada como a diversidade genética de duas subpopulações em um Algoritmo Evolucionário (AE); e, finalmente, (iii) no Capítulo 5, a técnica de estimação de *pdf* multivariada baseada em *kernel* é utilizada para modelar a população de um AE no espaço de objetivos no modelo de geração de diversidade proposto e, além disso, a *pdf* de uma ordem de Pareto em \mathcal{R}^m (equação (2.37)) constitui a base para as principais proposições elaboradas nessa dissertação.

3

Fundamentos de Otimização Evolucionária

Nós somos máquinas de sobrevivência – veículos robotizados cegamente programados para preservar moléculas egoístas chamadas de genes.

– Richard Dawkins

O propósito desse Capítulo é apresentar os conceitos básicos e o formalismo matemático dos modelos de otimização contínua, dinâmica e multiobjetivo, os quais permeiam o projeto das técnicas de busca evolucionária e geração de diversidade empregadas nessa dissertação. Os fundamentos da computação evolucionária e os Algoritmos Evolucionários (AEs) multiobjetivos implementados nesse trabalho também são detalhados.

3.1 Otimização Contínua

São introduzidos nessa seção os conceitos e fundamentos que permeiam a área da matemática conhecida como *otimização*, bem como uma de suas grandes subáreas, a saber, a que versa sobre a busca de máximos e/ou mínimos de funções real-valoradas, a *otimização contínua*.

Seja \mathbf{x} um vetor de variáveis, f uma função-objetivo a ser maximizada ou minimizada (a partir desse ponto, supõe-se por conveniência minimização, a menos quando apontado o contrário) e \mathbf{c} um vetor de restrições sobre as variáveis. A formulação matemática para o problema de otimização contínua é a seguinte (Nocedal e Wright, 2006):

$$\min_{\mathbf{x} \in \mathcal{R}^m} f(\mathbf{x}) \quad \text{sujeito a} \begin{cases} c_i = 0, & i \in \mathcal{E}, \\ c_i \geq 0, & i \in \mathcal{I}. \end{cases} \quad (3.1)$$

em que $f : \mathcal{R}^m \mapsto \mathcal{R}$, $c_i : \mathcal{R}^m \mapsto \mathcal{R}$ são funções contínuas, geralmente diferenciáveis, \mathcal{E} , \mathcal{I} são conjuntos de índices e m é a dimensionalidade do espaço de entrada. Como exemplo, considere o seguinte problema de otimização:

$$\min 2x_1 + 2,5x_2^2 \quad \text{sujeito a} \quad \begin{cases} 2x_1^2 - 16x_2 \leq 0 \\ -3x_1 - 2x_2 \leq -20 \\ x_1 \geq 4 \\ x_2 \leq 6 \end{cases} \quad (3.2)$$

Esse problema pode ser escrito na forma padrão da equação (3.1) como

$$f(\mathbf{x}) = 2x_1 + 2,5x_2^2, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

$$\mathbf{c}(\mathbf{x}) = \begin{bmatrix} c_1(\mathbf{x}) \\ c_2(\mathbf{x}) \\ c_3(\mathbf{x}) \\ c_4(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} -2x_1^2 + 16x_2 \\ 3x_1 + 2x_2 - 20 \\ x_1 - 4 \\ -x_2 + 6 \end{bmatrix}, \quad \mathcal{I} = \{1, 2, 3, 4\}, \quad \mathcal{E} = \emptyset. \quad (3.3)$$

A área hachurada na Figura 3.1 indica a região factível para o problema (3.3), i.e., o conjunto dos pontos do *espaço de busca* (ou de decisão) que são viáveis como solução, segundo as restrições fornecidas. A figura exhibe todas as restrições $c_i(\mathbf{x})$ que delimitam a região factível em \mathcal{R}^2 . Também é indicado no gráfico o ponto ótimo da função $f(\mathbf{x})$, o vetor $\mathbf{x}^* = [4, 77 \ 2, 84]^T$, pertencente à região factível, o qual fornece o menor valor quando avaliado pela função-objetivo, i.e., $f(\mathbf{x}^*) < f(\mathbf{x}), \forall \mathbf{x} \in \Omega \setminus \{\mathbf{x}^*\}$, em que Ω é o conjunto de vetores factíveis (ou viáveis).

Os problemas de otimização diferem em diversos aspectos. Por exemplo, se a função-objetivo e as restrições são lineares, tem-se um problema *linear*. Se ao menos uma restrição ou a função-objetivo for não-linear, tem-se um problema *não-linear*. Outro aspecto a se considerar é o domínio das variáveis. O problema do exemplo anterior pertence à classe de *otimização contínua*, uma vez que todas as variáveis assumem valores continuamente em \mathcal{R} . Apresentam-se a seguir algumas definições de interesse, sob o ponto de vista da otimização contínua, as quais são úteis para o entendimento de discussões posteriores. Para maiores detalhes, c.f. Loesch e Hein (2008).

Definição 3.1 (Bola aberta). Uma bola aberta B de centro \mathbf{x}_0 e raio $\varepsilon > 0$ é o conjunto dos pontos \mathbf{x} que possuam uma distância inferior a ε com respeito à \mathbf{x}_0 : $B(\mathbf{x}_0, \varepsilon) = \{\mathbf{x} \in \mathcal{R}^m : \|\mathbf{x} - \mathbf{x}_0\| < \varepsilon\}$.

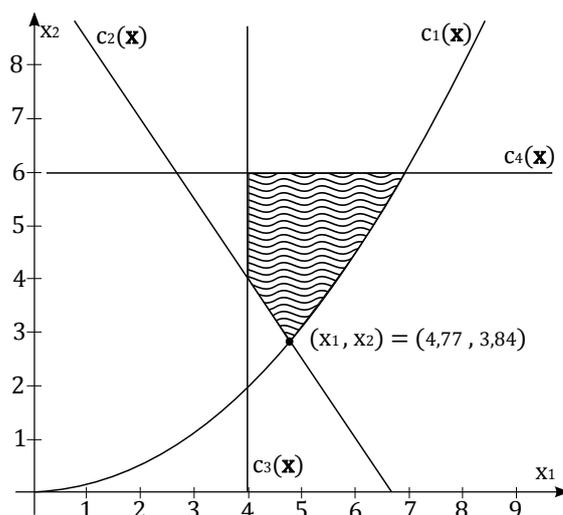


Figura 3.1 Representação do espaço de soluções factíveis para o problema da equação (3.3).

Definição 3.2 (Ponto interior). Um ponto $\mathbf{v} \in A \subset \mathcal{R}^m$ é um ponto interior se $\exists \varepsilon > 0 \mid B(\mathbf{v}, \varepsilon) \subset A$.

Definição 3.3 (Ponto de fronteira). Um ponto $\mathbf{w} \in \mathcal{R}^m$ é ponto de fronteira de A se $\Omega_1 \subset B(\mathbf{w}, \varepsilon)$ e $\Omega_2 \subset B(\mathbf{w}, \varepsilon)$, em que $\Omega_1 \subset A$ e $\Omega_2 \subset \mathcal{R}^m - A, \forall \varepsilon > 0$.

Definição 3.4 (Fronteira). A fronteira de A , $fr(A)$, é o conjunto de todos os pontos de fronteira. Dado um elemento $\mathbf{x} \in A$, \mathbf{x} é ou ponto interior ou ponto de fronteira de A .

Definição 3.5 (Segmento de reta). O conjunto $\{\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2 : \lambda \in [0, 1]\}$ é o segmento de reta $\overline{\mathbf{x}_1 \mathbf{x}_2}$ com pontos extremos $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{R}^m$.

Definição 3.6 (Conjunto convexo). Uma condição suficiente e necessária para que o conjunto $A \subset \mathcal{R}^m$ seja convexo é $\forall \mathbf{x}_1, \mathbf{x}_2 \in A \rightarrow \overline{\mathbf{x}_1 \mathbf{x}_2} \subset A$.

Definição 3.7 (Pontos extremos). Dado um conjunto convexo $A \subset \mathcal{R}^m$, o ponto $\mathbf{x} \in A$ é ponto extremo de A se a combinação convexa $\mathbf{x} = \lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2$, com $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{R}^m$ e $\lambda \in [0, 1]$, implicar $\mathbf{x} = \mathbf{x}_1 = \mathbf{x}_2$.

3.2 Otimização Multiobjetivo

A consideração de mais de uma função-objetivo no problema de otimização induz à necessidade de uma análise de custo-benefício das soluções obtidas. A depender do critério de importância de cada objetivo para o tomador de decisão, algumas soluções podem ser mais atraentes do que outras. As informações de custo-benefício, ou simplesmente de *tradeoff*, versam sobre como a melhoria em relação a um objetivo pode afetar negativamente os demais objetivos (Eichfelder, 2008). Nesse sentido, a otimização multiobjetivo é de fundamental importância, em especial se o problema necessitar a consideração de dois ou mais objetivos conflitantes entre si. A seguir são apresentados conceitos e definições úteis.

Formalmente, um problema de otimização multiobjetivo (MOP, *Multiobjective Optimization Problem*, do inglês) possui a seguinte descrição:

$$\min_{\mathbf{x} \in \Omega} \mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_M(\mathbf{x}) \end{bmatrix} \quad M \geq 2, \quad (3.4)$$

em que

$$\begin{aligned} \Omega \subseteq \mathcal{R}^m &= \{\mathbf{x} \in \mathcal{R}^m : \mathbf{h}(\mathbf{x}) = \mathbf{0}_p, \mathbf{g}(\mathbf{x}) \geq \mathbf{0}_q\}, \\ \mathbf{f} : \mathcal{R}^m &\mapsto \mathcal{R}^M, \mathbf{h} : \mathcal{R}^m \mapsto \mathcal{R}^p \text{ e } \mathbf{g} : \mathcal{R}^m \mapsto \mathcal{R}^q \quad (p, q \in \mathcal{N}) \end{aligned} \quad (3.5)$$

com \mathcal{N} denotando o conjunto dos naturais. Ademais, $f_i : \mathcal{R}^m \mapsto \mathcal{R}$, $i = 1, \dots, M$. As funções \mathbf{h} , \mathbf{g} e f_i são contínuas e duplamente diferenciáveis. Note-se que Ω é o conjunto de vetores factíveis, onde, idealmente, a busca é realizada. Ainda, m é a dimensionalidade do espaço de decisão e M é o número de funções-objetivo.

Dois conceitos úteis para a comparação da qualidade entre pares de solução obtidas em MOPs são a *dominância fraca de Pareto* e a *dominância de Pareto*, definidas no espaço de objetivos.

Definição 3.8 (Dominância fraca de Pareto). Sejam $\mathbf{v}, \mathbf{u} \in F_M$, em que F_M é o espaço M -dimensional de objetivos. O vetor \mathbf{v} domina fracamente \mathbf{u} , denotado por $\mathbf{v} \preceq \mathbf{u}$, $\iff (v_i \leq u_i \forall i \in \{1, \dots, M\})$.

Definição 3.9 (Dominância de Pareto). O vetor \mathbf{v} domina \mathbf{u} , denotado por $\mathbf{v} \prec \mathbf{u}$, $\iff (\mathbf{v} \preceq \mathbf{u}) \wedge (\exists j \in \{1, \dots, M\} \mid v_j < u_j)$.

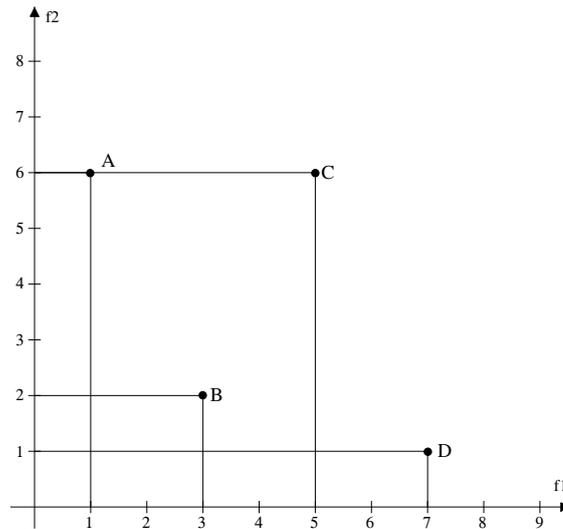


Figura 3.2 Representação do espaço de objetivos. Têm-se as seguintes relações de dominância: $A \prec C, B \prec C$; mas A e B não são comparáveis, i.e., $A \not\prec B$ e $B \not\prec A$.

Desse modo, diz-se que uma solução \mathbf{x}_1 *domina* \mathbf{x}_2 ($\mathbf{x}_1 \prec \mathbf{x}_2$) se $\mathbf{f}(\mathbf{x}_1) \prec \mathbf{f}(\mathbf{x}_2)$. A dominância de Pareto (Figura 3.2) estende-se naturalmente aos conjuntos de soluções não-dominadas em que $(A \subseteq \Omega \preceq B \subseteq \Omega) \iff (\forall \mathbf{y} \in B) (\exists \mathbf{x} \in A) : (\mathbf{x} \preceq \mathbf{y})$.

O conceito de otimalidade de Pareto é comumente utilizado para a análise das soluções obtidas em MOPs, dado a dificuldade de encontrar uma única solução que otimize todas as funções-objeto f_i simultaneamente. Em problemas multiobjetivo, as soluções ótimas são aquelas que não são dominadas por qualquer outra solução.

Definição 3.10 (Ótimo de Pareto). Um ponto $\mathbf{x}^* \in \Omega$ é um ótimo de Pareto, ou ponto eficiente, ou solução não-dominada $\iff \nexists \mathbf{x} \in \Omega \mid \mathbf{f}(\mathbf{x}) \prec \mathbf{f}(\mathbf{x}^*)$ (Das e Dennis, 1996).

A definição do ótimo de Pareto contrasta com a de uma solução ótima na otimização de um único objetivo no sentido de que, com apenas um objetivo a ser otimizado, apesar de ser possível a existência de várias soluções ótimas no espaço de decisão, os ótimos são mapeados para o mesmo valor da função-objeto. Isso significa que há apenas um único valor ótimo global no espaço de objetivos. Já em problemas multiobjetivo, os ótimos de Pareto podem ser mapeados para diferentes vetores da função-objeto. Em outras palavras, deseja-se obter um conjunto de ótimos de Pareto o qual represente o *tradeoff* entre os diferentes objetivos, em contraste com a obtenção de uma única solução na otimização mono-objetivo (Goh e Tan, 2009).

O conjunto de todas as soluções ótimas de Pareto no espaço de decisão, $\Omega^* \subseteq \Omega$, é chamado de *conjunto de Pareto*, enquanto a imagem de Ω^* no espaço de objetivos,

$F_M^* = \mathbf{f}(\Omega^*) \subseteq F_M$, é chamada de *Frente de Pareto (PF, Pareto Front, do inglês)*. Note-se que a *PF* é a Fronteira do conjunto de soluções factíveis no espaço de objetivos (ver Definição 3.4).

Deb (2001) sugere que a melhor abordagem para resolver MOPs consiste em duas etapas: (i) obter uma aproximação, \mathcal{F}_1 , para a *PF*; e (ii) escolher, dentre todas as soluções obtidas, aquela mais adequada em relação às preferências do tomador de decisão.

3.2.1 Indicadores de Proximidade, Espaçamento e Cobertura

Em otimização multiobjetivo, a questão fundamental na avaliação de desempenho de diferentes algoritmos é a descrição dos conjuntos de aproximação obtidos (compostos apenas por soluções mutuamente não-dominadas) em termos quantitativos. Para tanto, utilizam-se indicadores de qualidade, definidos como:

Definição 3.11 (Indicador de qualidade). Um indicador de qualidade k -ário \mathcal{I} é um mapeamento $\mathcal{I} : \underbrace{\Omega \times \cdots \times \Omega}_k \mapsto \mathcal{R}$ o qual atribui para cada vetor de k aproximações (A_1, \dots, A_k) um valor real $\mathcal{I}(A_1, \dots, A_k)$ (Zitzler *et al.*, 2003).

Em geral, indicadores unários e binários são utilizados para quantificar a qualidade absoluta de um ou relativa entre dois conjuntos de aproximação, respectivamente. Usualmente, os três principais itens considerados na avaliação são (i) a proximidade do conjunto de aproximação à *PF*; (ii) o espaçamento relativo das soluções presentes na aproximação; e (iii) o grau de cobertura da aproximação em relação à *PF*.

O indicador comumente utilizado para medir o grau de proximidade à *PF* é a distância geracional (*Generational Distance, do inglês*), \mathcal{I}_{GD} , definido como a média da raiz das distâncias quadráticas mínimas entre as n soluções não-dominadas obtidas e as soluções na *PF* (Zitzler *et al.*, 2003) real. Seja $\mathbf{z} = \mathbf{f}(\mathbf{x} \in \mathcal{F}_1)$, em que \mathcal{F}_1 é o conjunto de soluções não-dominadas na população. Então,

$$\mathcal{I}_{GD}(\mathcal{F}_1) = \frac{1}{|\mathcal{F}_1|} \sqrt{\sum_{i=1}^{|\mathcal{F}_1|} \min_{\mathbf{y} \in PF} d(\mathbf{z}_i, \mathbf{y})^2}, \quad (3.6)$$

em que $\mathbf{z}_i \in F_M$ e $|\mathcal{F}_1|$ é o número de soluções não-dominadas obtidas. Note-se que $\mathcal{I}_{GD}(\mathcal{F}_1) = 0 \iff \mathcal{F}_1 \subseteq \Omega^*$ e, portanto, quanto menor a distância geracional, mais próximo à *PF* está o conjunto de aproximação avaliado. A desvantagem do indicador \mathcal{I}_{GD} é que é necessário que um conjunto finito de pontos na *PF* seja conhecido *a priori*. Nessa dissertação, utiliza-se a distância geracional para avaliação da proximidade à *PF*

no conjunto de MOPs contínuos ZDT (ver Apêndice B), uma vez que a forma analítica da PF nos problemas ZDT é conhecida (Zitzler *et al.*, 2002).

O indicador de espaçamento, \mathcal{I}_S , mede o quão uniformemente distribuída é a imagem do conjunto de aproximação obtido no espaço de objetivos (Zitzler *et al.*, 2003):

$$\begin{aligned}\mathcal{I}_S &= \sqrt{\frac{1}{|\mathcal{F}_1|} \sum_{i=1}^{|\mathcal{F}_1|} (d_i - \bar{d})^2}, \\ d_i &= \min_{\mathbf{x} \in \mathcal{F}_1; \mathbf{x} \neq \mathbf{x}_i} \sum_{l=1}^M |f_l(\mathbf{x}_i) - f_l(\mathbf{x})|, \\ \bar{d} &= \frac{1}{|\mathcal{F}_1|} \sum_{i=1}^{|\mathcal{F}_1|} d_i,\end{aligned}\tag{3.7}$$

em que f_l denota a l -ésima função-objetivo, M é o número de funções-objetivo, \mathbf{x}_i é a i -ésima solução não-dominada presente em \mathcal{F}_1 . Note-se que $\mathcal{I}_S(\mathcal{F}_1) = 0$ implica que as soluções estão espaçadas uniformemente no conjunto de não-dominadas. Portanto, valores menores de espaçamento são associados a melhores níveis de diversidade na aproximação obtida.

Para medir o quão bem a PF é coberta por um conjunto de aproximação, o indicador de espalhamento máximo, I_{MS} , pode ser utilizado (Zitzler *et al.*, 2003):

$$\mathcal{I}_{MS} = \sqrt{\frac{1}{M} \sum_{l=1}^M \left(\frac{\min(\overline{\mathcal{F}_1^l}, \overline{PF^l}) - \max(\underline{\mathcal{F}_1^l}, \underline{PF^l})}{\overline{PF^l} - \underline{PF^l}} \right)^2},\tag{3.8}$$

em que $\overline{\mathcal{F}_1^l}$, $\underline{\mathcal{F}_1^l}$, $\overline{PF^l}$ e $\underline{PF^l}$ correspondem aos valores máximo e mínimo do l -ésimo objetivo das soluções em \mathcal{F}_1 e na PF , respectivamente, e M é o número de funções-objetivo. Um valor alto de espalhamento indica uma melhor cobertura da PF .

3.2.2 Teoria do Hipervolume

Um indicador que vem despertando grande interesse é o hipervolume da união dos hipercubos formados pelas regiões de dominância das soluções do conjunto de aproximação, \mathcal{I}_H (Auger *et al.*, 2009). Seja $\Lambda : I \subset \mathbb{R}^M \mapsto \mathbb{R}^+$, em que $I = \prod_{i=1}^M [a_i, b_i]^M$, com $a_i \leq b_i$, uma medida de Lebesgue, i.e., $\Lambda(I) = \prod_{i=1}^M (b_i - a_i)$, então, dado um ponto de referência

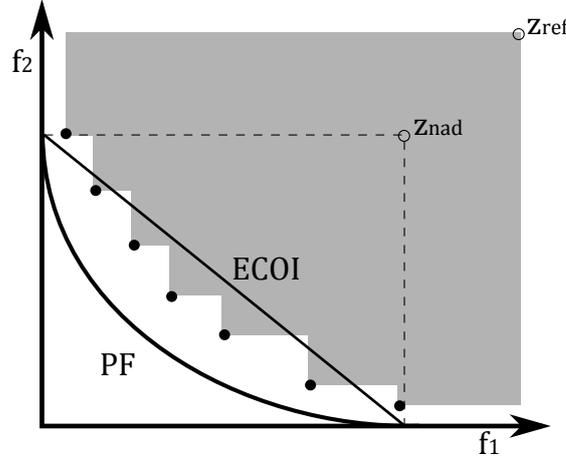


Figura 3.3 O indicador \mathcal{I}_H corresponde à área delimitada em cinza no gráfico.

$\mathbf{z}_{ref} \in F_M$ e um conjunto de não-dominadas \mathcal{F}_1 , o indicador de hipervolume é

$$\mathcal{I}_H(\mathcal{F}_1, \mathbf{z}_{ref}) = \Lambda \left(\bigcup_{\mathbf{x} \in \mathcal{F}_1} \{ \mathbf{z}' \in F_M : \mathbf{f}(\mathbf{x}) \prec \mathbf{z}' \prec \mathbf{z}_{ref} \} \right). \quad (3.9)$$

Esse indicador captura em um único valor real tanto a convergência quanto a distribuição das soluções não-dominadas. Ademais, \mathcal{I}_H é o único indicador conhecido que reflete a dominância de Pareto (Bader, 2010) no sentido de que, se um conjunto de aproximação domina fracamente outro, esse fato será refletido nos valores designados pelo hipervolume. Mais especificamente, $A \preceq B \wedge B \not\prec A \Rightarrow \mathcal{I}_H(A, \mathbf{z}_{ref}) > \mathcal{I}_H(B, \mathbf{z}_{ref})$. Em outras palavras, o hipervolume é estritamente monotônico com respeito à dominância de Pareto. Antes de apresentar o próximo resultado teórico relevante em relação ao hipervolume, define-se um ponto de Nadir:

Definição 3.12 (Ponto de Nadir). O vetor objetivo $\mathbf{z}^{nad} \in F_M$ é o ponto de Nadir se é formado pelos piores componentes dentre os vetores do conjunto $\{ \mathbf{z} : \mathbf{z} \in \text{ECOI} \cap PF \}$ (ver definição de ECOI no Apêndice A, página 145).

Demonstrou-se que, dado um Ω finito, se $\mathbf{z}_{ref} \prec \mathbf{z}^{nad}$, maximizar \mathcal{I}_H corresponde a encontrar o conjunto de Pareto Ω^* , quando $|\mathcal{F}_1| \rightarrow \infty$ (Fleischer, 2003). Também fora observado experimentalmente que a maximização do hipervolume para $|\mathcal{F}_1|$ finito produz subconjuntos da PF os quais apresentam altos níveis de diversidade. Portanto, o indicador \mathcal{I}_H fornece um arcabouço atraente para o projeto de algoritmos para MOPs baseados em seleção por indicadores. Um algoritmo evolucionário baseado em seleção

por hipervolume, cujo desempenho é o melhor conhecida para o conjunto de *benchmarks* ZDT (Apêndice B), é descrito em Beume *et al.* (2007).

Dada a crescente influência do indicador de Hipervolume na análise de resultados publicados na literatura de otimização multiobjetivo (Bader, 2010), a discussão dos resultados apresentados no Capítulo 6 concentra-se nesse indicador.

3.3 Otimização Dinâmica

Jin e Branke (2005) definem a *otimização dinâmica* como aquela em que um ou mais dos componentes a seguir mudam, de acordo com um índice temporal: (i) as funções-objetivos; (ii) a instância do problema¹; ou (iii) as restrições sobre as variáveis. Nessas condições, soluções ótimas encontradas podem não ser mais adequadas em estágios posteriores da otimização. Diversos problemas do mundo real possuem essa característica, como, por exemplo: o escalonamento de operações de trânsito (Fu *et al.*, 2003); o controle de veículos autônomos (Pereira, 2001); o roteamento de dados em redes móveis *ad hoc* (Bilgin e Khan, 2010); e a otimização de carteiras de investimento (Hassan, 2010).

A estratégia mais simples para lidar com mudanças temporais é descartar as soluções atuais e reinicializar o processo de otimização sempre que se detectar uma mudança em quaisquer dos componentes mencionados. No entanto, o custo desse procedimento pode ser proibitivo em muitas aplicações práticas. Por conseguinte, as principais abordagens para otimização dinâmica consideram uma *população de soluções* para aproveitar o máximo de informação do espaço de busca de forma a acelerar a velocidade com que novos ótimos são encontrados no novo estado temporal do problema (Jin e Branke, 2005). Nesse contexto, Algoritmos Evolucionários (AEs) constituem-se em uma técnica das mais apropriadas. Uma classe de AEs particularmente estudada nessa dissertação para a otimização dinâmica considera a incorporação de soluções atípicas por meio de *geradores de diversidade* ao longo do processo evolucionário² (ver Capítulo 4). A Seção 3.5.2 discute aspectos da otimização dinâmica multiobjetivo evolucionária.

¹Como exemplo, considere o problema do Caixeiro Viajante (Apêndice B): a lista de cidades a serem visitadas (instância do problema) pode mudar com o tempo, no entanto, as restrições de custo podem permanecer as mesmas, assim como o objetivo de minimizar a distância total percorrida.

²Jin e Branke (2005) consideram as abordagens que geram diversidade *ao longo* das gerações como “*manutenção de diversidade*”, em contraste com as abordagens as quais geram diversidade *após* mudanças no ambiente, denominadas “*geração de diversidade*”. O presente trabalho todavia considera a distinção entre ambas as categorias insuficiente, uma vez que se pode argumentar serem as estratégias de “*manutenção*” um caso especial das estratégias de “*geração*”. Ao longo da dissertação, portanto, adota-se o termo “*geração de diversidade*” para unificar as abordagens das duas categorias de Jin e Branke (2005). Ver Seção 4.4, página 64 para uma discussão mais aprofundada.

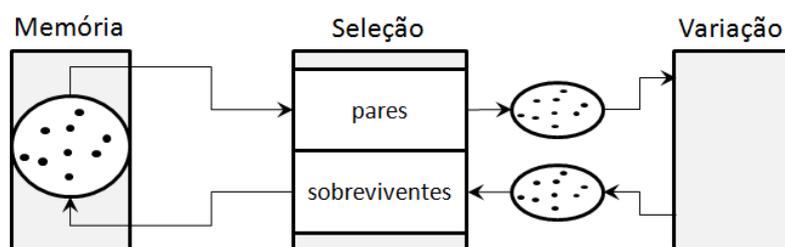


Figura 3.4 Módulos comuns em algoritmos evolucionários baseados em população (Adaptado de Zitzler *et al.* (2004)).

3.4 Fundamentos de Algoritmos Evolucionários

Métodos de busca estocástica (*metaheurísticas*), tais como *Simulated Annealing* (Kirkpatrick *et al.*, 1983), Otimização por Enxame de Partículas (PSO) (Kennedy e Eberhart, 1995), Estratégias Evolucionárias (EE) (Schwefel, 1965) e Algoritmos Genéticos (AG) (Goldberg, 1989), são métodos de otimização que fazem uso sistemático de decisões estocásticas para a amostragem de pontos em regiões promissoras do espaço de busca. Alguns desses métodos diferenciam-se dos demais por manterem um conjunto de soluções candidatas na memória — em vez de trabalhar com uma única solução —, denominando-se *metaheurísticas baseadas em população*, como é o caso do PSO, algumas variantes da EE e AGs. Nessa dissertação, são estudadas as metaheurísticas com inspiração biológica nos paradigmas da evolução Darwiniana e genética Mendeliana, metaheurísticas evolucionárias, as quais possuem componentes comuns entre si, conforme apontam Zitzler *et al.* (2004) (Figura 3.4):

1. Uma memória contendo as soluções candidatas consideradas (i.e., população);
2. Um módulo de seleção de soluções candidatas;
3. Um módulo de variação das soluções candidatas.

A memória armazena uma população finita de soluções candidatas, estando limitada aos recursos de armazenamento disponíveis no sistema computacional no qual seja implementada. Os elementos da população (indivíduos ou cromossomos) promissores, não apenas são selecionados para a etapa de variação — na qual novos indivíduos são gerados por meio de operadores especialmente projetados — mas também são avaliados por um critério de seleção de sobreviventes, o qual tem por finalidade manter nas próximas iterações do algoritmo, com maior probabilidade, aqueles indivíduos que

são mais promissores para a solução do problema. O operador de seleção de pares para reprodução utiliza, no caso geral, um critério de seleção baseado na qualidade da solução (aptidão), proporcionando, assim, a difusão das informações codificadas pelas melhores soluções encontradas. Por sua vez, os operadores de variação — recombinação (ou cruzamento) e mutação — desempenham o papel de geradores de diversidade, isto é, exploram novas regiões de pontos no espaço de busca por meio da criação de novos indivíduos, os quais herdam informações de outros indivíduos previamente selecionados, enquanto o operador de seleção de sobreviventes cria um nível de pressão seletiva que contribui para a melhoria na qualidade média da população ao longo das iterações do algoritmo (gerações).

Após um número suficiente de gerações, espera-se que o AE convirja para um subconjunto da região no espaço de soluções na qual poderá estar contida a solução ótima global, o que, principalmente em problemas multimodais, nem sempre ocorre. Embora os AEs reconhecidamente possuam a habilidade de escapar de ótimos locais (Goldberg, 1989), o sucesso em encontrar a solução global dependerá da relação *complexidade do problema vs. recursos computacionais disponíveis* e da natureza das restrições (se houver) sobre a(s) função(ões)-objetivo(s).

Descreve-se formalmente a seguir um Algoritmo Genético geracional mono-objetivo com base no modelo de Smith (1998), adaptado para representação com vetores reais, e apresentam-se: (i) os principais conceitos utilizados no projeto de um AE; e (ii) os operadores de recombinação, mutação e seleção utilizados nos experimentos dos Capítulos 4 e 6.

Definição 3.13 (Algoritmo Genético). Um Algoritmo Genético (AG) é a 8-tupla

$$\text{AG} = (\Theta^0, \Psi^0, N, \lambda, m, f, G, U), \quad (3.10)$$

em que $\Theta^0 \subset \mathcal{R}$ é um conjunto inicial de parâmetros; $\Psi^0 = (\mathbf{x}_1^0, \dots, \mathbf{x}_N^0) \in \mathcal{R}^{N \times m}$ é a população inicial; $N \in \mathcal{N}$ é o tamanho da população; $\lambda \in \mathcal{N}$ é o número de novas soluções (filhos); $m \in \mathcal{N}$ é a dimensão do espaço de soluções; $f: \mathcal{R}^m \mapsto \mathcal{R}^+$ é a função de aptidão; $G: \mathcal{R}^{N \times m} \times \Theta \mapsto \mathcal{R}^{\lambda \times m}$ é a função de variação; e $U: \mathcal{R}^{N \times m} \times \mathcal{R}^{\lambda \times m} \mapsto \mathcal{R}^{N \times m}$ é a função de seleção de sobreviventes.

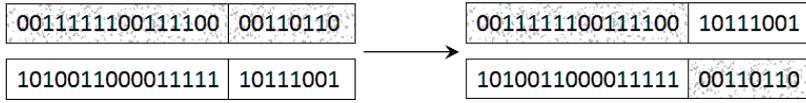


Figura 3.5 Exemplo de uma operação de cruzamento de 1 ponto em duas cadeias binárias.

Se $O \in \mathcal{R}^{\lambda \times m}$ é um conjunto de filhos, então uma geração t de um AG é descrita por:

$$\begin{aligned} O_i^g &= G(\Psi^t, \Theta^g) \\ \forall i &\in \{1, \dots, \lambda\} \\ \Psi^{g+1} &= U(\Psi^g \cup O^g), \end{aligned} \quad (3.11)$$

em que g denota a geração atual.

3.4.1 Operadores de Recombinação

A importância dos operadores de recombinação em AEs é reconhecida na obra de Goldberg (1989), que acreditava serem esses o principal mecanismo de funcionamento dos AGs com representação binária estudados à época. A principal função da recombinação em AGs é a de gerar novas soluções a partir das características herdadas de soluções pais. Em geral, os operadores são construídos de forma probabilística, de modo que n operações independentes realizadas com os mesmos indivíduos de entrada produzem n soluções potencialmente distintas. Embora operadores de recombinação de aridade maior que dois sejam possíveis, na prática, operadores binários (com respeito ao número de indivíduos de entrada) são utilizados:

Definição 3.14 (Operador de Recombinação). Um operador de recombinação é um mapeamento $R : \mathcal{R}^m \times \mathcal{R}^m \times \Theta \mapsto \mathcal{R}^m$.

Cabe ressaltar que a Definição 3.14 considera o caso de soluções candidatas codificadas com números reais. As primeiras aplicações de AGs para otimização contínua, no entanto, faziam uso de codificações binárias, em que o *espaço genótipo* era constituído de populações de cadeias de bits as quais eram mapeadas ao longo da execução para elementos do *espaço fenótipo*, que correspondem aos valores mapeados pela função-objetivo a partir das variáveis de decisão. Por exemplo, um dos primeiros operadores utilizados, a *recombinação de n-pontos*, é inspirado nas operações de *crossing over* biológicas, em que dois cromossomos se alinham em um ponto (quiasma) aleatório – *crossing over* simples – ou em dois pontos aleatórios – *crossing over* duplo – se quebrando em duas ou quatro

partes, respectivamente, e se recompondo no cromossomo homólogo. A Figura 3.4.1 mostra um exemplo para um cruzamento em um ponto aleatório, em que um inteiro aleatório q é escolhido no intervalo $[0, l - 1]$, de forma a separar as soluções pais nessa posição e produzir duas novas soluções trocando os últimos $l - q$ bits.

As dificuldades decorrentes da representação binária desencorajaram as primeiras aplicações de AGs na otimização contínua. Primeiro, um fenômeno conhecido como *penhascos de Hamming* requer a alteração de vários bits para certas transições entre soluções vizinhas em espaços de busca reais, o que prejudica a convergência dos AGs. Por exemplo, a transição entre os fenótipos 127 e 128 requer a alteração de 7 bits no espaço genótipo (01111111 \rightarrow 10000000). Outra dificuldade é a necessidade da escolha *a priori* da precisão numérica, a qual depende do tamanho dos cromossomos. Aumentos de precisão levam à adição de mais bits na representação, o que requer populações de tamanhos maiores, aumentando o custo computacional do processo de otimização. Essas e outras dificuldades podem ser contornadas com a codificação em ponto flutuante, o que exige a definição de operadores de recombinação que atuem sobre a representação real. No entanto, os primeiros operadores sugeridos, como a recombinação aritmética, não demonstraram o mesmo *poder de busca* (Definição 3.15) dos operadores sobre representação binária.

Definição 3.15 (Poder de Busca). O poder de busca de um operador de recombinação pode ser definido como a probabilidade da amostragem de um ponto arbitrário no espaço de busca a partir de duas soluções pais (Deb e Agrawal, 1995).

Deb e Agrawal (1995) propuseram um operador de recombinação para representação real equivalente ao cruzamento de 1-ponto no caso binário, em relação ao poder de busca daquele operador. Por simplicidade, considera-se o caso unidimensional: sejam $p_1, p_2 \in \mathcal{R}$ os genes dos dois pais envolvidos na recombinação e $c_1, c_2 \in \mathcal{R}$ os genes dos filhos gerados pela aplicação do operador $R(p_1, p_2, \Theta)$. Seja ainda $\beta = \frac{|c_1 - c_2|}{|p_1 - p_2|}$ a razão de espalhamento. O Cruzamento Binário Simulado (SBX, *Simulated Binary Crossover* do inglês) é então definido em termos da distribuição de probabilidade de β , aproximada por

$$p(\beta) = \begin{cases} \frac{1}{2}(\eta + 1)\beta^\eta, & \text{se } \beta \leq 1 \\ \frac{1}{2}(\eta + 1)\frac{1}{\beta^{\eta+2}}, & \text{c.c.} \end{cases}, \quad (3.12)$$

em que $\eta \in \mathcal{N}$. O procedimento completo para gerar filhos a partir do SBX é:

1. Amostrar um valor de uma distribuição uniforme no intervalo $[0, 1]$, $u \sim U(0, 1)$;

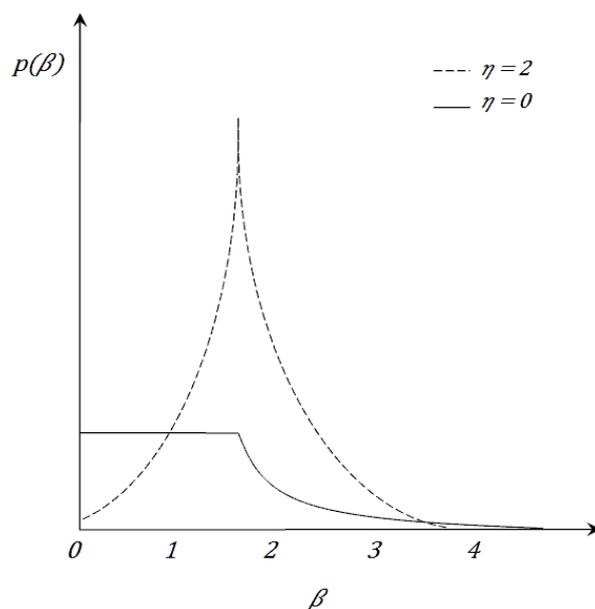


Figura 3.6 Função densidade de probabilidade da razão de espalhamento β no operador SBX.

2. Amostrar um valor β' a partir da distribuição da equação (3.12):

$$\text{Selecionar } \beta' \text{ tal que } \int_0^{\beta'} p(\beta) d\beta = u;$$

3. Gerar os dois filhos com espalhamento β' :

$$\begin{aligned} c1 &= \frac{1}{2} [(p_1 + p_2) - \beta' |p_2 - p_1|], \\ c2 &= \frac{1}{2} [(p_1 + p_2) + \beta' |p_2 - p_1|]. \end{aligned} \quad (3.13)$$

Resultados experimentais mostram que $2 \leq \eta \leq 5$ aproxima a distribuição da razão de espalhamento em cruzamentos de 1-ponto (Figura 3.4.1). Na prática, valores pequenos de η fazem com que o operador tenha maiores chances de gerar filhos distantes dos pais no espaço genótipo, enquanto valores grandes aumentam o grau de estacionariedade do SBX (Deb e Agrawal, 1995). Em espaços de busca multidimensionais, basta aplicar o procedimento anterior independentemente para cada gene. Note-se que, pela Definição 3.14, $\Theta = \{\eta\}$ para o operador SBX.

3.4.2 Operadores de Mutação

Na biologia evolutiva, mutações são erros introduzidos em uma sequência genética decorrentes de falhas nos processos de cópia e replicação cromossômica (Ridley, 2003). A computação evolucionária simula esse processo com o objetivo de melhor explorar o espaço de busca e aumentar a capacidade dos AEs de escapar de ótimos locais por meio do aumento da diversidade genética. Deb e Goyal (1996) propuseram um operador de mutação para AGs com codificação real que, assim como o operador SBX, utiliza uma distribuição de probabilidade polinomial para gerar indivíduos no entorno do vetor de entrada. A distribuição é calculada sobre um fator de perturbação $\delta = (c - p)/\Delta_{\max}$:

$$P(\delta) = \frac{1}{2}(\eta + 1)(1 - |\delta|)^\eta, \quad \delta \in (-1, 1), \quad (3.14)$$

em que c , p e Δ_{\max} representam respectivamente o gene da solução mutante, o gene da solução original e a perturbação máxima permitida para o gene p . O procedimento para gerar mutantes é:

1. Amostrar um valor de uma distribuição uniforme no intervalo $[0, 1]$, $u \sim \text{Uniforme}(0, 1)$;
2. Calcular o fator de perturbação $\bar{\delta}$ a partir da seguinte equação:

$$\bar{\delta} = \begin{cases} (2u)^{\frac{1}{\eta+1}} - 1, & \text{se } u < \frac{1}{2} \\ 1 - [2(1-u)]^{\frac{1}{\eta+1}}, & \text{se } u \geq \frac{1}{2} \end{cases}. \quad (3.15)$$

3. Gerar o gene mutante com perturbação $\bar{\delta}\Delta_{\max}$:

$$c = p + \bar{\delta}\Delta_{\max}. \quad (3.16)$$

Os operadores SBX e de mutação polinomial apresentados são utilizados na proposta original do AE multiobjetivo NSGA2 (Seção 3.5.1), bem como nos experimentos relatados nos Capítulos 4 e 6.

3.4.3 Operadores de Seleção

A principal função dos operadores de seleção em AEs é direcionar a busca em torno das regiões mais promissoras visitadas no espaço de soluções (Valdez-Peña *et al.*, 2009). Tanto a seleção para reprodução quanto a seleção de sobreviventes utilizam mecanismos

de decisão baseados na aptidão relativa dos indivíduos na população. A abordagem mais difundida, a *seleção proporcional à aptidão*, seleciona pais (ou sobreviventes) a partir de um processo estocástico baseado em uma distribuição de probabilidade discreta: à cada indivíduo é associado unicamente um número inteiro de 0 a N e uma probabilidade de seleção dada pela razão entre a aptidão desse em relação à aptidão total da população; os pais ou sobreviventes são selecionados por meio da amostragem de índices dessa distribuição. Essa estratégia apresenta problemas quanto à escala relativa de valores de aptidão. Por exemplo, se o desvio padrão da distribuição de aptidões for muito pequeno, cada indivíduo recebe aproximadamente a mesma probabilidade de ser selecionado, reduzindo o grau de competição entre eles (efeito conhecido como perda de *pressão de seleção*). Dessa forma, as aplicações mais recentes em computação evolucionária consideram outros tipos de operadores, invariantes quanto ao desvio padrão da distribuição de aptidões, com parâmetros ajustáveis capazes de controlar os níveis de pressão de seleção. A seguir, duas alternativas à seleção proporcional à aptidão são apresentadas.

Na *seleção por torneio*, k indivíduos são escolhidos ao acaso para competir entre si pela reprodução/sobrevivência. O vencedor dessa competição (torneio) é considerado aquele de maior aptidão dentre os competidores. Essa abordagem é interessante, pois permite o ajuste da pressão de seleção em função do tamanho do torneio: quanto maior o valor de k , maior as chances dos melhores indivíduos serem selecionados para a competição e de saírem vencedores, e vice-versa.

Finalmente, a *seleção por truncamento* é o método mais simples possível para a etapa de sobreviventes: consiste em selecionar os $\omega(g) \times N$ indivíduos de aptidão mais alta na população da geração g , em que $0 \leq \omega(g) \leq 1$ é frequentemente escolhida como uma função constante no tempo.

As seleções por torneio e por truncamento são comumente utilizadas no NSGA2 (Seção 3.5.1) nas etapas de reprodução e de sobrevivência, respectivamente.

3.5 Algoritmos Evolucionários Multiobjetivo (MOEAs)

Na aplicação de técnicas evolucionárias à otimização multiobjetivo, a distinção entre ótimos globais e locais perde clareza, uma vez que se deseja obter um conjunto de soluções, em vez de uma única solução definitiva. Matematicamente, uma única solução global para um MOP não existe, a menos que o ponto de utopia (ver Definição 3.16) pertença ao espaço de soluções factíveis:

Definição 3.16 (Ponto de utopia). Um ponto \mathbf{f}^o no espaço de objetivos é um *ponto de utopia* $\iff \mathbf{f}_i^o = \min_{\mathbf{x} \in F_M} f_i(\mathbf{x})$ com respeito a \mathbf{x} , $\forall i$.

Portanto, a abordagem de busca evolucionária na otimização multiobjetivo considera a determinação de soluções que sejam, globalmente, pontos ótimos de Pareto. Nesse sentido, a busca evolucionária considera a melhor aproximação possível para a Frente de Pareto (PF), determinada pelo conjunto de soluções não-dominadas, \mathcal{F}_1 . Nesse cenário, é importante ressaltar a existência de duas possibilidades para a aplicação da computação evolucionária em MOPs (Marler e Arora, 2004):

1. Na otimização de um único objetivo por vez nos métodos clássicos para MOPs (Apêndice A);
2. A otimização simultânea das funções-objetivo através de operadores especialmente projetados para geração e preservação de soluções não-dominadas.

Ao longo do restante da dissertação, apenas a segunda possibilidade é considerada, uma vez que as abordagens evolucionárias nessa categoria constituem o estado da arte na otimização multiobjetivo. Os AEs os quais empregam operadores especiais para a otimização multiobjetivo denominam-se na literatura Algoritmos Evolucionários Multi-objetivos (*Multiobjective Evolutionary Algorithms*, MOEAs, do inglês). Para entender a extensão com que os MOEAs têm sido utilizados em trabalhos envolvendo a solução de MOPs, Jones *et al.* (2002) informam que cerca de 90% desses utilizaram metaheurísticas como técnicas de solução, enquanto que 70% escolheram MOEAs como ferramentas de otimização. Nesse cenário, Konak *et al.* (2006) apontam três metas conflitantes as quais devem ser perseguidas por qualquer algoritmo que se proponha a aproximar o conjunto ótimo de Pareto:

1. Qualidade da aproximação: \mathcal{F}_1 deve estar o mais próximo possível da PF real. Na situação ideal, \mathcal{F}_1 deve ser um subconjunto da PF ;
2. Espalhamento: as soluções contidas em \mathcal{F}_1 devem ser diversificadas e uniformemente distribuídas sobre a PF , de forma a proporcionar ao tomador de decisão a melhor visualização possível das relações de custo-benefício existentes entre as funções-objetivo;
3. Cobertura: \mathcal{F}_1 deve capturar todo o espectro da PF . Para tanto, as regiões próximas aos pontos extremos (ver Definição 3.7) na PF devem ser exploradas.

Apresenta-se, a seguir, um dos principais MOEAs proposto na literatura, o qual é utilizado ao longo dos experimentos relatados nos próximos Capítulos da dissertação.

3.5.1 AG com Ordenação por Não-Dominação (NSGA2)

O *Fast Nondominated Sorting Genetic Algorithm* (NSGA2) propõe uma aproximação da *PF* com complexidade computacional reduzida, quando comparado a outros métodos que usam *rank* de soluções não-dominadas e técnicas de especiação (ver Apêndice C). Para tanto, o NSGA2 (Deb *et al.*, 2000) utiliza um método de ordenação de soluções não-dominadas de complexidade $\mathcal{O}(MN^2)$ – em que M é o número de objetivos e N é o tamanho da população – por meio de um novo operador de seleção e técnicas de elitismo, i.e., o arquivamento de soluções na *PF* aproximada. Ainda, diferentemente de MOEAs como o MOGA (Apêndice C), o NSGA2 elimina a necessidade da determinação de parâmetros de compartilhamento de aptidão (e.g. σ_{share}), pois apresenta um procedimento eficiente para estimar a densidade de soluções no entorno de um indivíduo.

O módulo de ordenação das soluções não-dominadas computa, a cada indivíduo \mathbf{x}_j , as quantidades $nd_j = nd(\mathbf{x}_j, g)$ (número de soluções que dominam \mathbf{x}_j na geração g) e S_j , o conjunto de soluções dominadas por \mathbf{x}_j . Cada indivíduo tal que $nd_j = 0$ é armazenado em uma lista \mathcal{F}_1 , chamada *frente atual*. Então, para cada indivíduo $\mathbf{x}_k \in S_j$, aqueles cujo $nd_k = 1$ são movidos para uma segunda lista \mathcal{H} . Nesse ponto, \mathcal{F}_1 é chamada de Frente 1 – ou *conjunto de não-dominadas* –, \mathcal{H} passa a ser a *frente atual* e o processo é repetido para construir $\mathcal{F}_2, \dots, \mathcal{F}_n$, supondo que $n \leq N$ frentes existam na população, em que N é o número de indivíduos na população. Esse processo possui complexidade $\mathcal{O}(mN^2)$ e retorna uma lista $\mathcal{F} = \{\mathcal{F}_1, \dots, \mathcal{F}_n\}$ de todas as frentes identificadas.

O processo de estimação de densidade de indivíduos no entorno de uma solução \mathbf{x}_j no NSGA2 é descrito a seguir. Primeiro, a metade do perímetro (normalizado) do cubóide (Figura 3.5.1), cujos dois dos vértices são as duas soluções vizinhas mais próximas que cercam \mathbf{x}_j no espaço de objetivos, é computada. Denota-se essa quantidade, cd_j , por *crowding distance*.

O NSGA2 define então um operador de comparação (denotado nessa dissertação por \succeq_{ND} , ND para “não-dominância”), que é uma ordem parcial sobre o espaço de soluções, para guiar o processo de seleção de modo que a população convirja para uma distribuição uniforme na direção da *PF*. A ordem parcial é definida da seguinte forma:

$$\mathbf{x}_j \succeq_{\text{ND}} \mathbf{x}_k \iff r_j < r_k \wedge (r_j = r_k \vee cd_j > cd_k), \quad (3.17)$$

em que $r_j = r(\mathbf{x}_j)$ é o rank de Pareto de \mathbf{x}_j , definido no NSGA2 como $r_j = i$, se $\mathbf{x} \in \mathcal{F}_i$. Portanto, vê-se que os valores de *crowding distance* computados servem como critério de desempate ao determinar qual solução precede outra, no sentido da dominância de Pareto

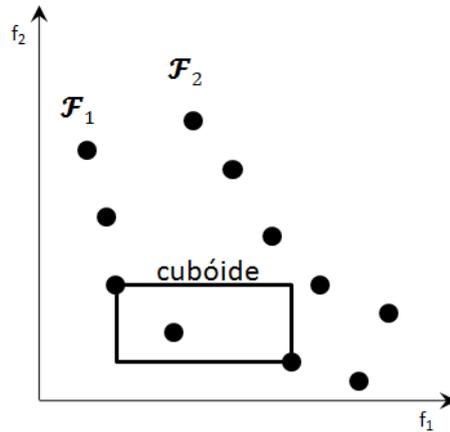


Figura 3.7 Representação do maior cubóide que contém uma dada solução no espaço F_m , conforme calculado no NSGA2 (adaptado de Deb *et al.* (2000)).

baseada em *rank*, mantendo a diversidade necessária na etapa de seleção de sobreviventes.

O processo de otimização proposto pelo NSGA2 completa-se com a introdução de um mecanismo de seleção elitista baseado nas relações de não-dominância entre as várias frentes construídas na geração atual. O algoritmo completo, contando com a seleção de soluções que sobreviverão à próxima geração, é descrito no Pseudocódigo 1.

3.5.2 MOEAs em Ambientes Dinâmicos

Um MOP dinâmico é assim definido (Farina *et al.*, 2004):

$$\begin{cases} \min_{\mathbf{x} \in \Omega} \mathbf{f} = [f_1(\mathbf{x}, t) \cdots f_m(\mathbf{x}, t)]^T, \\ \text{sujeito a } \mathbf{g}(\mathbf{x}, t) \leq 0, \mathbf{h}(\mathbf{x}, t) = 0 \end{cases} \quad (3.18)$$

Naturalmente, na otimização dinâmica multiobjetivo, as definições de otimalidade (ver Definição 3.10) devem incluir o índice temporal do *ambiente dinâmico*:

Definição 3.17 (Ótimo dinâmico de Pareto). Um ponto $\mathbf{x}^* \in \Omega$ é um ótimo dinâmico de Pareto $\iff \nexists \mathbf{x} \in \Omega \mid \mathbf{f}(\mathbf{x}, t) \prec \mathbf{f}(\mathbf{x}^*, t)$, em que t é o índice temporal das funções-objetivo.

De forma semelhante, a Frente de Pareto e o conjunto de Pareto também incluem o índice temporal: $PF(t)$ e $\Omega^*(t)$, respectivamente.

No que concerne ao projeto de MOEAs para ambientes dinâmicos, poucos estudos existem na literatura. Em um dos principais trabalhos, Farina *et al.* (2004) discutem os desafios trazidos pelas mudanças temporais para a aproximação da PF , além de proporem

Pseudocódigo 1 Fast Nondominated Sorting Genetic Algorithm (NSGA2)

```

InicializarPopulação( $\Psi^{(0)}$ ),  $g \leftarrow 0$ 
Ordenar( $\Psi^{(0)}, \geq_{ND}$ ) // Ordena  $\Psi^{(0)}$  pelo critério de não-dominância  $\geq_{ND}$ 
enquanto condição de término não satisfeita faça
     $Q^{(g)} \leftarrow$  OperadoresGenéticos( $\Psi^{(g)}$ ) // Seleção por torneio, cruzamento e mutação
     $R^{(g)} \leftarrow \Psi^{(g)} \cup Q^{(g)}$  //  $Q^{(g)}$  é a subpopulação de filhos e  $\Psi^{(g)}$  é a população atual.
     $\mathcal{F} \leftarrow$  ConstruirFrentes( $R^{(g)}$ )
     $\Psi^{(g+1)} \leftarrow \emptyset$ ,  $i \leftarrow 0$ 
    enquanto  $|\Psi^{(g+1)}| < N$  faça
        AtribuirCrowdingDistance( $\mathcal{F}_i$ )
         $\Psi^{(g+1)} \leftarrow \Psi^{(g+1)} \cup \mathcal{F}_i$ 
         $i \leftarrow i + 1$ 
    fim enquanto
    Ordenar( $\Psi^{(g+1)}, \geq_{ND}$ )
     $\Psi^{(g+1)} \leftarrow \Psi^{(g+1)}[0 : N]$  // Os primeiros  $N$  elementos de  $\Psi^{(g+1)}$  sobrevivem
     $g \leftarrow g + 1$ 
fim enquanto
retorne  $\Psi^{(g)}$ 

```

um conjunto de teste para a avaliação de MOEAs dinâmicos. Os testes buscam refletir as quatro combinações de mudanças que podem ocorrer nos espaços de busca e de objetivos:

1. O conjunto de Pareto Ω^* no espaço de decisão muda, enquanto a PF no espaço de objetivos se mantém;
2. Ambos, Ω^* e PF , mudam;
3. A PF muda e Ω^* se mantém inalterado;
4. Nem Ω^* nem PF mudam, embora a instância do problema possa mudar.

Dessa forma, soluções antes não-dominadas em um estado temporal podem não mais apresentar essa propriedade em estados subsequentes. O projeto de novos MOEAs robustos nos tipos de ambientes descritos, portanto, deve ser incentivado, especialmente por meio do uso de técnicas de geração de diversidade. O conjunto de cinco problemas propostos pelos autores, denominado FDA, é apresentado no Apêndice B. No Capítulo 4, página 70 discute-se um dos poucos MOEAs especialmente projetados para lidar com ambientes dinâmicos.

3.6 Discussão Final e Síntese

O presente Capítulo apresentou: (i) o formalismo matemático dos problemas de otimização dinâmica multiobjetivo, para os quais essa dissertação estuda novos métodos de solução; e (ii) os fundamentos dos Algoritmos Evolucionários Multiobjetivos (MOEAs) avaliados nos Capítulos 4 e 6.

Conforme discutido na Seção 3.5.2, a mudança das funções-objetivo ao longo do tempo constitui-se em um desafio para as estratégias de preservação de diversidade em MOEAs, as quais devem implementar ações para aumentar as chances de adaptação da população de soluções candidatas aos novos ambientes. Desse modo, a busca por mecanismos de geração de diversidade genética que atendam à esse requisito de forma robusta, i.e., com a maior correlação possível entre a diversidade gerada e os novos ambientes, deve ser promovida. Esse é o principal problema que a pesquisa descrita nessa dissertação se propõe a tratar.

O Algoritmo Genético com Ordenação por Não-Dominação, NSGA2 (ver Seção 3.5.1), é utilizado como o modelo base sob o qual novas técnicas de geração de diversidade são incorporadas. Como o foco da dissertação está nos mecanismos de geração de diversidade genética os quais, acredita-se nesse trabalho, devem ser estudados de forma separada dos mecanismos de convergência à *PF* utilizados pelos MOEAs, optou-se por incorporar as técnicas de geração de diversidade no MOEA mais referenciado na literatura, a fim de facilitar futuras análises e comparações com outras propostas. Resulta daí a escolha do NSGA2. Ressalta-se, entretanto, que as conclusões formuladas a partir dos estudos experimentais relatados no Capítulo 6 independem dessa escolha e que os métodos e propostas desenvolvidos nos próximos Capítulos são aplicáveis para qualquer MOEA escolhido.

A importância da preservação de diversidade da população e as técnicas encontradas na literatura para gerar diversidade no contexto da otimização dinâmica são discutidas no próximo Capítulo.

4

A Influência da Diversidade Genética

*In a way, this diversity is very exciting
but one has at some point to ask
are these real beginnings
or so many false starts?*

– **Juliet Mitchell**

Segundo a União Internacional para a Conservação da Natureza e dos Recursos Naturais (IUCN, do inglês), a diversidade genética é uma das três formas de biodiversidade que devem ser preservadas. A perda de diversidade genética em uma população biológica está associada à pressão de seleção – a qual naturalmente induz a reduções benéficas na frequência alélica de um determinado gene – e à deriva genética, que corresponde a flutuações na frequência alélica devido à aleatoriedade dos processos de reprodução e transmissão de características entre gerações subseqüentes.

Reed e Frankham (2003) apontam os dois efeitos mais evidentes resultantes da perda de diversidade genética em populações biológicas: (i) a redução da capacidade de resposta evolucionária de uma população às mudanças no ambiente; e (ii) a redução na aptidão média da população devido ao aumento da taxa de endocruzamento e conseqüente aumento na incidência de alelos recessivos deletérios, que bloqueiam ou reduzem passos metabólicos importantes na fisiologia do organismo devido à indução de homozigose. Portanto, não é surpresa que correlações estatísticas entre diversidade genética e aptidão (medida em termos de taxas de fecundidade ou expectativas de vida) tenham sido demonstradas em populações biológicas, como em Reed e Frankham (2003).

O objetivo deste capítulo é demonstrar a influência da diversidade genética na dinâmica de convergência de Algoritmos Evolucionários Multiobjetivo (MOEAs) à frente real de Pareto (*PF*). Para tanto, discutem-se caracterizações e propriedades da diversidade

genética em termos de métricas de dispersão populacional. A diversidade genética de um modelo de duas subpopulações é obtida e esse resultado é utilizado para ilustrar as limitações práticas das métricas de diversidade existentes. Esse modelo também é descrito com o propósito de caracterizar a operação das estratégias de *inserção de imigrantes*, utilizadas nessa dissertação para geração de diversidade em MOEAs. Dessa forma, uma *subpopulação de trabalho* é descrita como aquela obtida a partir dos operadores de variação e seleção, enquanto que a *subpopulação de imigrantes* é gerada independentemente. Ambas as subpopulações são descritas em termos de variáveis aleatórias e a diversidade genética esperada do modelo é obtida em termos da taxa de imigração, i.e., a proporção de imigrantes na população conjunta. Finalmente, uma revisão de literatura da inserção de imigrantes é apresentada e contextualizada por meio de uma taxonomia sugerida a partir do modelo de duas subpopulações proposto.

4.1 O Balanço Entre Exploração e Exploração

O balanço entre Exploração e Exploração (E_r/E_t) é um tema central em diversas áreas da inteligência computacional, como aprendizagem de máquina, modelagem e previsão, além de busca e otimização (Chen *et al.*, 2009).

Uma definição formal para os conceitos de E_r/E_t afirma que um processo estocástico é de exploração se, e somente se, o ponto de amostragem é gerado independentemente da informação adquirida no histórico de pontos amostrais; enquanto que a exploração utiliza explicitamente tal informação (Chen *et al.*, 2009):

Definição 4.1 (Exploração). Um processo de amostragem X_t no tempo t é de exploração se, e somente se, $P(X_t = X | X_{t-1}, \dots, X_0) = P(X_t = X)$.

Definição 4.2 (Exploração). Um processo de amostragem X_t no tempo t é de exploração se, e somente se, $P(X_t = X | X_{t-1}, \dots, X_0) \neq P(X_t = X)$.

As Definições 4.1 e 4.2 restringem o escopo de operações de E_r/E_t à influência do histórico do processo em termos da presença ou ausência de dependências estatísticas temporais.

Na computação evolucionária, a diversidade genética populacional encontra-se no cerne das discussões acerca de E_r/E_t . Embora não haja consenso e rigor em uma definição de E_r/E_t , para o presente trabalho, é suficiente considerar que operações as quais elevam os níveis de diversidade – aumentando a cobertura do espaço de busca – são associadas à exploração, enquanto que aquelas que reduzam os níveis de diversidade – restringindo

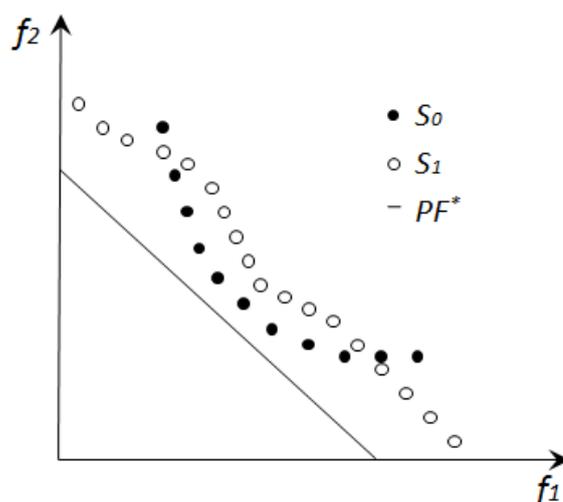


Figura 4.1 Duas aproximações da PF com diferentes níveis de diversidade e proximidade (Adaptado de Bosman e Thierens (2003)).

o escopo da busca – associam-se à exploração. Ademais, há uma linha de pensamento – e.g. Eshelman *et al.* (1989) – que associa as operações de exploração aos mecanismos de seleção de pais e de sobreviventes. Por outro lado, consideram-se as operações comumente associadas à exploração aquelas capazes de amostrar novos pontos do espaço de soluções, i.e., os operadores de variação (Eiben e Schippers, 1998). O contraste desse pensamento com a Definição 4.1 é evidente.

4.1.1 Exploração e Exploração em MOEAs

Uma diferença das estratégias de balanço de E_r/E_t em MOEAs em relação às aquelas de AEs de um único objetivo reside na preservação de um nível mínimo de diversidade ao longo de todo o processo de busca. Algoritmos evolucionários de um objetivo, em contraste, admitem perda de diversidade nos estágios finais do processo de busca de forma a acelerar a convergência. Apesar dessa diferença, o principal dilema no balanço entre E_r/E_t tanto em AEs quanto em MOEAs reside no fato de que uma estratégia de exploração excessiva contribui para a redução da velocidade de convergência, enquanto que uma estratégia de exploração excessiva pode levar à convergência prematura.

Além disso, Bosman e Thierens (2003) argumentam que mensurar o desempenho de MOEAs sem uma preferência por diversidade ou proximidade à PF leva a situações de incomparabilidade, como ilustra a Figura 4.1, em que a aproximação S_1 – a qual apresenta boa diversidade e cobertura, além de um bom nível de proximidade – não é preferível em

relação à aproximação S_0 – que apresenta diversidade e cobertura menores em relação às apresentadas por S_1 , mas com um melhor nível de proximidade – e vice-versa.

Um fator complicador reside no argumento de que manter e/ou elevar os níveis de diversidade no espaço de objetivos não deve ser mais difícil do que fazê-lo em sincronismo com uma caminhada suave em direção à PF . Nesse contexto, um MOEA é considerado eficiente se for capaz de encontrar um balanço adequado de E_r/E_t ao longo de todo o processo evolucionário de modo a aproximar a PF à medida que mantém bons níveis de diversidade e espalhamento. Cabe ressaltar, no entanto, que o balanço ideal de E_r/E_t depende das características da paisagem de aptidão e, logo, varia em relação ao problema de otimização considerado.

4.1.2 Métricas de Diversidade Genética

Uma variedade de algoritmos evolucionários computam medidas de diversidade ao longo da otimização para diversos fins, como a detecção de convergência para guiar critérios de término e reinicialização, ou o controle de parâmetros. Em estratégias de nicho (ver Apêndice C, página 162), a diversidade é utilizada como base para atribuição de aptidão. Em paisagens de aptidão dinâmicas, em que as funções-objetivo variam com o tempo, a diversidade pode ser medida com o objetivo de preservá-la por meio de ações que alteram a taxa de mutação para obter uma distribuição de diversidade alvo (Gouvêa Jr. e Araújo, 2010).

Em populações de soluções com representação por cadeias binárias, uma das métricas de diversidade mais comum é a Distância de Hamming dois a dois (Horn, 1997), definida para N indivíduos \mathbf{x} de tamanho l , a qual mensura para cada gene o número total de divergência entre alelos na população:

$$H = \sum_{j=1}^{N-1} \sum_{j'=j+1}^N \left(\sum_{i=1}^l |x_{ij} - x_{ij'}| \right), \quad (4.1)$$

em que $\mathbf{x} \in \{0, 1\}^l$ e $x_{ij} \in \{0, 1\}$. Uma desvantagem das métricas baseadas em computações dois a dois é a alta complexidade computacional, de ordem quadrática em relação ao tamanho da população $O(N^2)$.

Morrison e De Jong (2002) propuseram uma medida de diversidade genética de complexidade linear em relação à dimensionalidade do espaço de decisão (m) e ao tamanho da população (N) baseada no conceito físico de momento de inércia (I_M), o qual indica o grau de facilidade com que um objeto gira em torno de um eixo de rotação.

Quando o ponto de rotação corresponde ao centróide do objeto, o momento de inércia pode ser interpretado como o grau de afastamento do centro de gravidade sob o qual a massa do objeto está distribuída. O momento de inércia em um ponto de massa é dado como mr^2 , em que m é a massa do objeto e r^2 é o quadrado da distância ao ponto de rotação. A partir desse conceito, sendo $\mathbf{c} = [c_1 \cdots c_m]^T$ o centróide da população, em que $c_i = \sum_{j=1}^N x_{ij}/N$ e $x_{ij} \in \mathcal{R}$, a medida de diversidade baseada no momento de inércia é dada por:

$$I_M = \sum_{i=1}^m \sum_{j=1}^N (x_{ij} - c_i)^2, \quad (4.2)$$

A diversidade genética medida por I_M é de grande utilidade na otimização em espaços contínuos, pois pode ser mostrado que a distância de Hamming dois a dois (equação (4.1)) é a versão binária de I_M (Morrison e De Jong, 2002).

4.2 A Diversidade Genética de Duas Subpopulações

As métricas de diversidade genética utilizadas em AEs fornecem informação sobre o acúmulo e/ou média da dispersão dos indivíduos sobre o espaço de busca, mas, embora sejam medidas de localização importantes, elas não são capazes de caracterizar completamente a dispersão genética de uma população. Em particular, ao se considerar o momento de inércia na equação (4.2) como uma medida de dispersão populacional, pode-se demonstrar a existência de diferentes populações de soluções candidatas distribuídas de tal forma que a diversidade genética esperada, descrita em termos das diferenças quadráticas para os centros de massa das populações, não são distinguíveis entre si, o que pode levar a decisões desinformadas no controle da diversidade. Esse fato é demonstrado no modelo de população proposto a seguir.

4.2.1 O Modelo e Propriedades Estatísticas

Sejam Y e Z duas variáveis aleatórias (v.a.) contínuas. Um modelo de população de interesse para o presente trabalho é este:

$$X = \begin{cases} Y, & \text{com probabilidade } \alpha \\ Z, & \text{com probabilidade } 1 - \alpha. \end{cases} \quad (4.3)$$

Mais especificamente, $P\{X = Y\} = \alpha$ e $P\{X = Z\} = 1 - \alpha$, sendo Y, Z interpretadas como duas subpopulações locais, não necessariamente isoladas (disjuntas), e continu-

amente distribuídas com $0 \leq \alpha \leq 1$. Denote esse modelo por $X(\alpha)$. Sem perda de generalidade, se X representar a distribuição das soluções candidatas em um estado específico do processo evolucionário na reta real, então, ao considerarmos o centróide $C_{X(\alpha)}$ da população para a estimação da diversidade genética, têm-se

$$C_{X(\alpha)} = E[X(\alpha)] = \sum_{i=1}^2 x_i P\{X(\alpha) = x_i\} = \alpha Y + (1 - \alpha)Z. \quad (4.4)$$

Vê-se, então, que $C_{X(\alpha)}$ é uma combinação linear das v.a. Y e Z e, portanto, é, por si mesma, uma variável aleatória. Calculando-se o valor esperado e a variância de $C_{X(\alpha)}$, têm-se

$$\begin{aligned} E[C_{X(\alpha)}] &= E[\alpha Y + (1 - \alpha)Z] \\ &= E[\alpha Y] + E[(1 - \alpha)Z] \quad (\text{linearidade do operador esperança}) \\ &= \alpha E[Y] + (1 - \alpha)E[Z], \end{aligned} \quad (4.5)$$

$$\begin{aligned} \text{Var}[C_{X(\alpha)}] &= \text{Var}[\alpha Y + (1 - \alpha)Z] \\ &= \text{Var}[\alpha Y] + \text{Var}[(1 - \alpha)Z] + 2\text{Cov}(\alpha Y, (1 - \alpha)Z) \\ &= \text{Var}[\alpha Y] + \text{Var}[(1 - \alpha)Z] \quad (\text{pois } Y \text{ e } Z \text{ são independentes}) \\ &= \alpha^2 \text{Var}[Y] + (1 - \alpha)^2 \text{Var}[Z]. \end{aligned} \quad (4.6)$$

Calculando-se a variância de $X(\alpha)$, obtém-se a expressão

$$\begin{aligned} \text{Var}[X(\alpha)] &= E[(X(\alpha) - E[X(\alpha)])^2] \\ &= E[X(\alpha)^2] - (E[X(\alpha)])^2 = (\alpha Y^2 + (1 - \alpha)Z^2) - (\alpha Y + (1 - \alpha)Z)^2, \end{aligned} \quad (4.7)$$

a qual também é uma combinação entre Y e Z . Além disso, uma vez que a computação de $C_{X(\alpha)} = E[X(\alpha)]$ se faz necessária no cálculo de $\text{Var}[X(\alpha)]$, a incerteza da v.a. $C_{X(\alpha)}$ deve ser considerada, e, portanto, a *diversidade genética esperada de $X(\alpha)$* , $I_{X(\alpha)}$, pode ser expressa dividindo-a em duas componentes: o valor esperado da diferença quadrática entre as variáveis Y e Z , e a componente de incerteza dada pela variância de $C_{X(\alpha)}$:

$$I_{X(\alpha)} = E[\text{Var}[X(\alpha)]] + \text{Var}[C_{X(\alpha)}], \quad (4.8)$$

4.2. A DIVERSIDADE GENÉTICA DE DUAS SUBPOPULAÇÕES

Tabela 4.1 Resultados da validação por simulação estocástica da diversidade genética em função de α .

α	$E[X(\alpha)]$	$I_{X(\alpha)}$	$S_{X(\alpha)}^2$	Erro absoluto
0,0	2,50	0,0833	0,0835	0,0002
0,1	2,30	0,4433	0,4419	0,0014
0,2	2,10	0,7233	0,7235	0,0002
0,3	1,90	0,9233	0,9241	0,0008
0,4	1,70	1,0433	1,0438	0,0005
0,5	1,50	1,0833	1,0833	0,0000
0,6	1,30	1,0433	1,0437	0,0005
0,7	1,10	0,9233	0,9243	0,0010
0,8	0,90	0,7233	0,7237	0,0005
0,9	0,70	0,4433	0,4427	0,0006
1,0	0,50	0,0833	0,0834	0,0001

em que

$$E[\text{Var}[X(\alpha)]] = \alpha(1 - \alpha)E[(Y - Z)^2]. \quad (4.9)$$

Analisando a equação (4.8), percebe-se que a primeira componente, $E[\text{Var}[X(\alpha)]]$, é idêntica para as populações $X(\alpha)$ e $X(1 - \alpha)$. Nesse caso, em que os valores de α são complementares, vê-se também, da equação (4.6), que

$$\text{Var}[Y] = \text{Var}[Z] \Rightarrow \text{Var}[C_{X(\alpha)}] = \text{Var}[C_{X(1-\alpha)}], \quad (4.10)$$

e, por conseguinte, uma condição suficiente para que os valores esperados de diversidade genética de dois regimes de imigração $X(\alpha)$ e $X(1 - \alpha)$ sejam idênticos é a igualdade das variâncias das subpopulações Y e Z , i.e.,

$$\text{Var}[Y] = \text{Var}[Z] \Rightarrow \text{Var}[X(\alpha)] = \text{Var}[X(1 - \alpha)], \quad (4.11)$$

o que demonstra a inadequabilidade da diversidade genética para distinguir situações de simetria posicional no espaço de soluções.

4.2.2 Exemplo Ilustrativo e Validação

Exemplo (Centróides de $X(\alpha)$ e $X(1 - \alpha)$). A Figura 4.2 ilustra o posicionamento do centróide computado em duas populações com taxas complementares, $X(\alpha)$ e $X(1 - \alpha)$, quando as duas subpopulações locais, Y e Z , possuem variâncias idênticas (utiliza-se $\alpha = 0,7$ nesse exemplo). Na parte de cima da Figura, percebe-se que $C_{X(\alpha)}$ é atraído para a região que apresenta a maior densidade de pontos, nesse caso, a região onde se encontram os pontos amostrados a partir de Y . A parte inferior da Figura 4.2 mostra que, sob uma

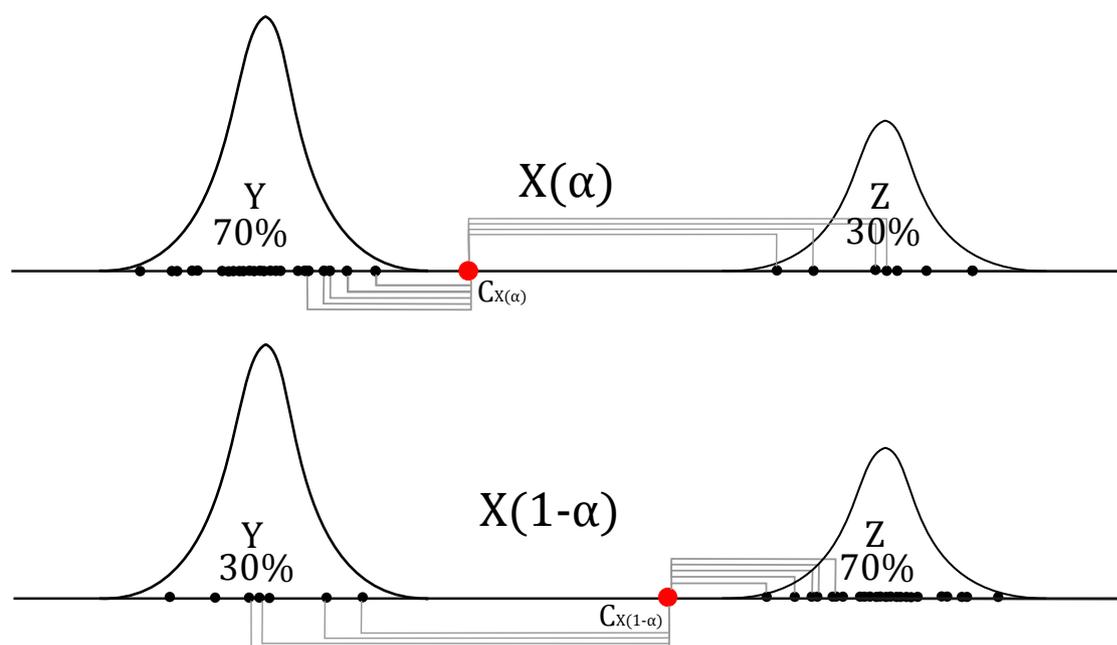


Figura 4.2 Ilustração de dois modelos de populações com taxas de imigração complementares.

taxa de imigração complementar, $C_{X(1-\alpha)}$ é atraído para a região de pontos amostrados a partir de Z. Assim, intuitivamente, vê-se que, devido à simetria das diferenças quadráticas calculadas entre os pontos amostrais e os centróides das populações $X(\alpha)$ e $X(1-\alpha)$, o valor esperado da soma dessas diferenças (i.e., a diversidade genética baseada em distância para o centróide) é aproximadamente o mesmo para taxas de imigração complementares. Essa simetria é, no entanto, quebrada quando as distribuições de Y e Z possuem variâncias diferentes, como ilustra a Figura 4.4: a assimetria em relação à diversidade genética esperada para taxas de imigração complementares aumenta conforme a diferença absoluta entre as variâncias das subpopulações cresce, estando de acordo com o que sugere a condição expressa na equação (4.11). Variâncias diferentes, portanto, implicam em diferentes balanços entre as incertezas associadas quanto às distribuições das diferenças quadráticas dos pontos amostrais gerados em cada subpopulação para o centróide, o que possivelmente constitui o fenômeno responsável pela assimetria observada da Figura 4.4.

Validação (Diversidade genética esperada). A expressão (4.9) é validada neste exemplo por meio de simulação estocástica. A Tabela 4.1 mostra o resultado de simulação de populações da forma $X(\alpha)$, com $Y \sim \text{Uniforme}(0, 1)$ e $Z \sim \text{Uniforme}(2, 3)$, para α

4.2. A DIVERSIDADE GENÉTICA DE DUAS SUBPOPULAÇÕES

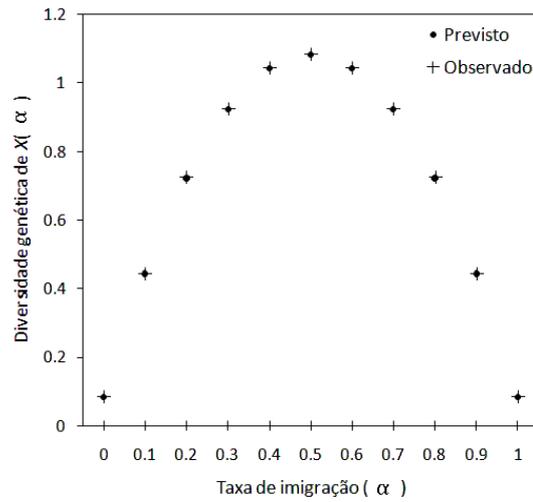


Figura 4.3 Diversidade genética prevista com variâncias das subpopulações idênticas.

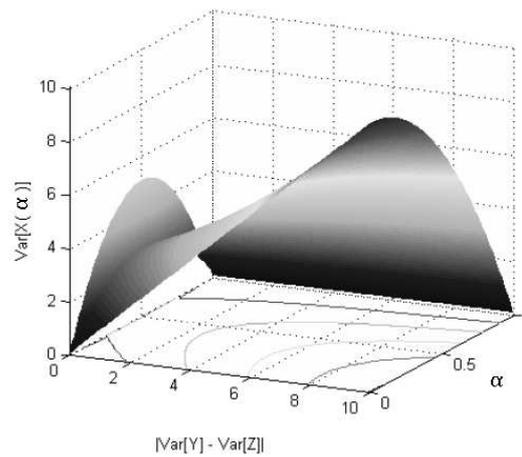


Figura 4.4 Variação de $I_X(\alpha)$ em função da diferença entre as variâncias das subpopulações.

variando de 0,0 a 1,0 em incrementos de 10^{-1} . Nesse exemplo, note-se que

$$\begin{aligned} \text{Var}[Y] &= \int_0^1 (y - \mu_Y)^2 dy = \int_2^3 (z - \mu_Z)^2 dz = \text{Var}[Z], \\ \mu_Y &= E[Y], \mu_Z = E[Z]. \end{aligned} \quad (4.12)$$

Os resultados validam o valor previsto para $\text{Var}[X(\alpha)]$ e ilustram a simetria da diversidade genética em função de α . As simulações consideram o tamanho da população $N = 1000$, em que a média da variância amostral, $S_{X(\alpha)}^2$, é computada sobre 1000 re-

petições. A simetria de $I_X(\alpha)$ é mais bem visualizada na Figura 4.4, a qual indica que a relação funcional entre $I_X(\alpha)$ e α é um polinômio de segundo grau com coeficiente quadrático negativo, para $|Var[Y] - Var[Z]| = 0$. Já a Figura 4.3 mostra a superfície descrita por $I_X(\alpha)$ em função da diferença absoluta entre $Var[Y]$ e $Var[Z]$ e α , onde pode ser observada a região descrita na Figura 4.4, quando as variâncias de Y e Z são iguais. Observa-se também que a assimetria de $I_X(\alpha)$ em relação à α aumenta à medida que a diferença entre as variâncias das subpopulações locais cresce.

Na Seção 4.4, interpreta-se α como a taxa de imigração na estratégia de geração de diversidade por meio de fluxos de imigração, em que Y representa o modelo de geração de imigrantes e Z representa o processo de geração de indivíduos a partir dos operadores convencionais em AEs. A medida de inércia da equação (4.2) é utilizada na Seção 4.3 para avaliar como a diversidade genética influencia a dinâmica evolucionária em MOEAs.

4.3 Correlação entre Diversidade e Hipervolume

Discute-se nesta Seção os resultados de uma série de experimentos conduzidos no intuito de avaliar a influência da diversidade da população na qualidade objetiva da aproximação da PF obtida em MOEAs operando sobre diferentes parametrizações. Em particular, avalia-se o algoritmo NSGA2 sob diferentes regimes de exploração e exploração, regimes esses influenciados pelas escolhas dos parâmetros empregados nos operadores de variação. A análise dos resultados busca determinar a existência de padrões de correlação entre a diversidade genética (equação (4.2)) – a qual representa uma medida de localização da distribuição de diversidade – e a qualidade da PF obtida, em termos do indicador de Hipervolume (Seção 3.2.2).

Cabe ressaltar que Burke *et al.* (2004) realizaram estudo semelhante no contexto da Programação Genética (GP, do inglês), em que a correlação de Spearman entre os valores de seis medidas de diversidade genética e entre valores da aptidão do melhor indivíduo presente na população foi computada a cada geração de um processo evolucionário sobre um conjunto de quatro problemas de otimização. O propósito daquele estudo, no entanto, é o de identificar as métricas de diversidade em GP com os maiores níveis de correlação com a paisagem de aptidão, já que os indivíduos em GP são representados por estrutura de dados não-lineares – e.g. árvores sintáticas em uma linguagem formal –, o que abre a possibilidade de interpretações diversas para a diversidade genética.

4.3. CORRELAÇÃO ENTRE DIVERSIDADE E HIPERVOLUME

Tabela 4.2 Setup experimental para o NSGA2(η_{SBX} , η_{mut}).

Parâmetro	Forma Assumida
Mutação	Polinomial (Seção 3.4.2)
Cruzamento	SBX (Seção 3.4.1)
Taxa de Mutação	$\frac{1}{m}$
Taxa de Cruzamento	0,9
Tamanho da População	$N = 100$
Seleção para Reprodução	Torneio Binário (Seção 3.4.3)

4.3.1 Metodologia e Planejamento dos Experimentos

O NSGA2 (Pseudocódigo 1) implementado utiliza a configuração descrita no trabalho de Deb *et al.* (2000), a qual se encontra na Tabela 4.2, em que m é a dimensionalidade dos vetores no espaço de decisão. O conjunto de problemas de otimização multiobjetivo contínua escolhido é composto pelo pacote de cinco problemas ZDT, cada qual com duas funções-objetivo a serem minimizadas. A formulação dos problemas ZDT1, ZDT2, ZDT3, ZDT4 e ZDT6 encontra-se no Apêndice B, com dimensionalidades variando em $m \in \{10, 30\}$, a depender do problema.

Como forma de avaliar o efeito de 200 gerações do NSGA2 no conjunto ZDT sob diferentes regimes de exploração e exploração, optou-se por um experimento fatorial 3^2 , i.e., com dois fatores e três níveis (Montgomery, 2004). Os fatores escolhidos são os parâmetros denominados η_{SBX} e η_{mut} , os quais influenciam as razões de espalhamento (ver Seções 3.4.1 e 3.4.2) dos operadores de cruzamento e mutação, respectivamente. Intuitivamente, valores menores assumidos por esses parâmetros promovem a geração de filhos distantes dos pais, produzindo, portanto, maiores níveis médios de diversidade e vice-versa. Tanto η_{SBX} quanto η_{mut} assumem os níveis $\{5, 10, 20\}$, produzindo um total de nove combinações por problema, em que a notação NSGA2(η_{SBX}, η_{mut}) é utilizada para indicar a escolha de parâmetros. Cada tratamento é replicado 50 vezes e o Teste-T bicaudal de Welch (Welch, 1947), o qual não supõe igualdade de variâncias populacionais, é utilizado para aferir a significância estatística das diferenças entre as médias dos resultados, ao nível de confiança 0,95.

4.3.2 Resultados e Discussão

Os valores médios de Hipervolume na última geração, observados sob diferentes regimes de exploração/exploração, sugerem que o desempenho do NSGA2 varia significativamente com a escolha dos parâmetros de espalhamento dos operadores de variação. Embora não se possa formular uma conclusão definitiva sobre todo o conjunto ZDT, observa-se que a

Tabela 4.3 Diversidade genética, Hipervolume e Correlação para os problemas ZDT.

$\eta_{\text{SBX}}/\eta_{\text{mut}}$	ZDT1			ZDT2			ZDT3			ZDT4			ZDT6		
	$\Delta\mathcal{S}_h$	$I_{M(R)}$	ρ												
5/5	0,16	1,00	-	0,38	1,00	+	0,29	1,00	-	3,82	1,00	-*	8,25	0,90	-
5/10	0,13	0,99	+	0,30	0,96	-	0,23	0,97	-	3,81	0,95	-*	7,28	0,78	+
5/20	0,11	0,99	+	0,26	0,91	-	0,20	0,95	-	3,48	0,87	-	6,15	0,70	+
10/5	0,17	0,99	-	0,46	0,97	+*	0,28	0,95	+	0,86	0,81	-	13,88	0,97	+
10/10	0,14	0,96	-	0,38	0,97	-	0,22	0,93	+	0,45	0,78	-	12,83	0,79	-
10/20	0,12	0,95	-	0,38	0,91	-	0,19	0,92	+*	0,43	0,73	+	11,00	0,67	-
20/5	0,21	0,97	-	0,59	0,95	+	0,28	0,92	+	0,32	0,74	+	15,73	1,00	-
20/10	0,17	0,96	+*	0,42	0,92	+	0,24	0,90	+	0,11	0,68	-	14,10	0,83	+
20/20	0,13	0,93	+	0,33	0,86	-*	0,19	0,88	+*	0,21	0,65	+	12,35	0,70	+

configuração NSGA2(5, 20) obteve os melhores resultados com significância estatística na maioria dos problemas (ZDT1, ZDT2 e ZDT6), além de um resultado próximo ao das melhores configurações no problema ZDT3. A Tabela 4.3 resume o desempenho médio de cada tratamento sobre os cinco problemas, em que são relatadas as diferenças para o Hipervolume da distribuição ótima de pontos sobre a PF , $\Delta\mathcal{S}_h$, juntamente com os valores de diversidade genética relativa ($I_{M(R)}$) e os sinais dos coeficientes de correlação de Pearson (ρ) entre os valores da última geração de Hipervolume e a diversidade acumulada ao longo das 200 gerações. Note-se que o coeficiente de correlação de Pearson entre duas v.a.'s, digamos X e Y , é definido como a razão entre a covariância das variáveis e o produto dos desvios-padrão, i.e.,

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}. \quad (4.13)$$

Define-se $I_{M(R)}$ como a razão entre a diversidade acumulada em consideração e o maior valor de diversidade acumulada observado entre as 50 réplicas de uma dada escolha de parâmetros. Note-se ainda que as entradas em negrito nas colunas $\Delta\mathcal{S}_h$ da Tabela 4.3 correspondem aos tratamentos que alcançaram os menores valores de $\Delta\mathcal{S}_h$, com diferenças estatisticamente significativas em relação às demais parametrizações avaliadas em um dado problema. Ainda, as entradas seguidas de asterisco indicam correlações estatisticamente diferentes de zero. Uma segunda observação da Tabela 4.3 é que, em geral, as combinações de fatores de espalhamento que resultam em maior diversidade acumulada ao longo do processo evolucionário são aquelas que utilizam valores menores de η_{SBX} e η_{mut} e vice-versa.

Nos estudos relatados ao longo da dissertação, reportam-se os coeficientes de Pearson *amostrais*, computados a partir da covariância amostral e dos desvios-padrão amostrais

4.3. CORRELAÇÃO ENTRE DIVERSIDADE E HIPERVOLUME

entre dois conjuntos de amostras obtidos de duas v.a.'s, a saber, diversidade genética acumulada e Hipervolume. Assim, da Tabela 4.3, com o auxílio da Figura 4.5 (a), percebe-se que, com exceção do problema ZDT3, há predominância de correlações estatísticas negativas na maior parte dos tratamentos (variações paramétricas em η_{SBX} e η_{mut}). Quando se analisam os tratamentos separadamente, observa-se que poucos casos apresentam correlações entre diversidade genética acumulada e Hipervolume significativamente diferentes de zero, como por exemplo, na entrada 20/20 do problema ZDT2 na Tabela 4.3. No entanto, ao se combinarem todos os resultados obtidos em um único conjunto de dados e computarem-se em separado os coeficientes de Pearson, percebe-se melhor o tipo de correlação (positiva ou negativa) induzida pela aplicação do NSGA2 a cada problema em separado, como ilustra a Figura 4.5 (a), em que apenas o MOP ZDT1 não apresentou correlação estatisticamente diferente de zero.

A Figura 4.5 resume as correlações entre diversidade acumulada e Hipervolume final (a) para cada problema, considerando todos os tratamentos; e (b) para cada tratamento, considerando todo o conjunto ZDT, em que as linhas tracejadas indicam os níveis críticos que delimitam a região onde não há significância estatística para os coeficientes de Pearson observados. Na parte (a) da figura, observa-se que o problema ZDT3 é o único a apresentar correlação positiva significativa entre diversidade e Hipervolume. Esse resultado é coerente com o fato do ZDT3 ser o único dentre os problemas testados que possui descontinuidades na PF . Uma hipótese possível é a de que a influência da diversidade genética das últimas gerações no ZDT3 cresce em importância, ao contrário do que ocorre nos demais problemas, visto que as soluções candidatas devem ser agrupadas em cada uma das cinco partições que compõem a PF . Essa hipótese encontra suporte na observação da Figura 4.6, discutida no parágrafo seguinte. Para os demais problemas – com exceção do ZDT1, o qual não possui correlação estatisticamente diferente de zero –, e em especial para o ZDT4 – o qual alcançou a correlação de maior magnitude de todo o conjunto ($\approx -0,57$) –, as correlações negativas sugerem que níveis excessivos de diversidade genética no início do processo repercutem negativamente na velocidade de convergência do NSGA2. Isso porque, como o efeito da perda de diversidade é acentuado à medida que a população se aproxima da PF , os valores atômicos de diversidade genética que mais contribuem para a diversidade acumulada são aqueles das gerações iniciais. Já na Figura 4.5 (b), observa-se que prevalece o efeito do operador de recombinação sobre aquele do operador de mutação em todos os problemas, pois há, para os tratamentos NSGA2(5, η_{mut}), correlações negativas entre diversidade e Hipervolume, enquanto que, para as combinações NSGA2(10, η_{mut}) e NSGA2(20, η_{mut}), há correlações positivas

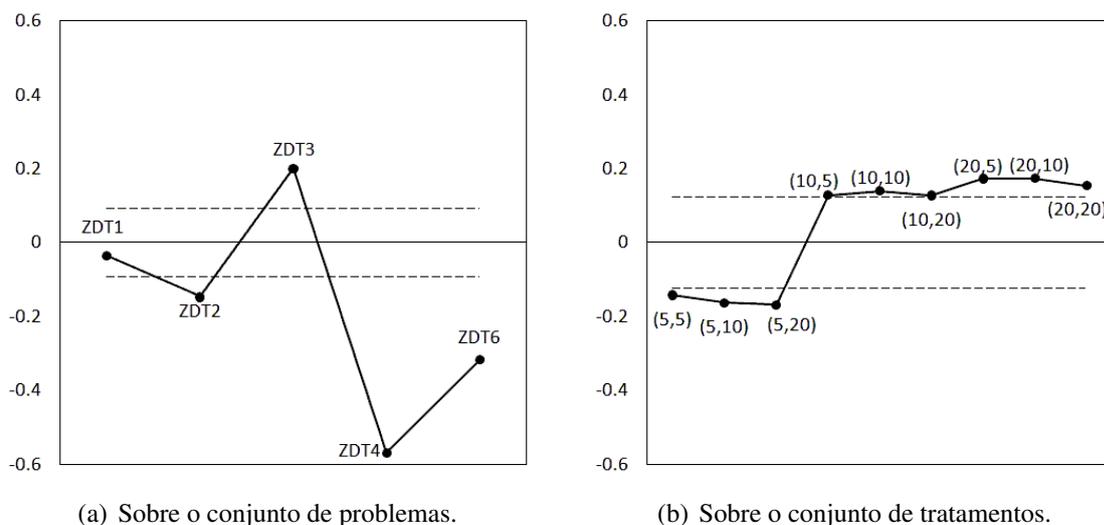
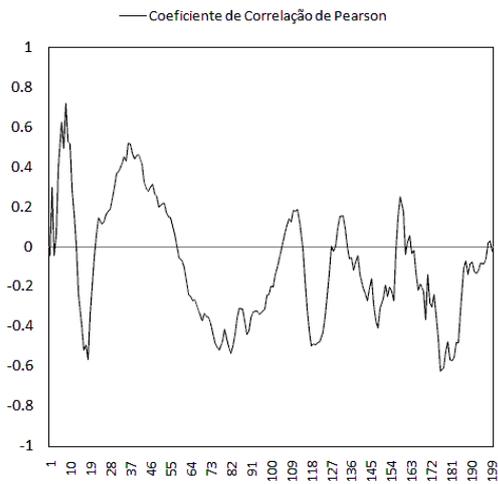


Figura 4.5 Correlação entre diversidade acumulada e Hipervolume na última geração.

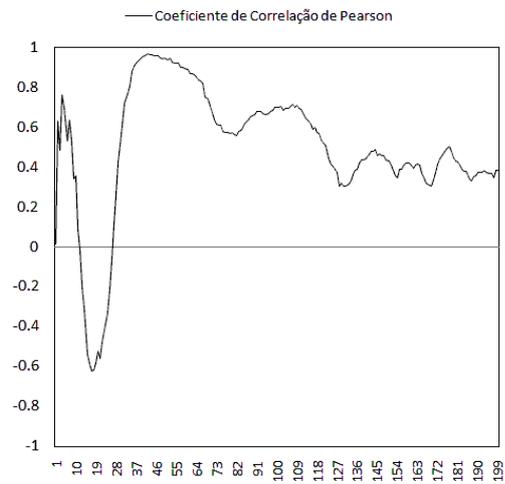
quando todo o conjunto ZDT é considerado. Isso é de certa forma esperado, uma vez que a taxa de cruzamento utilizada é muito maior que a de mutação. No entanto, o fato de todas as correlações observadas serem significativamente diferentes de zero demonstra que, em geral, a diversidade genética possui um papel relevante na qualidade de \mathcal{F}_1 , a aproximação obtida da PF , independentemente da escolha de parâmetros, sendo capaz de explicar grande parte das diferenças de desempenho observadas entre os tratamentos.

Também se estudou a dinâmica das correlações temporais entre diversidade e aptidão ao longo da evolução, ilustrada na Figura 4.6. Nesse estudo, computaram-se os coeficientes de Pearson entre os valores de diversidades genética (momento de inércia) acumulados em janelas temporais de gerações passadas – de tamanhos $J_{\text{tam}} \in \{1, 10, 20\}$ – e o valor observado de Hipervolume em uma geração de referência, ao longo de todo o processo evolucionário. Observou-se, todavia, que, principalmente na fase de ajuste fino das últimas gerações – quando os ganhos no indicador de Hipervolume entre gerações subsequentes são menores –, a correlação entre gerações g e $g + 1$ expõe-se a um nível excessivo de ruído ocasionado pela natureza estocástica dos operadores de variação e das escolhas de sobreviventes, conforme mostra a série de coeficientes de Pearson da Figura 4.7 (a) para o problema ZDT1, cuja função de autocorrelação é plotada na parte (b). Note-se que, com $J_{\text{tam}} = 1$, as correlações entre os retardos temporais (*lags*) mais distantes no tempo oscilam dentro da região delimitada pelos valores críticos de significância estatística. Esse problema é atenuado para tamanhos de janela maiores, em que se espera da diversidade acumulada um papel mais importante na explicação da

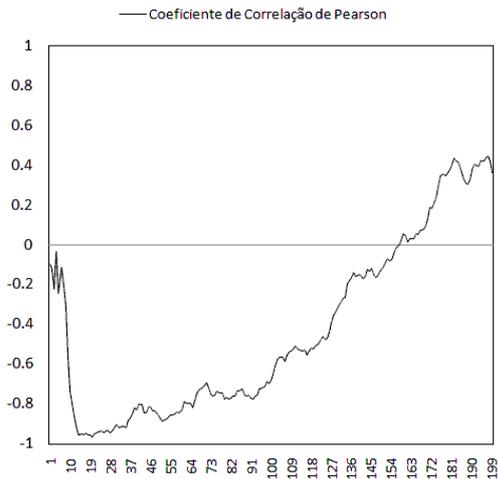
4.3. CORRELAÇÃO ENTRE DIVERSIDADE E HIPERVOLUME



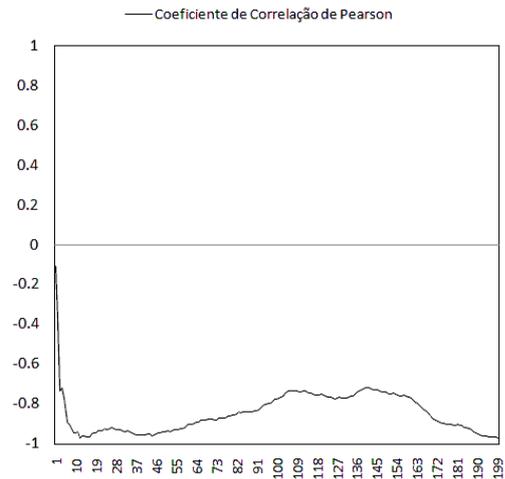
(a) ZDT1.



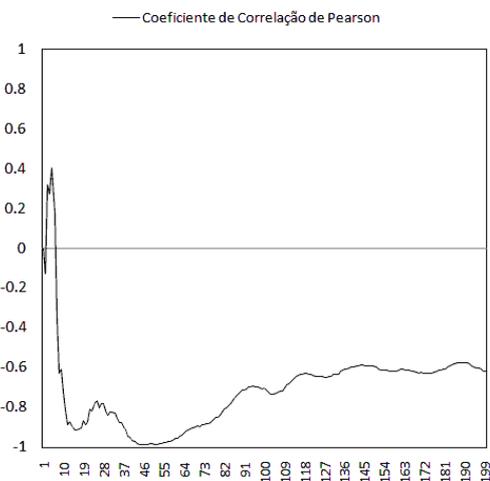
(b) ZDT2.



(c) ZDT3.



(d) ZDT4.



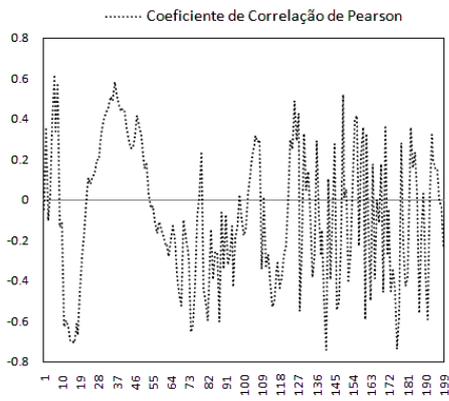
(e) ZDT6.

Figura 4.6 Evolução do coeficiente de Pearson ($J_{tam} = 10$) para o conjunto ZDT.

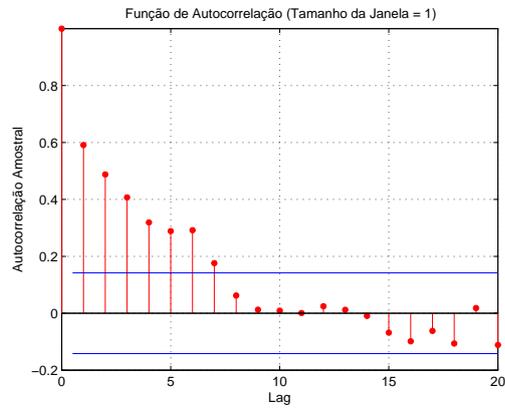
dinâmica do NSGA2 à medida que o ruído é atenuado (partes (c) a (f) da Figura 4.7). Dessa forma, as séries temporais do coeficiente de Pearson ao longo das gerações são computadas para $J_{tam} = 10$ a fim de facilitar a visualização dos padrões temporais de correlação. Ressalta-se, entretanto, que as conclusões da inspeção visual das mesmas séries computadas para outros valores de J_{tam} não se alterariam substancialmente, uma vez que as séries temporais conservam padrões de crescimento/decrescimento discerníveis sob várias escolhas de tamanhos de janelas diferentes. Interessante é observar a existência de padrões de oscilação na evolução do coeficiente sobre o tempo. Em todos os cinco problemas, vê-se na Figura 4.6 que se parte de situações cuja magnitude da correlação diversidade/Hipervolume é praticamente zero, aumentando subitamente nas gerações seguintes. Ainda, em torno da vigésima geração, um ponto de inflexão é observado, após o qual a magnitude da correlação começa a decrescer para, então, seguir tendências específicas em cada problema. Nesse sentido, conforme a hipótese discutida no parágrafo anterior, observa-se que o ZDT3 é o único dentre o conjunto de problemas que apresenta uma tendência de transição entre fortes correlações negativas no início do processo para correlações positivas moderadas na fase de ajuste fino, quando a população encontra-se mais próxima à *PF*. Esse fato sugere que, conforme discutido em parágrafos anteriores, a diversidade genética pode, de fato, influenciar positivamente o desempenho do NSGA2 no ZDT3 nos estágios finais do processo de otimização. Finalmente, cabe ressaltar que as funções de autocorrelação para as séries temporais de coeficientes de Pearson sugerem alto poder preditivo da dinâmica de interação entre diversidade/Hipervolume.

A Figura 4.8 mostra a função de autocorrelação amostral (ACF) computada para os demais MOPs do conjunto ZDT a um tamanho de janela igual a dez. Notadamente, devido à forte tendência de crescimento apresentada na série temporal de coeficientes de Pearson na Figura 4.6 (c), a ACF computada para o problema ZDT3 é aquela em que se visualizam as maiores magnitudes de autocorrelações estatisticamente diferentes de zero para os 20 retardos temporais (*lags*), visualizados na Figura 4.8 (b). O MOP ZDT4, assim como o ZDT3, também apresenta autocorrelações positivas estatisticamente diferentes de zero para todos os *lags* mostrados na Figura 4.8 (c), apesar das magnitudes não serem tão elevadas quanto àquelas observadas para o ZDT4. O MOP ZDT6 também se comporta de forma interessante em relação à computação da ACF: embora não apresente autocorrelação significativamente diferente de zero já a partir do nono *lag*, os *lags* 15 em diante mostram recuperação em direção às intensidades de autocorrelação positivas. A inspeção visual de todos os gráficos de ACF, portanto, aponta para um alto poder preditivo das séries amostrais de coeficientes de Pearson computadas para cada MOP, ao

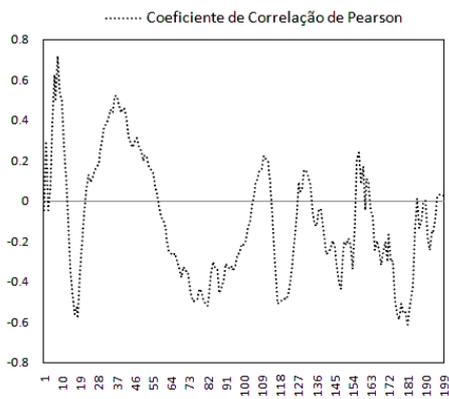
4.3. CORRELAÇÃO ENTRE DIVERSIDADE E HIPERVOLUME



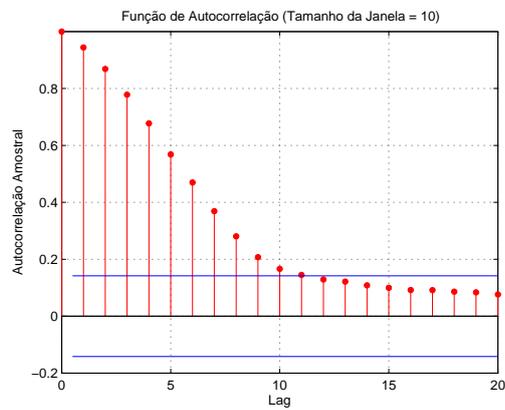
(a) Série ($J_{tam} = 1$)



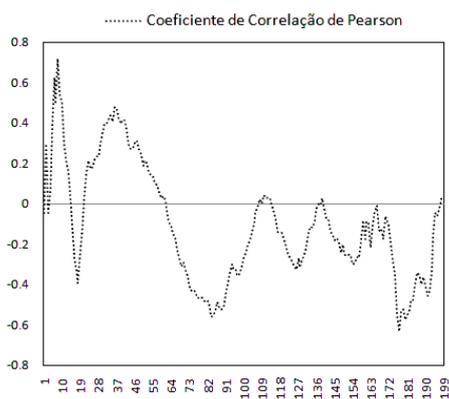
(b) Autocorrelação ($J_{tam} = 1$)



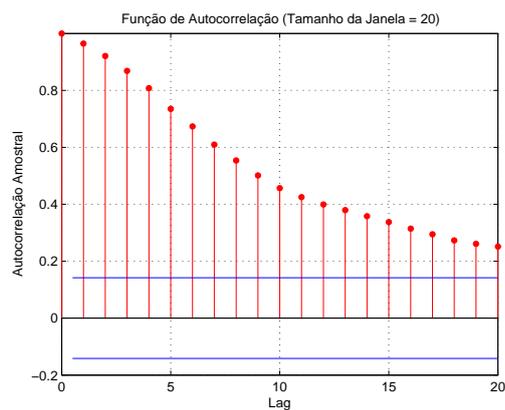
(c) Série ($J_{tam} = 10$)



(d) Autocorrelação ($J_{tam} = 10$)



(e) Série ($J_{tam} = 20$)



(f) Autocorrelação ($J_{tam} = 20$)

Figura 4.7 Correlações temporais e autocorrelação para diferentes tamanhos de janela no ZDT1.

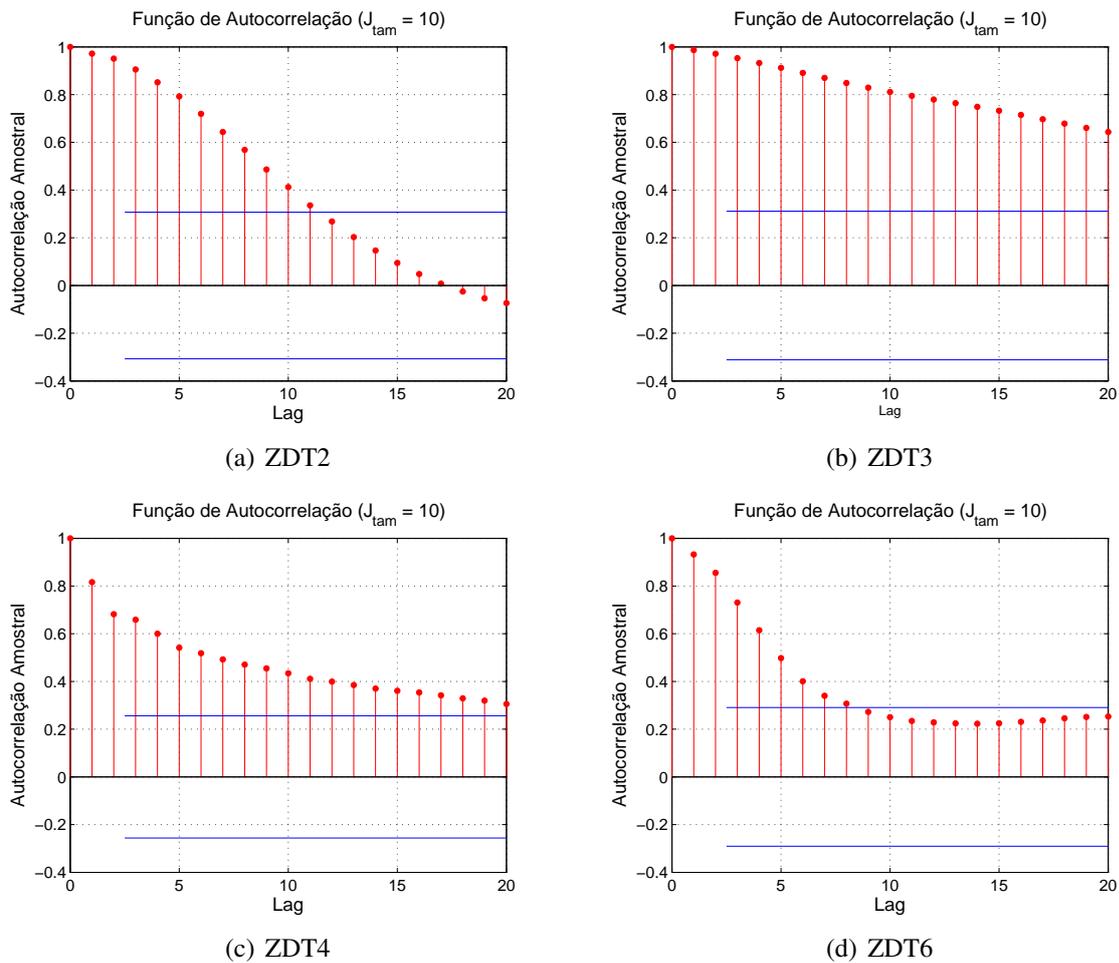


Figura 4.8 Autocorrelação amostral para $J_{tam} = 10$ nos demais MOPs ZDT.

se considerar todos os cenários de variação paramétrica possíveis (combinações de η_{SBX} e η_{mut}), permitindo a representação das séries por meio de modelos algébricos lineares com boa acurácia. Esse resultado sugere a possibilidade do desenvolvimento de estratégias de controle de diversidade em MOEAs baseadas na computação dos coeficientes de Pearson do histórico do processo evolucionário como forma de obter indicadores de quando os níveis de diversidade deveriam ser elevados ou atenuados.

4.4 Geração de Diversidade por Inserção de Imigrantes

Em populações biológicas, a presença de regimes de imigração permite maiores possibilidades de trocas genéticas e, logo, promove novos caminhos evolucionários. Como observado em Reed e Frankham (2003), a perda de diversidade genética reduz o grau

4.4. GERAÇÃO DE DIVERSIDADE POR INSERÇÃO DE IMIGRANTES

de adaptabilidade de uma população em resposta às mudanças do ambiente. Portanto, a inserção de imigrantes é um mecanismo o qual naturalmente pode aumentar os níveis de diversidade genética de uma população. Regimes de imigração biológica são estudados, por exemplo, em Altrock *et al.* (2010), em que se descreve a dinâmica de um sistema com duas populações locais sob regime de imigração assimétrica, em respeito a um único gene. Por exemplo, na Figura 4.9, em se escolhendo $\alpha_1 = \alpha_2$, têm-se um regime simétrico, em que o fluxo de migração é igual em ambas as vias. O regime assimétrico caracteriza-se, portanto, pela escolha de $\alpha_1 \neq \alpha_2$. Altrock *et al.* (2010), relatam os níveis críticos de frequências alélicas relativas conjuntas os quais levam a estados internos estáveis do sistema, em que um alelo heterozigoto (dois alelos diferentes para o mesmo gene), menos apto, é preservado em uma das populações. Os autores relatam que o equilíbrio do sistema é controlado por taxas críticas de imigração, que dependem da aptidão relativa do alelo heterozigoto. Esse estudo é importante porque demonstra a manutenção de diversidade em uma população biológica, mesmo que o alelo heterozigoto possua aptidão menor que os equivalentes homozigotos (alelos iguais para o mesmo gene).

Nessa Seção, o modelo apresentado na equação (4.3) é utilizado para a contextualização das estratégias existentes de manutenção de diversidade genética em AEs por meio de inserção de soluções atípicas (*imigrantes*) na otimização dinâmica, em que Y é interpretada como o modelo de uma subpopulação de imigrantes, Z como o modelo da *subpopulação de trabalho*, evoluída por meio dos operadores de variação e seleção, e α como a *taxa de imigração*. Destacam-se três observações importantes:

1. Optou-se por dividir as abordagens de geração de imigrantes em três categorias. Segundo esta taxonomia, adotada por clareza, o processo de geração de imigrantes é *não-correlacionado* quando Y e Z são v.a.'s estatisticamente independentes; *correlacionado* quando a subpopulação de imigrantes possui dependência estatística em relação à subpopulação de trabalho; e *híbrido* quando há inclusão de K subpopulações de imigrantes adicionais, cada qual com uma taxa de imigração $\alpha_1, \dots, \alpha_K, \sum_{j=1}^K \alpha_j \leq 1$ associada, em que ao menos duas subpopulações difiram quanto à presença de dependências estatísticas em relação à população de trabalho. Todavia, ressalta-se que, para a abordagem híbrida, apenas os casos de duas subpopulações de imigrantes são estudados na literatura, sendo uma gerada a partir da população de trabalho e outra gerada aleatoriamente. Note-se que mecanismos de geração que aproveitam informações da subpopulação de trabalho incluem-se na segunda categoria.
2. A equação (4.3) é um caso especial de um modelo de imigração mais geral, em que

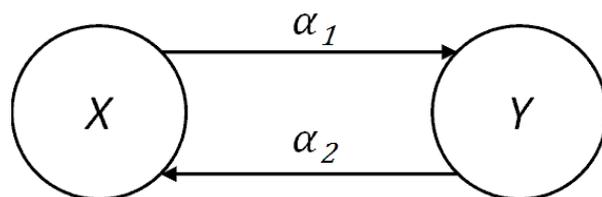


Figura 4.9 Representação de um modelo de migração genérica com taxas assimétricas (Adaptado de Altrock *et al.* (2010)).

duas ou mais subpopulações apresentam fluxos de migração assimétricos de duas vias ao longo do tempo (ver a Figura 4.9). Em relação ao modelo da Figura 4.9, o modelo proposto nesse trabalho (equação (4.3)) considera um fluxo de uma via, uma vez que a taxa de migração de X para Y é zero, ou seja, $\alpha_1 = 0$ e $\alpha \equiv \alpha_2$.

3. É natural pensar-se que a estratégia de inserção de imigrantes seja um caso especial de AEs com isolamento por distância (Azevedo e Gordon, 2009), como o modelo de AGs com ilhas. No entanto, esse não é o caso, pois nas estratégias de isolamento, cada subpopulação evolui independentemente, enquanto que nos modelos discutidos nessa Seção, as subpopulações de imigrantes não estão sujeitas à evolução genética. Ou seja, o modelo de geração de imigrantes não se caracteriza por um processo evolucionário. Os modelos de isolamento, por outro lado, são utilizados para aumentar o grau de paralelismo do processo evolucionário no espaço de busca (Azevedo e Gordon, 2009), ao encorajar a exploração de regiões possivelmente distintas naquele espaço, enquanto que a abordagem de imigrantes visa fornecer meios à população de melhor se adaptar frente às mudanças do ambiente.

A seguir, uma síntese da literatura recente, relevante para o modelo de geração de imigrantes proposto no Capítulo 5, é apresentada.

4.4.1 Imigrantes Não-Correlacionados

A lógica subjacente à inserção de imigrantes aleatórios é a de causar inserção de informação gênica não-correlacionada com o material genético da subpopulação de trabalho. Dessa forma, na ocorrência de mudanças da paisagem de aptidão, a presença da subpopulação de imigrantes aumenta as chances de adaptação da população como um todo ao novo ambiente. O primeiro AE mono-objetivo proposto a fazer uso dessa abordagem é o *Random Immigrants Genetic Algorithm* (RIGA) (Cobb e Grefenstette, 1993), descrito no Pseudocódigo 2. Apesar do modelo descrito na equação (4.3) ser, em princípio, suficien-

Pseudocódigo 2 Random Immigrant Genetic Algorithm

- 1: InicializarPopulação($\Psi^{(0)}$), $t \leftarrow 0$
 - 2: **enquanto** condição de término não satisfeita **faça**
 - 3: Avaliar($\Psi^{(t)}$)
 - 4: Gerar a subpopulação $Y^{(t)}$ composta de $\alpha * N$ imigrantes aleatórios
 - 5: Substituir os piores $\alpha * N$ indivíduos em $\Psi^{(t)}$ por $Y^{(t)}$
 - 6: $\Psi^{(t+1)} \leftarrow$ Seleção(Variação($\Psi^{(t)}$))
 - 7: **fim enquanto**
 - 8: **retorne** $\Psi^{(t)}$
-

temente amplo para considerar *pdfs* arbitrárias na geração dos imigrantes, o RIGA e as demais abordagens derivadas geram imigrantes a partir da v.a.

$$Y \sim \text{Uniforme} \left[(x_1^{\text{inf}}, x_1^{\text{sup}}) \times \dots \times (x_m^{\text{inf}}, x_m^{\text{sup}}) \right], \quad (4.14)$$

isto é, vetores aleatórios são amostrados uniformemente no hipercubo que compreende todo o espaço de busca e tomados como a subpopulação de imigrantes.

No modelo de aprendizagem incremental baseada em população com inserção de imigrantes proposto por Yang e Yao (2005), um vetor de probabilidades é utilizado ao longo do processo evolucionário de forma que todas as soluções são criadas a partir de amostragens desse vetor, o qual é iniciado em $p_i = \frac{1}{2}$ (p_i é a probabilidade do *bit* 1 para a i -ésima variável binária do problema). A cada geração, o vetor de probabilidade é atualizado mediante o aprendizado competitivo na direção da melhor solução presente na população e uma proporção dos piores indivíduos da população é substituída por imigrantes aleatórios. Os autores relatam que a repercussão da introdução de imigrantes aleatórios nos AGs estudados para otimização dinâmica é “*dependente do problema*” sem, no entanto, demonstrarem a existência de intervalos críticos de proporção de imigrantes fora dos quais o desempenho dos AGs seria degradada. Por outro lado, argumentam que um nível “*intermediário*” em torno de 40% produz resultados superiores com significância estatística. Também fora observado que a introdução de imigrantes degrada o desempenho dos AGs estudados em fases estacionárias do processo dinâmico.

Em um dos raros exemplos da aplicação de imigrantes aleatórios em ambientes estáticos (i.e., funções-objetivo invariantes no tempo), Hornby (2009) propõe estruturar a população de um AG para otimização contínua utilizando o conceito de *aging* (envelhecimento). A estrutura agrupa indivíduos com idades semelhantes em duas ou mais subpopulações (camadas) com restrição por idade máxima, em que a idade é uma medida do número de gerações no qual um indivíduo é mantido na população, por exemplo, por

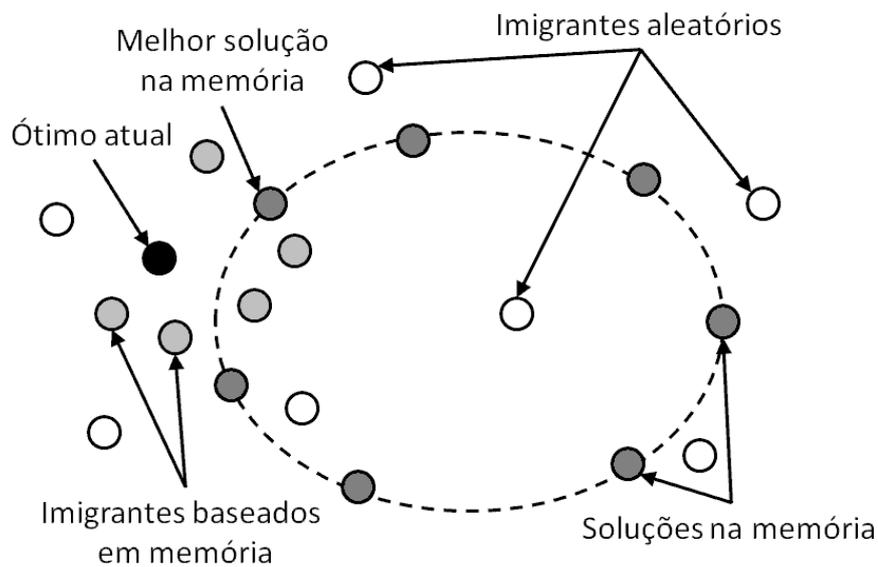


Figura 4.10 Geração de imigrantes correlacionados com base em memória (Adaptado de Yang (2008)).

meio de elitismo. Indivíduos gerados por operadores de variação herdam a idade do pai mais velho. A troca de material genético entre indivíduos pertencentes a diferentes camadas se dá por meio de seleção para reprodução em camadas subsequentes ou quando a idade de um indivíduo extrapola a idade máxima permitida para aquela camada, situação em que uma cadeia de eventos é disparada de forma a mover esse indivíduo e os indivíduos que ele irá substituir para camadas superiores. Imigrantes aleatórios são gerados regularmente em um intervalo pré-definido de gerações e introduzidos na camada mais baixa, substituindo-a por completo. Dessa forma, mantém-se bons níveis de diversidade à medida que se tenta preservar bons indivíduos, candidatos à promoção às camadas superiores. Os resultados demonstram que essa estratégia mantém altos níveis de diversidade na população e leva o AG a explorar diferentes regiões promissoras em problemas multimodais.

4.4.2 Imigrantes Correlacionados

O objetivo principal das estratégias de geração de imigrantes correlacionados com a subpopulação de trabalho é o de manter diversidade genética em função das regiões exploradas em estados anteriores do processo evolucionário. O *modus operandi* dessas técnicas consiste em manter atualizado um arquivo externo dos melhores pontos visitados no espaço de busca de tal forma que as melhores memórias avaliadas de acordo com o

4.4. GERAÇÃO DE DIVERSIDADE POR INSERÇÃO DE IMIGRANTES

ambiente atual são utilizadas para a geração de imigrantes (Branke, 1999). O principal efeito dessas abordagens é o de viesar a geração de diversidade no ambiente atual, uma vez que os novos imigrantes são introduzidos em regiões próximas às melhores memórias armazenadas em relação ao novo ambiente (ver Figura 4.10).

Yang (2008) compara o desempenho de AGs utilizando duas estratégias de imigrantes correlacionados: (1) baseada em elitismo e (2) baseada em memória. Em ambas as estratégias, os imigrantes são gerados a partir de uma série de mutações de soluções passadas, aumentando as chances de adaptação ao novo ambiente. Na estratégia baseada em elitismo, os melhores indivíduos do estado anterior do processo são usados como base enquanto que, na estratégia baseada em memória, os imigrantes são gerados a partir das melhores memórias armazenadas. Note-se que a abordagem de elitismo pode ser vista como um caso especial da baseada em memória, em que a memória é atualizada a cada geração e cujas soluções armazenadas são retiradas da memória sem reposição. O autor argumenta que, nas duas abordagens, a diversidade da população é mantida e a busca se dá com maior eficiência (i) se as mudanças no ambiente forem pequenas (elitismo); ou (ii) se os novos ambientes forem correlacionados com estados passados do processo evolucionário (memória).

Yu *et al.* (2008) propõem a geração de imigrantes segundo a estimação da distribuição de frequências alélicas em problemas dinâmicos com representação binária. Nessa abordagem, os piores indivíduos da população são substituídos por imigrantes amostrados do modelo probabilístico da população atual. Para lidar com mudanças abruptas no ambiente, propõem a geração de imigrantes por meio da amostragem da distribuição alélica complementar. Os resultados demonstram que o aproveitamento de informações da população atual melhora o desempenho do AG implementado, pois introduz imigrantes com níveis de aptidão correlacionados com a população no ambiente atual. Entretanto, se o ambiente muda rapidamente ou se o grau de mudança é elevado, a estratégia proposta é superada em desempenho pelas estratégias convencionais de geração de imigrantes não-correlacionados com o estado anterior à mudança.

Finalmente, Yu *et al.* (2009) apresentam uma categorização para os mecanismos de geração de imigrantes correlacionados. Os *mecanismos diretos* são aqueles cujos imigrantes são gerados a partir dos indivíduos existentes na população. Já os *mecanismos indiretos* constroem um modelo (não necessariamente probabilístico) a partir do qual imigrantes são gerados. Os autores utilizam novas medidas de robustez e diversidade. A *robustez média* é definida como a razão entre os níveis médios de aptidão de populações subsequentes em relação ao índice temporal do ambiente. Os autores observam que

os mecanismos de geração de imigrantes indiretos produzem populações mais robustas em relação àsquelas produzidas pelos mecanismos de geração diretas em quase todo o conjunto de casos de teste. Por outro lado, o *desempenho médio* (aptidão média da população acumulada entre todas as gerações) é maior nas abordagens diretas.

4.4.3 Estratégias Híbridas

Na abordagem baseada em memória associativa, um modelo probabilístico associado ao vizinho mais próximo presente na memória é reativado. Yang e Yao (2008) estudaram as interações desse modelo com diversas estratégias híbridas para problemas dinâmicos as quais combinam múltiplas populações, imigrantes aleatórios e reinicializações. Diversas conclusões são elaboradas em termos da adequação de combinações específicas de mecanismos de imigrantes para variados níveis de dificuldade dos problemas. Como sugestão de trabalhos futuros, os autores propõem a investigação da estratégia de memórias associativas em algoritmos de estimação de distribuição.

Cheng e Yang (2010) propõem melhoramentos nas estratégias de imigrantes para lidarem com soluções inviáveis em problemas combinatórios. As propostas são experimentalmente avaliadas no problema da construção de redes *ad hoc multicast* com topologia variável no tempo. Três estratégias para lidar com soluções inviáveis são descritas, aplicadas a cada mudança de ambiente detectada: (1) substituir os indivíduos inviáveis por imigrantes aleatórios; (2) reparar os indivíduos inviáveis, tornando-os viáveis no novo ambiente, e atualizar a elite; e (3) com probabilidade $\frac{1}{2}$, utilizar a estratégia (1) ou (2). Os resultados demonstram que a estratégia híbrida (3) foi a mais promissora para mudanças suaves no ambiente, enquanto que a estratégia (2) comportou-se melhor para ambientes com mudanças abruptas. Os autores atribuem o bom desempenho da estratégia elitista ao mecanismo de reparação de soluções.

4.4.4 Imigrantes na Otimização Multiobjetivo

As investigações concernentes à inserção de imigrantes em MOEAs encontram-se em fase incipiente. Em um estudo recente, Goh e Tan (2009) apresentam um paradigma competitivo/cooperativo para a otimização multiobjetivo em ambientes dinâmicos. Os autores adotam a estratégia de inserção de imigrantes não-correlacionados, em que um processo de amostragem por hipercubos latinos (Tang, 1993) é aplicado para gerar os imigrantes uniformemente no espaço de decisão. Os imigrantes, chamados de “*competidores estocásticos*”, são produzidos quando mudanças no ambiente são detectadas. A contribuição

dos imigrantes para o ambiente é avaliada em um mecanismo de competição, em que competidores são escolhidos das demais subpopulações. Se os imigrantes emergirem como a subpopulação vencedora, a subpopulação competidora é reinicializada na região em que os imigrantes foram amostrados. Um arquivo externo de soluções não-dominadas também é utilizado. Sempre que uma mudança de ambiente é detectada, parte das soluções não-dominadas anteriores à mudança é armazenada na memória. As soluções não-dominadas escolhidas são os extremos em relação à cada objetivo e, caso haja espaço na memória, o restante das soluções são escolhidas aleatoriamente. Em seguida, toda a primeira Frente, \mathcal{F}_1 , é descartada e atualizada de acordo com a reavaliação das soluções contidas no arquivo externo.

4.5 Discussão Final e Síntese

Nesse Capítulo, discutiu-se a existência de correlações estatísticas significativas entre diversidade genética e aptidão na dinâmica de convergência em MOEAs. Esse resultado sugere que o desempenho dos MOEAs é influenciado pela natureza dos mecanismos de geração de diversidade, e que essa influência varia em conformidade com o estágio do processo evolucionário. Também se observou que a diversidade gerada pelos operadores de variação não é suficiente para evitar a perda de diversidade ao longo das gerações. Dessa forma, discutiram-se os modelos de geração de diversidade baseados em imigrantes. Para a contextualização das técnicas de inserção de imigrantes, propôs-se um modelo de duas subpopulações e obteve-se uma expressão, dada pela equação (4.8), a qual permite a predição do valor esperado de diversidade genética, com médias e variâncias das subpopulações conhecidas, em função da taxa de imigração α . Ademais, por meio de um exemplo ilustrativo, demonstrou-se a inadequabilidade da medida de diversidade genética para distinguir simetrias distribucionais no espaço de busca. Também se comentou sobre o papel da diversidade genética na busca por balanços adequados entre exploração e exploração em AEs. Dessa forma, o Capítulo constitui um elo lógico entre os conceitos fundamentais até aqui apresentados e a proposta de um modelo algorítmico de geração de diversidade genética em MOEAs, formulado no próximo Capítulo.

Como contribuições, listam-se: (i) a caracterização da influência da diversidade na dinâmica de convergência do NSGA2 em direção à PF ; (ii) a derivação analítica da diversidade genética de um modelo de v.a.'s de duas subpopulações locais e a ilustração das limitações da medida de Inércia (equação (4.2)); e (iii) uma taxonomia alternativa para as técnicas de geração de imigrantes em AEs com base nesse modelo.

5

Diversidade Dirigida na Otimização Multiobjetivo

*Os problemas significativos que enfrentamos não podem ser resolvidos
no mesmo nível de pensamento de quando nós os criamos.*

– Albert Einstein

Em problemas cujas funções-objetivo variam no tempo, a manutenção de diversidade genética acima de certos limiares deve ser promovida. No entanto, o excesso de diversidade em períodos de estacionariedade do problema repercute negativamente na velocidade de convergência da população de soluções (Yang, 2008), exigindo maior esforço computacional (i.e., um número maior de gerações) para alcançar as regiões onde se encontram os pontos ótimos no espaço de decisão. Para atenuar esse efeito, o processo de geração de diversidade pode aproveitar a distribuição de soluções existentes na população, de forma a simultaneamente gerar diversidade e explorar regiões promissoras. Como visto no Capítulo anterior, os mecanismos de geração de diversidade por inserção de imigrantes vêm sendo aprimorados em trabalhos recentes para esse fim – sobretudo na otimização evolucionária mono-objetivo em ambientes dinâmicos. O propósito desse Capítulo, portanto, é o de contribuir para o estado da arte na geração de diversidade genética em MOEAs, um tema ainda pouco estudado na literatura, por meio da proposta de utilização de estatísticas de ordem multivariadas nos mecanismos de geração de imigrantes, capazes de descrever probabilisticamente as relações de dominância em diferentes regiões do espaço de objetivos e de promover a obtenção de novas soluções não-dominadas para a exploração da Frente de Pareto (PF).

5.1 Probabilidades de Dominância

Os principais MOEAs existentes, como o NSGA2 e o SPEA2 (ver Apêndice C), implementam critérios de seleção baseados nas relações de dominância entre soluções. Mais especificamente, como visto no Capítulo 3, a dominância de Pareto expressa quais vetores no espaço de objetivos são preferíveis em relação aos demais. Nesse contexto, exemplificam-se a seguir duas situações em que estatísticas de ordem multivariadas podem ser utilizadas para inferir informações de dominância entre soluções.

Exemplo 1. Por não ser uma ordem total, um dos problemas da tomada de decisões baseadas nas relações de ordem de Pareto em MOEAs é a perda de pressão de seleção em certos estágios do processo evolucionário, já que, em situações onde há apenas soluções mutuamente não-dominadas, não há como expressar preferências em relação à população. Dessa forma, a continuidade do caminhamento em direção à PF depende da descoberta de novas soluções capazes de dominar (soluções *dominantes*) as demais. A existência de apenas soluções não-dominadas na população, no entanto, não impede a identificação de regiões em torno das quais há maiores chances de que novas soluções dominantes sejam encontradas. Nesse caso, pode ser interessante computar a probabilidade

$$P\{\mathbf{x} \text{ dominar } j \text{ ou mais soluções da população}\} \quad (5.1)$$

como suporte à decisão em critérios de seleção de pais e de sobreviventes, em que \mathbf{x} é uma solução arbitrária e $j < N$ um escalar menor que o tamanho da população, N .

Exemplo 2. Na geração de diversidade por inserção de imigrantes em MOEAs é desejável estimar as chances de sobrevivência dos imigrantes na próxima geração. A extensão do tempo de vida dos imigrantes é importante, pois permite o compartilhamento de informações genéticas com a subpopulação de trabalho, atenuando o efeito de perda de diversidade da população. Desse modo, pode ser interessante computar a probabilidade

$$P\{\mathbf{x} \text{ ser dominada por uma ou mais soluções da população}\}. \quad (5.2)$$

As probabilidades de dominância indicadas nos dois exemplos podem ser calculadas por meio de estatísticas de ordem de Pareto em \mathcal{R}^M (M é o número de objetivos em um

MOP), discutidas no Capítulo 2. No Exemplo 2, a equação (5.2) é expressa por

$$\begin{aligned} P_{D_j}(\mathbf{x}) &= P\{\mathbf{x} \text{ ser dominada por } j \text{ ou mais soluções da população}\} \\ &= \sum_{k=j}^N \binom{N}{k} [P(\mathbf{x})]^k [1 - P(\mathbf{x})]^{N-k}, \end{aligned} \quad (5.3)$$

para $j = 1$ (i.e., $P_{D_1}(\mathbf{x})$), em que $P(\mathbf{x})$ é a *cdf* da distribuição da população no espaço de objetivos e N é o tamanho da população, $(\mathbf{X}_1, \dots, \mathbf{X}_N)$.

Demonstração. Considere \preceq uma ordem de Pareto e defina $C = \sum_{k=1}^N I(\mathbf{X}_{(k)}, \mathbf{x})$, com

$$I(\mathbf{v}, \mathbf{u}) = \begin{cases} 1, & \mathbf{v} \preceq \mathbf{u} \\ 0, & \text{caso contrário.} \end{cases} \quad (5.4)$$

Então, $C \sim \text{Binomial}[N, P(\mathbf{x})]$. Ainda, uma vez que $\mathbf{X}_{(j)} \preceq \mathbf{x} \iff C \geq j$, então

$$\begin{aligned} P_{D_j}(\mathbf{x}) &= P\{\mathbf{X}_{(j)} \preceq \mathbf{x}\} \\ &= P\{C \geq j\} \\ &= \sum_{k=j}^N P\{C = k\} \\ &= \sum_{k=j}^N \binom{N}{k} [P(\mathbf{x})]^k [1 - P(\mathbf{x})]^{N-k}, \end{aligned} \quad (5.5)$$

em que $\mathbf{X}_{(j)}$ denota a j -ésima estatística de ordem de Pareto no espaço de objetivos. \square

Como $P_{D_j}(\mathbf{x})$ é também a *cdf* da estatística de j -ésima ordem da população no espaço de objetivos, resulta que o resultado anterior também poderia ter sido obtido por meio da integração da *pdf* de $\mathbf{X}_{(j)}$, expressa pela equação (2.37), Capítulo 2:

$$P_{D_j}(\mathbf{x}) = \int \cdots \int_{\mathbf{y} \preceq \mathbf{x}} p_{\mathbf{x}_{(j)}}(\mathbf{x}) d\mathbf{x}, \quad (5.6)$$

em que $p_{\mathbf{x}_{(j)}}(\mathbf{x})$ denota a *pdf* da estatística de $\mathbf{X}_{(j)}$ (ver Capítulo 2).

De forma semelhante, pode-se mostrar que

$$\begin{aligned} P_{Dr_j}(\mathbf{x}) &= P\{\mathbf{x} \text{ dominar } j \text{ ou mais soluções da população}\} & (5.7) \\ &= \sum_{k=j}^N \binom{N}{k} [1 - P(\mathbf{x})]^k [P(\mathbf{x})]^{N-k}. \end{aligned}$$

Nota-se das equações (5.3) e (5.7) que as probabilidades das relações de dominância de Pareto podem ser modeladas por v.a.'s binomiais. Por meio de $P_{D_j}(\mathbf{x})$ e $P_{Dr_j}(\mathbf{x})$, obtém-se

$$\begin{aligned} P_{C_j}(\mathbf{x}) &= P\{\mathbf{x} \text{ ser comparável a } j \text{ ou mais soluções da população}\} & (5.8) \\ &= P\{\mathbf{x} \text{ ou ser dominada por, ou dominar } j \text{ ou mais soluções da população}\} \\ &= P_{D_j}(\mathbf{x}) [1 - P_{Dr_j}(\mathbf{x})] + P_{Dr_j}(\mathbf{x}) [1 - P_{D_j}(\mathbf{x})] \\ &= P_{D_j}(\mathbf{x}) + P_{Dr_j}(\mathbf{x}) - 2 [1 - P_{D_j}(\mathbf{x})] [1 - P_{Dr_j}(\mathbf{x})], \end{aligned}$$

em que duas soluções arbitrárias, digamos \mathbf{x} e \mathbf{y} , são ditas comparáveis se \mathbf{x} dominar \mathbf{y} ou se \mathbf{y} dominar \mathbf{x} . Em outras palavras, duas soluções são consideradas comparáveis nessa dissertação se e somente si pertencerem a Frentes diferentes:

Definição 5.1 (Soluções Comparáveis). Sejam duas soluções pertencentes a Frentes arbitrárias, $\mathbf{x} \in \mathcal{F}_i$ e $\mathbf{y} \in \mathcal{F}_j$, com $i, j \leq k \leq N$, em que k é o número de Frentes identificadas na população de tamanho N . As soluções \mathbf{x} e \mathbf{y} são ditas *comparáveis* $\iff i \neq j$.

As três equações anteriores permitem a obtenção da distribuição de *ranks* de Pareto, com $Rank(\mathbf{x}) = j \iff \mathbf{x} \in \mathcal{F}_j$, em que \mathcal{F}_j é a j -ésima Frente identificada utilizando o procedimento descrito no NSGA2 (página 42). Por exemplo, a probabilidade de \mathbf{x} pertencer à N -ésima Frente é pequena, uma vez que, para tanto, uma ordem total $\mathbf{x}_1 \preceq \mathbf{x}_2 \preceq \dots \preceq \mathbf{x}_N$ deve ocorrer. Essa probabilidade é expressa por

$$\begin{aligned} P\{\mathbf{x} \text{ ser dominada por } N - 1 \text{ soluções e cada } \mathbf{x}_i \text{ ser comparável a } N - 1 \text{ soluções}\} = \\ P\{\mathbf{x} \in \mathcal{F}_N\} = P_{D_{N-1}}(\mathbf{x}) \prod_{i=1}^N P_{C_{N-1}}(\mathbf{x}_i). \end{aligned} \quad (5.9)$$

Uma expressão de particular interesse para a geração de diversidade dirigida é esta:

$$\begin{aligned} P\{\mathbf{x} \in \mathcal{F}_1\} &= P\{\mathbf{x} \text{ ser não-dominada na população}\} & (5.10) \\ &= 1 - P_{D_1}(\mathbf{x}), \end{aligned}$$

denotada por *Probabilidade de Não-Dominância* (PND), a qual é incorporada nos modelos propostos de inserção de imigrantes em MOEAs na Seção 5.3.

5.2 Paisagens de Não-Dominância (NDLs)

Define-se a seguir um conceito chave para guiar a amostragem de soluções que irão compor a subpopulação de imigrantes em problemas de otimização multiobjetivo.

Definição 5.2 (Paisagem de Não-Dominância). Uma *paisagem de não-dominância* corresponde à superfície que descreve a variação de probabilidade de não-dominância (equação (5.10)) em termos dos valores das funções-objetivo para uma dada população.

A visualização das Paisagens de Não-Dominância (doravante NDL, de *Non-Dominance Landscapes*) indica as regiões no espaço de objetivos onde um vetor amostrado aleatoriamente possui maiores chances de ser não-dominado em relação à população atual. A Figura 5.1 mostra diferentes NDLs determinadas ao longo da evolução para o problema ZDT1 com o algoritmo NSGA2. Percebe-se que a forma das Paisagens muda em conformidade com o caminhar à *PF*. Por exemplo, comparando-se as Figuras 5.1 (a) e 5.1 (d), percebe-se que, além da área das regiões de mais alta probabilidade ser menor na geração $g = 100$ do que em $g = 25$, a curvatura da Paisagem parece ser mais consoante com o formato da *PF* real, a qual pode ser visualizada na Figura B.2 do Apêndice B. Observa-se também que as soluções no entorno do ponto de utopia (ver Definição 3.16) – no caso dos problemas ZDT, o ponto $(0, 0)$ – possuem probabilidade de não-dominância, PND, igual a um. Contudo, tal região não é factível e, portanto, nunca estará contida na região onde se encontram as populações em quaisquer estágios do processo evolucionário. Embora a maior parte das regiões com probabilidade unitária seja não-factível, ou ainda, inalcançável a partir dos operadores de variação sobre a população atual, as NDLs são capazes de indicar regiões alternativas onde há maiores chances de se encontrarem soluções não-dominadas. Uma vez que o mapeamento entre os espaços de decisão e o de objetivos é conhecido, mesmo que, em ambientes incertos¹, de forma aproximada, essa observação sugere o uso das Paisagens para revelar as regiões no espaço de busca mais promissoras para conduzir a população à *PF*.

¹Problemas de otimização cuja avaliação das funções-objetivo está sujeita a ruídos.

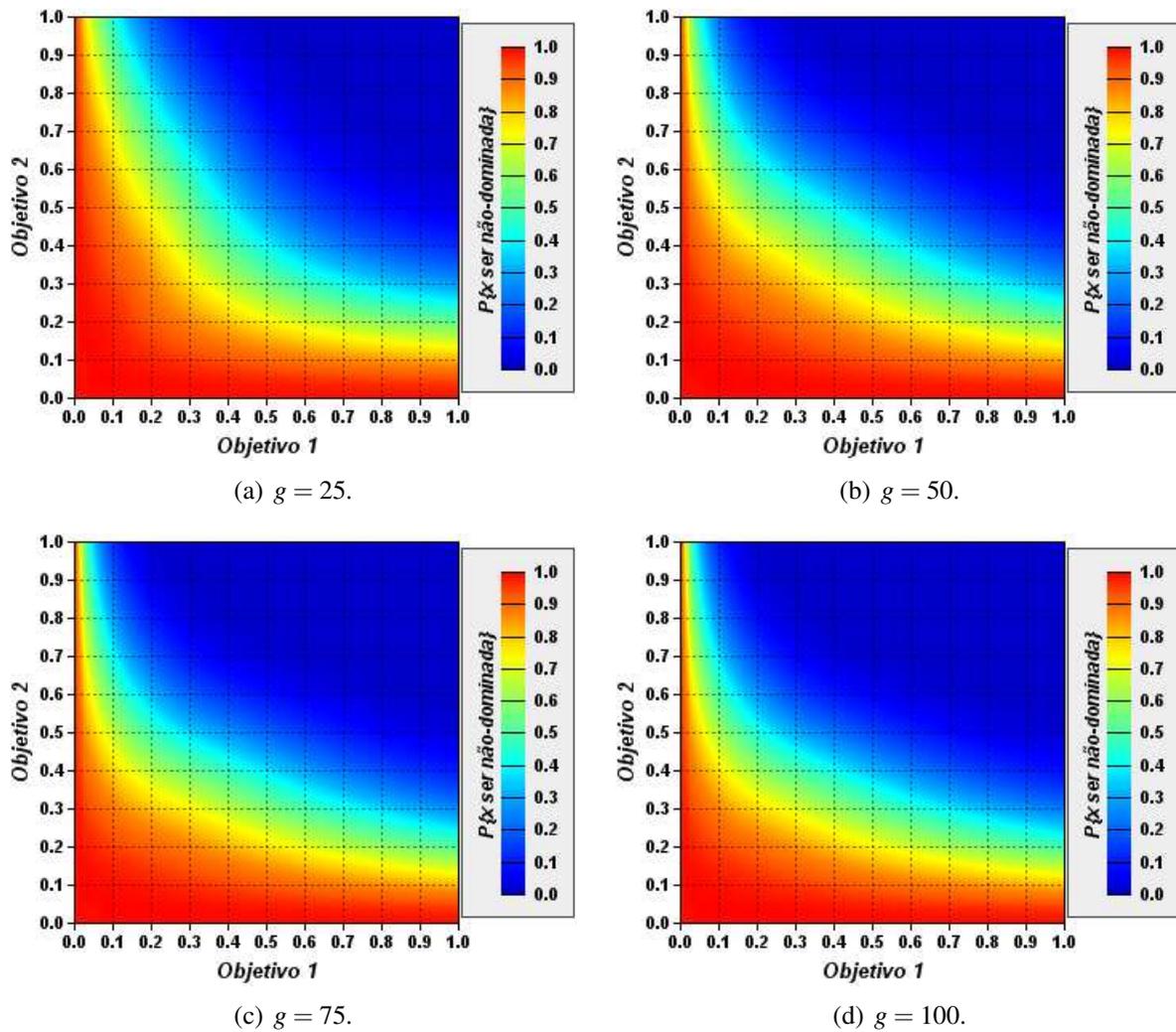


Figura 5.1 Paisagens de Não-Dominância em diferentes gerações (g) do problema ZDT1.

5.2.1 Teorema da Monotonicidade de Elementos Maximais

As Frentes \mathcal{F}_j ($1 \leq j \leq N$) existentes durante a evolução da população – as quais podem ser identificadas pelo procedimento empregado no NSGA2 (ver página 42) – se caracterizam pela inexistência de soluções dominantes em relação às Frentes superiores. Mais precisamente, dada um Frente j arbitrária, vale a propriedade:

$$(\nexists \mathbf{x} \in \mathcal{F}_j), (\mathbf{x} \preceq \mathbf{y}, \mathbf{y} \in \bigcup_{i=1}^{j-1} \mathcal{F}_i), \quad (5.11)$$

em que \mathcal{F}_j corresponde a uma das $k \leq N$ Frentes identificadas na população, com \mathcal{F}_1 denotando o conjunto de soluções mutuamente não-dominadas. Antes de prosseguir ao resultado relacionado para as NDLS, define-se um *elemento maximal* de uma Frente \mathcal{F}_j como aquele de maior PND (equação (5.10)) em relação aos demais elementos de \mathcal{F}_j :

Definição 5.3 (Elemento maximal). Um elemento $\mathbf{x}_{\max}^j \in \mathcal{F}_j$ é maximal em \mathcal{F}_j se

$$P\{\mathbf{x}_{\max}^j \in \mathcal{F}_j\} \geq P\{\mathbf{y} \in \mathcal{F}_j\}, \forall (\mathbf{y} \in \mathcal{F}_j). \quad (5.12)$$

Uma propriedade das NDLS que reflete a propriedade da equação (5.11) é descrita no *Teorema da Monotonicidade de Elementos Maximais*, formulado a seguir.

Teorema 1 (Monotonicidade das PNDs em relação aos elementos maximais). *Dadas as Frentes $\mathcal{F}_1, \dots, \mathcal{F}_k$ e os respectivos elementos maximais, $\mathbf{x}_{\max}^1, \dots, \mathbf{x}_{\max}^k$, com $k \leq N$, vale a desigualdade $P\{\mathbf{x}_{\max}^1 \in \mathcal{F}_1\} \geq P\{\mathbf{x}_{\max}^2 \in \mathcal{F}_1\} \geq \dots \geq P\{\mathbf{x}_{\max}^k \in \mathcal{F}_1\}$.*

Demonstração. Sejam $\mathbf{x}_{\max}^{j-1}, \mathbf{x}_{\max}^j$ os maximais de duas Frentes $\mathcal{F}_{j-1}, \mathcal{F}_j$ consecutivas. Então, têm-se apenas duas possibilidades excludentes:

$$\begin{cases} \mathbf{x}_{\max}^{j-1} \preceq \mathbf{x}_{\max}^j & (I), \text{ ou} \\ \mathbf{x}_{\max}^{j-1} \not\preceq \mathbf{x}_{\max}^j \text{ e } \mathbf{x}_{\max}^{j-1} \not\preceq \mathbf{x}_{\max}^j & (II). \end{cases} \quad (5.13)$$

(I) Se $\mathbf{x}_{\max}^{j-1} \preceq \mathbf{x}_{\max}^j$, por monotonicidade da *cdf* $P_{D_1}(\mathbf{x}) = \int \dots \int_{\mathbf{x}_i \preceq \mathbf{x}} p_{\mathbf{x}(1)}(\mathbf{x}) d\mathbf{x}$, tem-se:

$$\begin{aligned} P_{D_1}(\mathbf{x}_{\max}^{j-1}) &\leq P_{D_1}(\mathbf{x}_{\max}^j) \\ \Rightarrow 1 - P_{D_1}(\mathbf{x}_{\max}^{j-1}) &\geq 1 - P_{D_1}(\mathbf{x}_{\max}^j) \\ \Rightarrow P\{\mathbf{x}_{\max}^{j-1} \in \mathcal{F}_1\} &\geq P\{\mathbf{x}_{\max}^j \in \mathcal{F}_1\}. \end{aligned} \quad (5.14)$$

(II) Se, do contrário, $\mathbf{x}_{\max}^{j-1} \not\leq \mathbf{x}_{\max}^j$ e $\mathbf{x}_{\max}^{j-1} \not\prec \mathbf{x}_{\max}^j$, então $\exists \mathbf{y} \in \mathcal{F}_{j-1}$, ($\mathbf{y} \preceq \mathbf{x}_{\max}^j$). Logo, por monotonicidade de $P_{D_1}(\mathbf{x})$,

$$P\{\mathbf{y} \in \mathcal{F}_1\} \geq P\{\mathbf{x}_{\max}^j \in \mathcal{F}_1\}, \quad (5.15)$$

e como \mathbf{x}_{\max}^{j-1} é maximal em \mathcal{F}^{j-1} ,

$$P\{\mathbf{x}_{\max}^{j-1} \in \mathcal{F}_1\} \geq P\{\mathbf{y} \in \mathcal{F}_1\}. \quad (5.16)$$

Logo, por transitividade,

$$P\{\mathbf{x}_{\max}^{j-1} \in \mathcal{F}_1\} \geq P\{\mathbf{y} \in \mathcal{F}_1\} \geq P\{\mathbf{x}_{\max}^j \in \mathcal{F}_1\}. \quad (5.17)$$

□

Uma implicação do Teorema 1 é a indicação de que as soluções próximas do elemento maximal da primeira Frente, \mathcal{F}_1 , possuem maiores chances de serem não-dominadas na população. O Teorema também sugere que as PNDs aumentam à medida que se consideram regiões mais próximas à \mathcal{F}_1 e, portanto, mais próximas à PF . Visualizando-se a Figura 5.1, constata-se que as PNDs também crescem com a proximidade ao ponto de utopia. Portanto, conclui-se serem as NDLs uma ferramenta plausível para direcionar a exploração de regiões promissoras em MOEAs.

5.3 Geração de Diversidade Dirigida por NDLs

Nesta Seção, propõe-se a incorporação das NDLs em modelos de inserção de imigrantes em MOEAs, com o objetivo de gerar diversidade dirigida, capaz de promover exploração e/ou exploração do espaço de soluções que utilize informações sobre as relações de dominância da população atual. Para tanto, propõe-se um mecanismo o qual seja direto o suficiente de modo que o foco da investigação concentre-se nos efeitos da utilização pura e simples das Paisagens, sem recorrer a métodos de amostragem estatística sofisticados, cujos detalhes procedimentais poderiam vir a influenciar os resultados obtidos, além de implicarem em custos computacionais mais elevados. Assim, nesse contexto, incorporam-se as Paisagens em critérios de seleção cujo objetivo primordial é direcionar a inserção de imigrantes com maiores PNDs na população. Desta forma, espera-se atender ao requisito conflitante de manter diversidade em MOEAs sem prejudicar o caminhamento

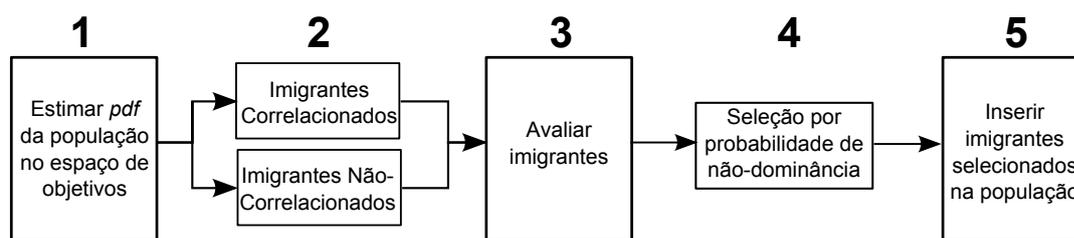


Figura 5.2 Diagrama de fluxo do mecanismo de geração de diversidade dirigida NDL-DG.

em direção à *PF*. Pelo contrário, procura-se mostrar que a proposta de geração de diversidade dirigida por NDLS é capaz de obter conjuntos de soluções não-dominadas (\mathcal{F}_1) de maior qualidade em termos de Hipervolume, quando comparadas aos conjuntos obtidos com mecanismos de diversidade sem a utilização das Paisagens.

Para a incorporação das NDLS aos geradores de diversidade, são necessárias as seguintes operações:

1. Estimação da *pdf* da população no espaço de objetivos – essa etapa é necessária para a computação das probabilidades de não-dominância dos imigrantes gerados, para que esses sejam inseridos na população conforme uma política que considere tais probabilidades;
2. Geração dos imigrantes – soluções atípicas devem ser geradas de forma a contribuir para o aumento dos níveis de diversidade da população. Dessa forma, duas estratégias são consideradas: (i) obter amostras aleatórias, a partir de uma *pdf* não-correlacionada com a distribuição atual da população no espaço de objetivos; ou (ii) modificar as soluções existentes na população de forma aleatória (e.g., por meio de mutações);
3. Avaliação dos imigrantes – os imigrantes gerados devem ser avaliados quanto à contribuição esperada para a obtenção de novas soluções não-dominadas;
4. Seleção dos imigrantes – com base nas PNDs avaliadas para cada imigrante gerado, um processo de seleção aponta os imigrantes que deverão ser inseridos;
5. Inserção dos imigrantes – como o tamanho da população é limitado, cada imigrante inserido deve substituir uma solução existente. Portanto, a operação de inserção considera uma política para gerenciar quais soluções devem ser descartadas para darem lugar aos novos imigrantes.

Pseudocódigo 3 Gerador de Diversidade Dirigida por Paisagens de Não-Dominância (NDL-DG)

- 1: Inicializar a subpopulação de imigrantes, Y_{NDL}
 - 2: Estimar a *pdf* da população no espaço de objetivos, $\hat{p}(\mathbf{x})$ (equação (2.63), página 7).
 - 3: **para** $i = 1$ até $\alpha \times N$ **faça**
 - 4: Inicializar lista temporária de imigrantes, Y .
 - 5: **para** $k = 1$ até K_{NDL} **faça**
 - 6: **se** Estratégia de imigrantes não-correlacionados **então**
 - 7: Gerar o imigrante \mathbf{y} a partir de uma *pdf* uniforme (equação (4.14), página 67).
 - 8: **senão se** Estratégia de imigrantes correlacionados **então**
 - 9: Gerar \mathbf{y} com base na elite da população utilizando o Pseudocódigo 4.
 - 10: **fim se**
 - 11: Avaliar o imigrante \mathbf{y} utilizando o Pseudocódigo 5.
 - 12: Armazenar \mathbf{y} na lista temporária de imigrantes, Y .
 - 13: **fim para**
 - 14: Selecionar \mathbf{y}_{NDL} tal que $P\{\mathbf{y}_{\text{NDL}} \in \mathcal{F}_1\} > P\{\mathbf{y} \in \mathcal{F}_1\}, \forall \mathbf{y} \in Y$.
 - 15: Inserir \mathbf{y}_{NDL} na subpopulação de imigrantes, Y_{NDL} .
 - 16: **fim para**
 - 17: Inserir os imigrantes em Y_{NDL} na população, Ψ (Pseudocódigo 6).
-

As operações descritas compõem o modelo de geração de diversidade dirigida, denominado Gerador de Diversidade por Paisagem de Não-Dominância (doravante NDL-DG, de *Non-Dominance Landscape Diversity Generator*), o qual é composto por cinco módulos básicos (Figura 5.3), cada qual correspondendo a uma das operações apresentadas: (1) estimacão da *pdf* da população no espaço de objetivos; (2) geração dos imigrantes; (3) avaliação dos imigrantes; (4) seleção por probabilidade de não-dominância; e (5) inserção dos imigrantes selecionados na população.

Ressalta-se que, os módulos (1) e (4) não constam nas estratégias de geração de diversidade por inserção de imigrantes convencionais, constituindo, dessa forma, o meio pelo qual as NDLs são incorporadas nos geradores de diversidade propostos. Outra diferença encontra-se no módulo (2), no qual a estratégia de geração de imigrantes correlacionados considera mutações da elite da população, estendendo as estratégias existentes de imigrantes baseados na elite para MOEAs. Ainda, a computação das PNDs dos imigrantes gerados é realizada no módulo (3), de acordo com a equação (5.10), constituindo outra novidade dos geradores propostos em relação às estratégias convencionais para AEs mono-objetivo, levantadas no Capítulo 4.

A proposta do NDL-DG é formalizada no Pseudocódigo 3, em que os parâmetros de entrada são a taxa de imigração (α) e o número de imigrantes gerados que concorrem entre si para serem inseridos na população (K_{NDL}). Dessa forma, sendo N o número de

5.3. GERAÇÃO DE DIVERSIDADE DIRIGIDA POR NDLS

Tabela 5.1 Quadro comparativo para as diferentes versões possíveis do NDL-DG.

Versão	K_{NDL}	Descrição
DG _R	1	Imigrantes aleatórios, sem seleção por PNDs.
DG _E	1	Imigrantes baseados na elite, sem seleção por PNDs.
NDL-DG _R	> 1	Imigrantes aleatórios, com seleção por PNDs.
NDL-DG _E	> 1	Imigrantes baseados na elite, com seleção por PNDs.

indivíduos na população, Ψ , o NDL-DG gera diversidade por meio da inserção de $\alpha \times N$ imigrantes em Ψ , em que $0 \leq \alpha \leq 1$. Note-se que, em se escolhendo $\alpha = 0$, o NDL-DG não insere imigrantes, enquanto para $\alpha = 1$, a população Ψ é completamente substituída pelos imigrantes gerados. O efeito da variação de K_{NDL} , por outro lado, não se dá sobre o número de imigrantes efetivamente inseridos em Ψ , mas sim sobre quantos imigrantes são gerados para cada imigrante inserido em Ψ . Assim, em se escolhendo $K_{\text{NDL}} = 1$, todo imigrante gerado será automaticamente inserido na população sem a consideração das PNDs. Portanto, deve-se escolher K_{NDL} inteiro maior que um para que as NDLS sejam efetivamente consideradas. Por exemplo, para $K_{\text{NDL}} = 2$, dois imigrantes são gerados e apenas aquele de maior PND é inserido em Ψ .

No Pseudocódigo 3, linha 2, é realizada a estimação não-paramétrica baseada em Kernel multivariado da *pdf* da população no espaço de objetivos, conforme descrito no Capítulo 2. Em seguida, os $\alpha \times N$ imigrantes a serem inseridos são escolhidos no procedimento compreendido pelas linhas 4 a 15. Para cada imigrante a ser inserido, K_{NDL} soluções atípicas são geradas e armazenadas em uma lista temporária, Y , entre as linhas 6 e 12. Na linha 15, as soluções em Y são selecionadas por PNDs, em que apenas aquela com a maior PND é escolhida para compor a subpopulação Y_{NDL} cujos elementos, na linha 17, são finalmente inseridos em Ψ . Ainda, cabe ressaltar que as duas formas possíveis de gerar imigrantes – não-correlacionados, linha 7, e correlacionados, linha 9 – levam a dois tipos de geradores distintos, denominados NDL-DG_R e NDL-DG_E, respectivamente. A Tabela 5.1 apresenta um quadro comparativo das diversas combinações possíveis de parâmetros do modelo e estratégias de geração de imigrantes para o NDL-DG. Note-se que os modelos descritos na Tabela são avaliados experimentalmente e suas performances são comparadas entre si no Capítulo 6.

Detalha-se a seguir cada módulo considerado no diagrama de fluxo da Figura 5.3, o qual descreve a operação geral do NDL-DG.

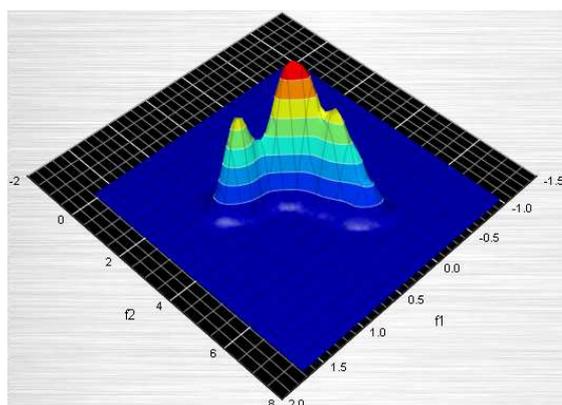


Figura 5.3 Visualização da *pdf* da população no espaço de objetivos estimada por Kernel Gaussiano multivariado nas primeiras gerações do NSGA2 no problema ZDT1.

5.3.1 Estimação da *pdf* da População no Espaço de Objetivos

Conforme discutido no início do Capítulo, a computação das probabilidades de não-dominância (PNDs) requer a utilização de estatísticas de ordem multivariada, as quais dependem do conhecimento da *pdf* que governa a distribuição das soluções candidatas no espaço de objetivos. No Pseudocódigo 3, esse processo é realizado na linha 2, utilizando a estimação por *kernel* multivariado da equação (2.63), página 7. Como resultado, obtêm-se a *pdf* estimada $\hat{p}(\mathbf{x})$. A escolha da estimação baseada em *kernel* dá-se principalmente por ser desnecessário supor modelos iniciais de *pdfs*, os quais poderiam levar a aproximações pobres da verdadeira distribuição dos vetores no espaço de objetivos (ver justificativas completas na página 22). Embora diversas escolhas sejam possíveis, $\hat{p}(\mathbf{x})$ é estimada no NDL-DG por meio de *kernels* Gaussianos multivariados, com matriz de suavização proporcional à raiz quadrada da matriz de covariância amostral (ver Seção 2.3.4, página 22), objetivando-se capturar as dependências estatísticas entre os valores das funções-objetivo. No entanto, cabe ressaltar que a escolha da função *kernel* na estimação não é tão importante quanto a influência da matriz de suavização, \mathbf{H} . Em testes preliminares, constatou-se que, em se escolhendo $\mathbf{H} \propto \hat{\Sigma}^{1/2}$, visualizam-se *pdfs* consoantes com a distribuição de vetores no espaço de objetivos (Figura 5.3).

Na implementação do NDL-DG realizada, utilizando a linguagem de programação C++, as *pdfs* estimadas são armazenadas em estruturas de dados na forma de matrizes de pontos flutuantes, 20×20 , em que cada entrada corresponde à densidade de probabilidade avaliada em um ponto escolhido com espalhamento uniforme em um *grid* de tamanho 20×20 no espaço de objetivos. Nesse caso, observou-se em testes preliminares que a

granularidade da representação de $\hat{p}(\mathbf{x})$ na memória, que depende do tamanho do *grid*, não é tão importante para a computação das PNDs: apesar de quanto maior o *grid*, mais precisa a representação de $\hat{p}(\mathbf{x})$ na memória, os erros de estimação para a escolha 20×20 são irrelevantes ao se compararem as PNDs computadas para duas soluções arbitrárias. Dessa forma, a escolha do tamanho dos *grids* justifica-se, principalmente, pelos relativos baixos custo de armazenamento em memória e de processamento. A representação da *pdf* na memória é então delimitada, para cada dimensão, pelos valores máximos e mínimos das funções-objetivo na população atual, respectivamente. De posse de $\hat{p}(\mathbf{x})$, e, uma vez que o número de objetivos é pequeno nos MOPs avaliados nessa dissertação (dois), calculam-se as integrais correspondentes às etapas da computação das PNDs por regras de quadratura convencionais, i.e., métodos numéricos como a regra do trapézio para integrais duplas. Ressalta-se, entretanto, que há diversas possibilidades de implementação para as etapas de estimação da *pdf* e de integração numérica, com diferentes graus de complexidade e precisão. Por exemplo, um método alternativo seria o cálculo das PNDs por amostragem Monte-Carlo, o que manteria o incremento no número necessário de operações polinomial em relação ao aumento do número de objetivos.

5.3.2 Geração dos Imigrantes em MOEAs

O modelo NDL-DG proposto pode incorporar diferentes paradigmas de geração de imigrantes, conforme a taxonomia discutida no Capítulo 4, página 64. Nessa dissertação, duas versões do NDL-DG são propostas e avaliadas experimentalmente no Capítulo 6. Na primeira, a geração de imigrantes não-correlacionados é utilizada da mesma forma como no algoritmo RIGA (página 67), em que os imigrantes são amostrados a partir de uma *pdf* uniforme multidimensional (linha 7 do Pseudocódigo 3). Isso significa que os imigrantes gerados são aleatoriamente distribuídos sobre o espaço de decisão. No quadro da Tabela 5.1, essa estratégia equivale aos modelos DG_R e $NDL-DG_R$, dependendo da escolha de K_{NDL} . Já a segunda versão utiliza a geração de imigrantes baseados na elite da população, semelhantemente ao AE mono-objetivo *Elitism-based Immigrants GA* (EIGA) investigado em Yang (2008), em que os imigrantes são gerados a partir de mutações dos melhores indivíduos da população. Nesse caso, duas estratégias são possíveis: DG_E e $NDL-DG_E$ (Tabela 5.1).

No $NDL-DG_E$, por conta da natureza multiobjetiva dos problemas para os quais o modelo deve ser aplicado, o conceito de elite (i.e., melhores indivíduos) precisa ser bem definido. Nessa dissertação, adota-se o operador de ordenação \geq_{ND} empregado pelo algoritmo NSGA2 (ver página 44), baseado em não-dominância e desempate por

Pseudocódigo 4 Gerar Imigrantes Baseados na Elite

- 1: Ordenar Ψ pelo critério de não-dominância, \succeq_{ND} (página 42).
 - 2: **se** $K \geq |\mathcal{F}_1|$ **então**
 - 3: Selecionar os K primeiros elementos de Ψ para compor a Elite.
 - 4: **senão**
 - 5: Tomar o conjunto de não-dominadas, \mathcal{F}_1 , para compor a Elite.
 - 6: **fim se**
 - 7: Inicializa a lista temporária de imigrantes, Y .
 - 8: **para** $k = 0$ até K **faça**
 - 9: Selecionar aleatoriamente \mathbf{x} da Elite, com reposição.
 - 10: Gerar imigrante \mathbf{y} por meio da mutação de \mathbf{x} .
 - 11: Inserir \mathbf{y} na lista Y .
 - 12: **fim para**
 - 13: **retorne** os K imigrantes gerados em Y .
-

estimação de *crowding distance* (CD). Suponha que se deseje obter os K melhores indivíduos em relação a um espaço de objetivos. No Pseudocódigo 4, primeiro, ordena-se a população utilizando a ordem parcial \succeq_{ND} (linha 1). Depois, na linha 3, se $K \geq |\mathcal{F}_1|$, seleciona-se a elite como os K primeiros indivíduos da população ordenada ($|\mathcal{F}_1|$ é o número de soluções não-dominadas obtidas na população). Já no caso em que $K < |\mathcal{F}_1|$, todas as soluções não-dominadas são selecionadas para compor a elite (linha 5). Finalmente, escolhida a elite, geram-se os imigrante a partir de mutações de elementos da elite selecionados aleatoriamente, com reposição (linhas 9 a 11), o que implica que cada membro da elite pode servir de base para a geração de um ou mais imigrantes. Note-se que o número de imigrantes a serem gerados (K) é fornecido como parâmetro de entrada.

Justifica-se a escolha para $K < |\mathcal{F}_1|$ pelo desejo de dar chances iguais às soluções não-dominadas de serem selecionadas para a geração de imigrantes, o que reduz a pressão de seleção induzida pela atribuição de CD. Essa decisão é motivada pelos resultados apresentados em Ishibuchi e Narukawa (2005), nos quais o aumento do tamanho dos torneios na seleção para reprodução no NSGA2 leva ao aumento de diversidade e degrada os valores obtidos para os indicadores de proximidade à *PF*. Embora os autores tenham classificado esse resultado como “*contra-intuitivo*”, inspeções adicionais dos resultados experimentais relatados no Capítulo 4 com o NSGA2 mostram que, por volta da setuagésima geração, todas as soluções na população tendem a ser não-dominadas, o que faz com que as seleções para reprodução e de sobreviventes baseiem-se majoritariamente nos valores de CD. Logo, como a seleção por CD promove diversidade no espaço de objetivos, não é desejável escolher a elite para a geração de imigrantes primordialmente com base nesse critério, como ocorreria para $K < |\mathcal{F}_1|$ se fosse seguida a escolha ordinária descrita

Pseudocódigo 5 Avaliar Imigrante por PND

-
- 1: **para** $i = 1$ até M **faça**
 - 2: Computar $v = f_i(\mathbf{y})$ e armazenar v no atributo $\mathbf{y}.f_i$;
 - 3: **fim para**
 - 4: Computar $p_{ND} = P_{D_1}(\mathbf{y})$ (equação (5.10)) e armazenar p_{ND} no atributo $\mathbf{y}.p_{ND}$.
 - 5: **retorne** o imigrante \mathbf{y} avaliado.
-

para $K \geq |\mathcal{F}_1|$, já que isso concentraria a geração de imigrantes exclusivamente em torno das regiões em que há baixa densidade de soluções no espaço de objetivos.

Note-se que, tanto no NDL-DG_R quanto no NDL-DG_E, são gerados $\alpha \times N \times K_{\text{NDL}}$ imigrantes a cada geração. Desses, $\alpha \times N$ são efetivamente inseridos na população. Note-se que $K_{\text{NDL}} \in \mathcal{N}^+$ controla a pressão de seleção por PND, ou seja, quanto maior o valor de K_{NDL} , maior o viés da inserção de imigrantes pela utilização de NDLS e vice-versa.

5.3.3 Avaliação dos Imigrantes

No Pseudocódigo 3, o procedimento de avaliação da subpopulação Y_{NDL} de imigrantes é realizado na linha 11, e consiste de duas etapas para cada imigrante (Pseudocódigo 5): primeiro, computam-se os valores de cada função-objetivo, f_i , com M denotando o número de objetivos do problema em consideração (linha 2). Então, a probabilidade de não-dominância em relação à população atual é computada, avaliando-se o imigrante \mathbf{y} pela equação (5.10), de modo que se associa o atributo $\mathbf{y}.p_{\text{ND}}$ a \mathbf{y} (linha 4). O procedimento então retorna o imigrante \mathbf{y} avaliado. Ressalta-se que essa é uma etapa importante, pois permite a seleção dos imigrantes gerados por PNDs.

5.3.4 Seleção por Probabilidade de Não-Dominância (PND)

No Pseudocódigo 3, a etapa de seleção de imigrantes por PNDs é realizada na linha 14, escolhendo-se \mathbf{y}_{NDL} tal que $P\{\mathbf{y}_{\text{NDL}} \in \mathcal{F}_1\} > P\{\mathbf{y} \in \mathcal{F}_1\}$, $\forall \mathbf{y} \in Y$, i.e., tomando as soluções atípicas geradas com maiores chances de serem não-dominados na população atual, Ψ . A escolha do valor K_{NDL} controla a pressão de seleção por PNDs no NDL-DG para a geração de diversidade. Para $K_{\text{NDL}} = 1$, a NDL não é efetivamente utilizada, e todos os imigrantes gerados são inseridos na população atual sem a consideração das probabilidades de não-dominância. Como exemplo, suponha que $N = 100$, $\alpha = 0,2$ e $K_{\text{NDL}} = 2$. Então, $\alpha \times N \times K_{\text{NDL}} = 40$ soluções atípicas são geradas, dentre as quais as $\alpha \times N = 20$ soluções com maiores probabilidade de não-dominância são efetivamente inseridas na subpopulação de imigrantes, Y_{NDL} . Dessa forma, o incremento de K_{NDL}

Pseudocódigo 6 Inserir Imigrantes

- 1: **para** $i = 1$ até $|Y_{\text{NDL}}|$ **faça**
 - 2: Descartar o i -ésimo pior indivíduo da população ordenada, Ψ .
 - 3: Inserir o imigrante $y_i \in Y_{\text{NDL}}$ em Ψ .
 - 4: **fim para**
-

aumenta o viés da inserção de imigrantes por PNDs na população, Ψ . Ressalta-se, no entanto, que essa não é a única estratégia possível para selecionar imigrantes por PNDs (ver Capítulo 7 para uma discussão acerca de outras possibilidades), mas é certamente a mais direta possível para o propósito de avaliar o efeito da influência de seleção de imigrantes por PNDs, sem recorrer a métodos de amostragem estatística sofisticados.

5.3.5 Estratégia de Inserção dos Imigrantes na População

Uma vez selecionada a subpopulação de imigrantes Y_{NDL} , resta escolher quais indivíduos serão substituídos na população Ψ atual. Essa é uma decisão importante para a qual, da mesma forma que nos outros módulos do NDL-DG, diversas estratégias são possíveis. Assim, com o propósito de manter todos os processos realizados por cada módulo os mais diretos possíveis, opta-se por descartar os $|Y_{\text{NDL}}| = \alpha \times N$ piores indivíduos de Ψ , para darem lugar aos novos imigrantes. Para tanto, no Pseudocódigo 6, com a Ψ já ordenada pelo operador de não-dominância \geq_{ND} , descartam-se os $\alpha \times N$ últimos indivíduos presentes nas Frentes inferiores e com menores valores de *crowding distance* (CD). Cabe ressaltar que, se porventura todos os indivíduos na população forem não-dominados, i.e., se $|\mathcal{F}_1| = N$, então, devido à ordenação de Ψ , os $\alpha \times N$ indivíduos com os menores valores de CD associados (ou seja, aqueles presentes em regiões mais densas no espaço de objetivos) são descartados para darem entrada aos novos imigrantes. Note-se que, nesse caso, embora seja possível que soluções não-dominadas sejam descartadas, por conta do NDL-DG considerar a inserção dos imigrantes com maiores probabilidade de serem não-dominados na população, espera-se que a inserção de imigrantes, como descrita, contribua para o aumento dos níveis de diversidade da população de soluções candidatas, sem comprometer muito a qualidade de \mathcal{F}_1 .

Resultados de validação experimental demonstrando o potencial da seleção baseada em PNDs para a obtenção de soluções não-dominadas são descritos na próxima Seção.

5.4 Validação do Modelo Proposto

Para validar o NDL-DG como um modelo capaz de gerar diversidade genética e de, ao mesmo tempo, promover o caminhamento em direção à PF , apresentam-se resultados de simulação envolvendo a geração de imigrantes não-correlacionados no problema multiobjetivo ZDT1, utilizando o NDL-DG_R. Nas simulações, escolheu-se a taxa de imigração $\alpha = 1$, de modo que a população é reinicializada a cada geração com $\alpha \times N = 100$ imigrantes gerados.

Dos resultados obtidos, três aspectos são analisados: (i) o desvio da pdf estimada da subpopulação de imigrantes no espaço de objetivos – $\hat{p}_{\text{imig}}(\mathbf{x})$ – em relação a uma pdf de referência – $\hat{p}_{\text{ref}}(\mathbf{x})$ –, também estimada no espaço de objetivos, sobre soluções geradas uniformemente no espaço de decisão com o DG_R (equivalente a $K_{\text{NDL}} = 1$); (ii) indicadores de qualidade e de diversidade das populações de imigrantes geradas; e (iii) a capacidade do NDL-DG de produzir imigrantes mutuamente não-dominados. O objetivo da análise é responder as seguintes perguntas:

1. Seria a amostragem realizada pelo NDL-DG_R aleatória, isto é, equivalente àquela realizada pelo DG_R?
2. Há ganhos qualitativos para os imigrantes gerados com o NDL-DG_R em relação àqueles gerados com o DG_R?
3. Obtêm-se imigrantes não-dominados mais facilmente com o NDL-DG_R, quando comparado com os imigrantes gerados pelo DG_R?

Primeiro, busca-se avaliar se o processo de geração de imigrantes no NDL-DG_R difere significativamente de uma amostragem aleatória no espaço de soluções. Para tanto, utiliza-se o procedimento descrito no Pseudocódigo 3 para a geração de 100 imigrantes selecionados por PNDs. Essa etapa de validação é importante porque permite quantificar as diferenças observadas na distribuição de imigrantes gerados pelo NDL-DG_R em relação a estratégias as quais não utilizam Paisagens de Não-Dominância (DG_R). A metodologia para essa etapa consiste em mensurar a divergência de Kullback-Leibler (D_{KL}) entre as pdf s $\hat{p}_{\text{imig}}(\mathbf{x})$ e $\hat{p}_{\text{ref}}(\mathbf{x})$ sobre diferentes níveis de pressão de seleção por PNDs, atribuindo-se ao parâmetro K_{NDL} os valores $\{1, 3, 6, 9\}$. Utiliza-se a notação NDL-DG_R(K_{NDL}) para fazer referência às diferentes parametrizações testadas, em que $\text{DG}_R \equiv \text{NDL-DG}_R(1)$.

Tabela 5.2 p -values computados para as divergências de Kullback-Leibler em função de K_{NDL} .

K_{NDL}	1	3	6	9
1	–	–	–	–
3	0,23	–	–	–
6	$3,04 \times 10^{-8}$	$1,16 \times 10^{-5}$	–	–
9	$2,21 \times 10^{-12}$	$1,96 \times 10^{-10}$	$1,72 \times 10^{-4}$	–

Define-se a D_{KL} entre duas pdf s p e q como (Kullback e Leibler, 1951)

$$\begin{aligned}
 D_{\text{KL}}(p||q) &= \int_{-\infty}^{+\infty} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} \\
 &= \int_{-\infty}^{+\infty} p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x} - \int_{-\infty}^{+\infty} p(\mathbf{x}) \log q(\mathbf{x}) d\mathbf{x}.
 \end{aligned} \tag{5.18}$$

A divergência de Kullback-Leibler foi primeiramente introduzida no contexto da teoria da informação para mensurar o número esperado de *bits* adicionais requeridos na codificação de um sinal gerado por uma v.a. com pdf $p(\mathbf{x})$ usando um código baseado em uma segunda v.a. aleatória com pdf $q(\mathbf{x})$ (Kullback e Leibler, 1951). Em probabilidade e estatística, $p(\mathbf{x})$ é geralmente um modelo teórico de referência e $q(\mathbf{x})$ é uma aproximação para $p(\mathbf{x})$. Note-se que D_{KL} não é matematicamente uma métrica, pois $D_{\text{KL}}(p||q) \neq D_{\text{KL}}(q||p)$, além de não ser válida a desigualdade triangular. Apesar disso, $D_{\text{KL}}(p||q) \geq 0$, com a igualdade ocorrendo se, e somente se, $p(\mathbf{x}) = q(\mathbf{x}), \forall \mathbf{x}$. Portanto, a equação (5.18) serve como uma medida adequada para o desvio entre duas pdf s arbitrárias. Observam-se os seguintes valores médios para $D_{\text{KL}}(\hat{p}_{\text{ref}}||\hat{p}_{\text{imig}})$:

- Para $K_{\text{NDL}} = 1$, $D_{\text{KL}}(\hat{p}_{\text{ref}}||\hat{p}_{\text{imig}}) = 2,30 \pm 0,79$;
- Para $K_{\text{NDL}} = 3$, $D_{\text{KL}}(\hat{p}_{\text{ref}}||\hat{p}_{\text{imig}}) = 2,43 \pm 0,81$;
- Para $K_{\text{NDL}} = 6$, $D_{\text{KL}}(\hat{p}_{\text{ref}}||\hat{p}_{\text{imig}}) = 2,96 \pm 0,82$;
- Para $K_{\text{NDL}} = 9$, $D_{\text{KL}}(\hat{p}_{\text{ref}}||\hat{p}_{\text{imig}}) = 3,63 \pm 1,55$.

A distribuição dos valores D_{KL} obtida em cada parametrização é melhor visualizada no *box plot* da Figura 5.4 (a).

A Tabela 5.2 apresenta os p -values computados para os valores de D_{KL} observados em 100 repetições. Da tabela, conclui-se que, com exceção do par NDL-DG_R(1)/NDL-DG_R(3), para o qual o p -value é 0,23, todas as médias de divergência de Kullback-Leibler diferem entre si com significância estatística a nível de confiança de 0,95, segundo o Teste-T de Welch (Welch, 1947). Portanto, a partir dos resultados apresentados, demonstra-se que o processo de geração de diversidade dirigida por NDLs proposto não é fruto

simplesmente de uma amostragem aleatória do espaço de decisão, pelo contrário: a divergência para o modelo de referência uniforme cresce com o aumento da pressão de seleção por probabilidade de não-dominância, controlada pelo parâmetro K_{NDL} .

Para responder a segunda pergunta, recorre-se ao indicador de Hipervolume (equação (3.9), página 32) e à diversidade genética da população (equação (4.2), página 51). As partes (b) e (c) da Figura 5.4 descrevem os resultados obtidos para Hipervolume e diversidade genética, respectivamente. Observa-se que o Hipervolume mediano das populações geradas pelo NDL-DG aumenta em conformidade com valores maiores de K_{NDL} . Ademais, testes estatísticos confirmam a significância das diferenças entre as médias de Hipervolume para cada combinação dois a dois de K_{NDL} com nível de confiança 0,95. Já a parte (c) da figura mostra que os níveis medianos de diversidade genética obtidos para as diferentes parametrizações NDL-DG(K_{NDL}) se mantêm aproximadamente constantes.

Apesar dessa observação, apenas as médias de diversidade entre o par NDL-DG(3)/NDL-DG(9) não diferem com significância estatística, de modo que se pode afirmar que há uma leve diminuição de diversidade genética ao se utilizarem as Paisagens de Não-Dominância para guiar a geração de diversidade. Dessa análise, conclui-se que, de fato, o NDL-DG é capaz de gerar diversidade a qual contribui para a melhoria na qualidade da subpopulação de imigrantes, aumentando, portanto, as chances de que os imigrantes sobrevivam na população em gerações subsequentes, sem comprometer muito os níveis de diversidade genética obtidos com o modelo proposto.

Finalmente, responde-se a terceira pergunta por meio da análise do número de imigrantes mutuamente não-dominados gerados, $|\mathcal{F}_1|$, por cada parametrização NDL-DG_R(K_{NDL}). Os *box plots* da Figura 5.4 (d) mostram o crescimento do número mediano de imigrantes não-dominados obtidos com o aumento de K_{NDL} , em que se atesta a significância estatística entre as médias para todas as combinações possíveis a nível 0,95 de confiança. Observa-se um aumento substancial de $\approx 100\%$ no número médio de imigrantes não-dominados obtidos com o NDL-DG_R(9) (24,67), quando comparado ao valor médio obtido (≈ 12) sem a influência das NDLs, quando $K_{\text{NDL}} = 1$.

Uma observação pertinente para a análise qualitativa de Hipervolume e diversidade genética da população é a de que os resultados descritos na Figura 5.4 foram obtidos para a dimensionalidade $m = 2$ no espaço de decisão. Cumpre mencionar que essa escolha se deu por conta das características dos problemas do conjunto ZDT, os quais computam os valores para o primeiro objetivo em função apenas da primeira componente dos vetores de soluções, enquanto que os valores para o segundo objetivo são computados em função das m componentes. Desse modo, para altas dimensões, otimizar o segundo objetivo

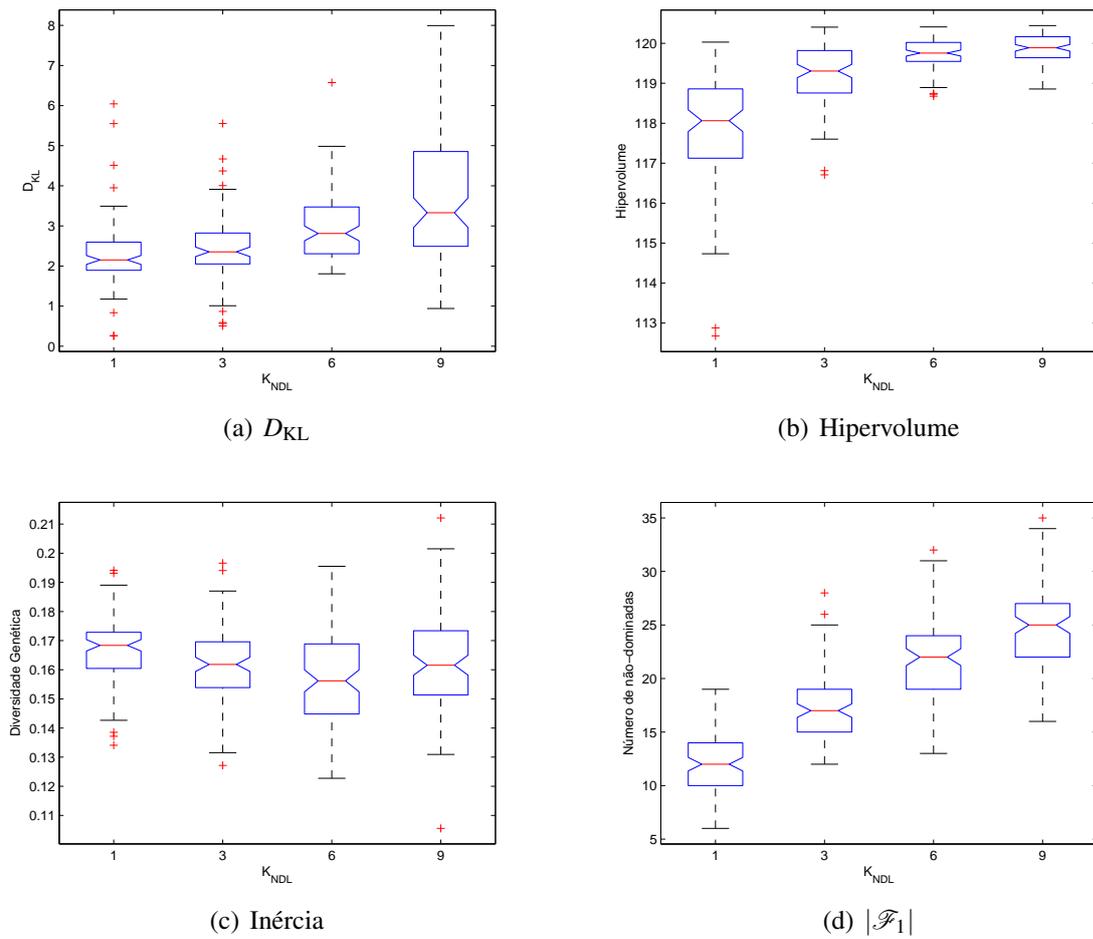


Figura 5.4 Box plots para análise qualitativa dos imigrantes gerados pelo NDL-DG_R.

torna-se muito mais complexo do que otimizar o primeiro: basta, por exemplo, que uma mutação bem-sucedida altere o primeiro gene de um indivíduo para um valor próximo de zero para que o mutante possua alta probabilidade de ser não-dominado. Em outras palavras, é muito mais fácil obter soluções não-dominadas por meio da alteração do primeiro gene do que pela alteração parcial dos demais $m - 1$ genes.

Em simulações preliminares (cujos resultados não são relatados), devido a essa característica dos problemas ZDT em altas dimensionalidades, observou-se que o uso intensivo de NDLs não é apropriado nas primeiras gerações de um MOEA projetado para otimizar o conjunto ZDT. Uma vez que o objetivo das simulações nesse Capítulo é de validar o NDL-DG, escolheu-se $m = 2$ para a validação apresentada objetivando-se a minimização dos efeitos do desbalanceamento entre as funções-objetivo, pois, desse modo, o número de variáveis que devem ser ajustadas para a otimização é distribuída da forma mais justa possível entre as duas funções.

5.5 Discussão Final e Síntese

O presente Capítulo introduziu contribuições originais para a geração de diversidade em algoritmos baseados em população aplicados à otimização multiobjetivo. Primeiro, derivaram-se expressões as quais refletem as probabilidades de dominância e comparabilidade entre soluções em um espaço de objetivos multidimensional arbitrário. Utilizaram-se tais expressões para computar probabilidades de Frentes na população, em especial, a probabilidade de uma solução ser não-dominada (i.e., pertencer à Frente \mathcal{F}_1) e a probabilidade da ocorrência de uma ordenação total das soluções no espaço de objetivos (i.e., a probabilidade de $\mathcal{F}_N \neq \emptyset$). Essas expressões levam à definição das Paisagens de Não-Dominância (NDLs), as quais foram incorporadas na proposta de geradores de diversidade dirigida, denominados NDL-DG. Por meio do Teorema 1, demonstrou-se a plausibilidade do uso das NDLs para dirigir o processo de inserção de imigrantes em populações de soluções candidatas, de modo a gerar diversidade genética ao mesmo tempo em que se promove o caminhamento em direção à PF por meio da exploração de regiões promissoras. Subdividiu-se, então, o NDL-DG em termos do processo de geração de imigrantes, originando dois modelos específicos: o NDL-DG_R, o qual considera imigrantes aleatórios não-correlacionados, e o NDL-DG_E, que promove a geração de imigrantes correlacionados com base em uma elite definida pela ordem parcial de Pareto.

Avaliaram-se os potenciais benefícios da aplicação do modelo NDL-DG à *geração de imigrantes em MOEAs* – assunto pouco estudado na literatura – por meio de vali-

dação experimental. Com o aumento da pressão de seleção de imigrantes baseada em probabilidade de não-dominância, controlada pelo parâmetro K_{NDL} , mostrou-se: (i) que a amostragem realizada pelo gerador NDL-DG_R diverge de um modelo de referência uniforme no espaço de decisão; (ii) que há ganhos qualitativos em termos do indicador de Hipervolume nas populações de imigrantes geradas pelo NDL-DG_R em comparação àquelas geradas pelo DG_R; e (iii) que se obtêm imigrantes não-dominados com maior frequência com o uso das NDLs. Portanto, espera-se que, com a incorporação do NDL-DG em MOEAs, esses se tornem robustos o suficiente para serem utilizados em situações em que o aumento e/ou manutenção de níveis mínimos de diversidade seja um requisito importante, como na otimização dinâmica multiobjetivo.

No próximo Capítulo, incorporam-se os geradores NDL-DG propostos ao NSGA2 e avalia-se experimentalmente a adequabilidade dos mesmos para a geração de diversidade na otimização multiobjetivo. O NDL-DG é especialmente avaliado em problemas de alta dimensionalidade. Cabe ressaltar que a escolha do NSGA2 para a avaliação experimental das variantes do NDL-DG é justificada na discussão final do Capítulo 3.

6

Resultados Experimentais

A esmagadora maioria das teorias são rejeitadas porque contêm explicações ruins, não porque falham em testes experimentais.

– David Deustch

Busca-se nesse Capítulo obter uma prova de princípio a qual demonstre a viabilidade do modelo proposto de geração de diversidade baseado em Paisagens de Não-Dominância (NDLs), denominado NDL-DG, para o problema conflitante de gerar diversidade à medida que se caminha em direção à Frente de Pareto (*PF*). Para tanto, avaliam-se duas versões diferentes da incorporação do NDL-DG ao NSGA2, em relação ao método de geração de imigrantes: (i) não-correlacionados (NDL-DG_R) – em que os imigrantes são gerados a partir de uma *pdf* distinta e independente daquela da subpopulação de trabalho (ver página 64)– e (ii) correlacionados (NDL-DG_E) – em que o processo de geração de imigrantes depende da distribuição da subpopulação de trabalho. Ambos, NDL-DG_R e NDL-DG_E, são aplicados e avaliados em cenários diversos da otimização contínua multiobjetivo, tanto em ambientes estáticos quanto em dinâmicos.

6.1 Metodologia e Arranjo Experimental

Os experimentos relatados nesse capítulo se dividem em duas partes. Primeiro, uma vez que, até onde se sabe, a inserção de imigrantes – como definida em e.g. Yang (2008) – não fora estudada sistematicamente na otimização multiobjetivo¹, avaliam-se as estratégias convencionais de inserção de imigrantes não-correlacionados e correlacionados sem o uso de paisagens de não-dominância sob diferentes taxas de imigração, tanto em ambientes

¹Existem, no entanto, exemplos de aplicação, bem como estudos relacionados como em Deb *et al.* (2007).

Pseudocódigo 7 NSGA2 com Geração de Diversidade

```

1: // Entrada:  $K$ , o tamanho da subpopulação de imigrantes
2: InicializarPopulação( $\Psi^{(0)}$ ),  $g \leftarrow 0$ 
3: Ordenar( $\Psi^{(0)}, \geq_{ND}$ ) // Ordena  $\Psi^{(0)}$  pelo critério de não-dominância  $\geq_{ND}$ 
4: // Aplicação dos operadores de variação e de seleção de sobreviventes
5: enquanto condição de término não satisfeita faça
6:    $Q^{(g+1)} \leftarrow$  OperadoresGenéticos( $\Psi^{(g+1)}$ )
7:   NDL-DG( $\Psi^{(g)}, K$ ) // Gera a subpopulação de imigrantes e os insere na população
8:    $R^{(g)} \leftarrow \Psi^{(g)} \cup Q^{(g)}$  //  $Q^{(g)}$  é a subpopulação de filhos e  $\Psi^{(g)}$  é a população atual.
9:    $\mathcal{F} \leftarrow$  ConstruirFrentes( $R^{(g)}$ )
10:   $\Psi^{(g+1)} \leftarrow \emptyset$ ,  $i \leftarrow 0$ 
11:  enquanto  $|\Psi^{(g+1)}| < N$  faça
12:    AtribuirCrowdingDistance( $\mathcal{F}_i$ )
13:     $\Psi^{(g+1)} \leftarrow \Psi^{(g+1)} \cup \mathcal{F}_i$ 
14:     $i \leftarrow i + 1$ 
15:  fim enquanto
16:  Ordenar( $\Psi^{(g+1)}, \geq_{ND}$ )
17:   $\Psi^{(g+1)} \leftarrow \Psi^{(g+1)}[0 : N]$  // Os primeiros  $N$  elementos de  $\Psi^{(g+1)}$  sobrevivem
18:   $g \leftarrow g + 1$ 
19: fim enquanto
20: retorne  $\Psi^{(g)}$ 

```

estáticos quanto em dinâmicos. Nesse estudo, varia-se a taxa de imigração α no conjunto discreto $\{0, 0; 0, 3; 0, 5; 0, 7; 1, 0\}$, em que $\alpha = 0, 0$ corresponde à ausência de geração de imigrantes e o processo convencional realizado pelo MOEA é utilizado. Já para $\alpha = 1, 0$, a cada geração, tem-se que a população é completamente reinicializada.

Na segunda parte dos experimentos, incorporam-se separadamente os geradores NDL-DG_R e NDL-DG_E ao MOEA estudado e busca-se avaliar o efeito do aumento da pressão de seleção dos imigrantes por probabilidade de não-dominância por meio da variação sistemática de K_{NDL} no conjunto discreto $\{1, 3, 6, 9\}$, em que $K_{NDL} = 1$ corresponde à ausência das NDLs no processo de seleção de imigrantes.

6.1.1 Incorporação do NDL-DG ao NSGA2

Escolheu-se incorporar os geradores de diversidade propostos ao algoritmo NSGA2 para os experimentos relatados nesse Capítulo. Dessa forma, têm-se quatro versões modificadas do NSGA2:

- **NSGA2(R)** – corresponde à versão que incorpora a geração de imigrantes não-correlacionados pela amostragem uniforme do espaço de decisão (DG_R);

- **NSGA2(E)** – corresponde à versão que incorpora a geração de imigrantes correlacionados com base na elite da população $(DG_R)^2$;
- **NSGA2(R_{NDL})** – corresponde à incorporação do NDL-DG_R ao NSGA2;
- **NSGA2(E_{NDL})** – corresponde à incorporação do NDL-DG_E ao NSGA2.

O arcabouço geral da incorporação de geradores de diversidade ao NSGA2 pode ser visualizado no Pseudocódigo 7. Destaca-se, na linha 7, o ponto em que o NDL-DG é incorporado ao NSGA2: após a etapa de reprodução, em que os operadores genéticos são aplicados (linha 6), e antes da etapa de seleção de sobreviventes. Essa escolha é condizente com o ponto em que os geradores de diversidade empregados em AEs mono-objetivo são incorporados (ver Pseudocódigo 2, página 67). Note-se que, em se escolhendo a taxa de imigração como $\alpha = 0,0$, todas as quatro versões equivalem ao NSGA2 padrão (Deb *et al.*, 2000). Nas versões que incorporam o NDL-DG, na escolha de $K_{NDL} = 1$ têm-se que: (i) o NSGA2(R_{NDL}) equivale ao NSGA2(R); e (ii) o NSGA2(E_{NDL}) equivale ao NSGA2(E).

Finalmente, utiliza-se a parametrização padrão do NSGA2 em todas as versões testadas nesse Capítulo, a qual é encontrada na Tabela 4.2, página 57, com os parâmetros de espalhamento relativos aos operadores de variação escolhidos como $\eta_{SBX} = \eta_{mut} = 20$. Ressalta-se que a escolha para os parâmetros η_{SBX} e η_{mut} baseou-se na observação dos resultados do Capítulo 4, em que $\eta_{SBX} = \eta_{mut} = 20$ mostrou-se a combinação a qual leva aos menores valores de diversidade genética acumulada em todos os cenários de otimização avaliados com o algoritmo NSGA2. Desse modo, espera-se que os níveis de diversidade observados nas populações evoluídas com as versões do NSGA2 avaliadas nesse Capítulo reflitam predominantemente a diversidade produzida pelos geradores de imigrantes propostos, reduzindo ao máximo o efeito da diversidade gerada pelos operadores de variação comumente utilizados no NSGA2.

Um quadro descritivo de todas as versões do NSGA2 estudadas nesse Capítulo encontra-se na Tabela 6.1.

6.1.2 Problemas de Otimização Multiobjetivo Selecionados

Escolheu-se o conjunto ZDT composto de cinco problemas de otimização multiobjetivo (MOPs) para os experimentos com ambientes estáticos. As dimensionalidades do espaço

²Ver definição de elite em um MOEA na página 85.

Tabela 6.1 Quadro comparativo dos algoritmos considerados para os experimentos.

Algoritmo	Valor de α	Valores de K_{NDL}	Descrição
NSGA2	$= 0,0$	--	Corresponde à implementação padrão em Deb <i>et al.</i> (2000).
NSGA2(R)	$> 0,0$	$= 1$	Corresponde à incorporação do DG_R . Não utiliza Paisagens de Não-Dominância.
NSGA2(E)	$> 0,0$	$= 1$	Corresponde à incorporação do DG_E . Não utiliza Paisagens de Não-Dominância.
NSGA2(R_{NDL})	$> 0,0$	> 1	Corresponde à incorporação do $NDL-DG_R$. Seleciona os imigrantes por PNDs.
NSGA2(E_{NDL})	$> 0,0$	> 1	Corresponde à incorporação do $NDL-DG_E$. Seleciona os imigrantes por PNDs.

de decisão variam em relação ao problema, sendo $m = 10$ para o ZDT4 e $m = 30$ para os demais problemas do conjunto.

Para os experimentos com ambientes dinâmicos, utilizam-se os dois primeiros problemas do conjunto FDA – com modificações quanto ao número de variáveis consideradas para cada objetivo – e propõe-se um novo MOP dinâmico baseado em um problema apresentado em Goh e Tan (2009). As descrições dos conjuntos de problemas ZDT e FDA encontram-se no Apêndice B.

Portanto, são estes os MOPs dinâmicos utilizados para a investigação experimental:

- FDA1 – consiste em um problema Tipo 1 (ver Capítulo 3, Seção 3.5.2) com $PF(t)$ convexa e fixa para qualquer valor de t . Na versão implementada para os experimentos, escolheram-se $X_I = \{x_1, \dots, x_{10}\} \in [0, 1]$ e $X_{II} = \{x_{11}, \dots, x_{20}\} \in [-1, 1]$, em que o primeiro objetivo é computado por $f_1(\mathbf{x}) = \sum_{m=1}^{10} x_m$ e as demais funções são computadas normalmente como em Farina *et al.* (2004);
- FDA2 – consiste em um problema Tipo 3 em que a $PF(t)$ muda progressivamente com o estado temporal de um formato convexo, para um formato côncavo. Note-se que o conjunto ótimo de Pareto no espaço de decisão, Ω^* , entretanto, não muda. Escolheram-se os conjuntos $X_I = \{x_1, \dots, x_7\} \in [0, 1]$, $X_{II} = \{x_8, \dots, x_{15}\} \in [-1, 1]$ e $X_{III} = \{x_{16}, \dots, x_{30}\} \in [-1, 1]$. Realizaram-se, ainda, as seguintes modificações no FDA2 – obtidas em conformidade com as sugestões de Farina *et al.* (2004) e extraídas de Sola (2010): define-se o primeiro objetivo como $f_1(\mathbf{x}) = 1 + \sum_{m=1}^7 x_m$ e as funções auxiliares, as quais geram a dinâmica temporal no FDA2, como $h(x_{III}, f_1, g) = 1 - \left(\frac{f_1}{g}\right)^{\left(H(t) + \sum_{x_i \in X_{III}} (x_i - H(t)/2\right)^2}$ e $H(t) = 5^{-\cos(\pi t/4)}$;

- dMOP1 – problema Tipo 1, definido como (adaptado de Goh e Tan (2009)):

$$\begin{aligned}
 f_1(x_r) &= x_r, & (6.1) \\
 f_2(\mathbf{x} \setminus x_r) &= g \cdot h, \\
 g(\mathbf{x} \setminus x_r) &= 1 + \sum_{i=1}^{\mathbf{x} \setminus x_r} (x_i - G(t))^2, \\
 h(f_1, g) &= 1 - \sqrt{\frac{f_1}{g}}, \\
 H(t) &= 0,75 \cdot \text{sen}(0,5 \cdot \pi \cdot t) + 1,25, \\
 G(t) &= \text{sen}(0,5 \cdot \pi \cdot t), \\
 r &= v \cdot m \cdot (g/G) \pmod{m}. & (6.2)
 \end{aligned}$$

em que $m = 10$ é a dimensionalidade do problema e v corresponde à periodicidade com que a variável a qual é considerada para a computação do primeiro objetivo é trocada. Para os resultados relatados nas próximas seções, $v = 3$, o que resulta na troca sequencial de variáveis, iniciando em x_1 , a cada 5 gerações. Ressalta-se que, após a variável x_{10} ser utilizada, a definição de r em termos do módulo do m resulta que x_1 é a próxima variável a ser utilizada.

Além disso, os parâmetros de severidade (n_g) e periodicidade (t_g) das mudanças temporais nos ambientes assumem valores no conjunto $\{5, 10\}$ – resultando em quatro combinações possíveis de ambientes para cada problema –, enquanto o número máximo de gerações é fixado em $G = 200$. Utiliza-se então a notação Problema(n_g, t_g) para fazer referência a um dos três problemas estudados, sobre uma parametrização específica de severidade e periodicidade. Por exemplo, FDA2(5,10) indica que estamos interessados no problema FDA2 com $n_g = 5$ e $t_g = 10$. Cabe ressaltar que a dinâmica temporal de todos os ambientes experimentados é expressa por

$$t = \frac{1}{n_g} \left\lfloor \frac{g}{t_g} \right\rfloor, \quad (6.3)$$

em que g representa a geração atual do processo evolucionário.

Finalmente, o Teste-T de Welch (Welch, 1947) é utilizado para aferir a significância estatística das diferenças entre as médias dos indicadores de desempenho, sempre no nível de significância de 0,05. Note-se que as médias são computadas para $E = 30$ repetições em relação a cada tratamento (variação de parâmetros).

6.1.3 Métricas de Avaliação

Para os experimentos, faz-se uso das seguintes métricas, algumas das quais utilizadas no estudo de correlação entre diversidade e Hipervolume no NSGA2, apresentado no Capítulo 4, e na validação experimental do modelo NDL-DG no Capítulo 5:

- $\Delta \mathcal{I}_h$ – diferença para o Hipervolume da distribuição ótima na PF ;
- \mathcal{I}_{GD} – distância entre a PF atual e a PF ótima (ver equação (3.6), página 30);
- $I_{M(R)}$ – os valores de diversidade genética relativa (ver definição na página 57);
- $|\mathcal{F}_1|_R$ – a proporção do número de soluções não-dominadas presentes ao longo de todo o processo evolucionário.

Como forma de caracterizar o efeito denominado nessa dissertação como *perda de Frentes*, propõe-se ainda a computação, a cada geração, da entropia normalizada do número de soluções particionadas em cada frente \mathcal{F}_i identificada na população:

$$H_N = -\frac{1}{\log_2 |\mathcal{F}|} \sum_{i=1}^{|\mathcal{F}|} p_i \log_2 p_i, \quad 0 \leq H_N \leq 1, \quad (6.4)$$

em que $|\mathcal{F}|$ corresponde ao número de Frentes identificadas e $p_i \equiv P\{\mathbf{x} \in \mathcal{F}_i\}$ é estimada a partir da frequência relativa do número de soluções \mathbf{x} identificadas na i -ésima frente \mathcal{F}_i . Note-se que

$$\begin{aligned} H_N = 0 &\iff |\mathcal{F}_1| = N, \text{ e } \mathcal{F}_j = \emptyset, (j = 2, \dots, N), \\ H_N = 1 &\iff |\mathcal{F}_1| = |\mathcal{F}_2| = \dots = |\mathcal{F}_k|, (k \leq N), \end{aligned} \quad (6.5)$$

ou seja, $H_N = 0$ ocorre apenas se todas as N soluções da população são não-dominadas, enquanto que $H_N = 1$ ocorre apenas se cada uma das $|\mathcal{F}| \leq N$ Frentes identificadas na população contiver o mesmo número de soluções das demais Frentes. As duas propriedades anteriores fazem da métrica proposta, H_N , uma medida adequada para aferir o que se denomina nessa dissertação por *diversidade de Frentes*. A perda de Frentes é então definida como o fenômeno ao longo da evolução de MOEAs em que o número de soluções não-dominadas na população aumenta, extinguindo antigas Frentes identificadas em estágios passados da evolução.

As métricas para a avaliação qualitativa da dinâmica de aproximação à PF nos algoritmos estudados em ambientes dinâmicos objetivam avaliar o desempenho dos

MOEAs em diferentes instantes de tempo. Na literatura, não há consenso em relação a quais instantes de tempo devem ser levados em consideração na análise. Uma abordagem muito utilizada é computar a *aptidão offline* (Jin e Branke, 2005), a qual considera a média tomada sobre o maior valor de aptidão presente na população a cada instante do processo de otimização. Mais especificamente, define-se

$$f_{\text{offline}} = \frac{1}{G} \sum_{t=1}^G f_t^*, \quad (6.6)$$

em que f_t^* representa a melhor aptidão na população desde a última mudança temporal. No entanto, uma vez que a presente pesquisa lida com otimização multiobjetivo e porque não é pertinente a atribuição de valores de aptidão individuais para soluções possivelmente não-comparáveis segundo a ordem de Pareto, escolheu-se computar a medida Hipervolume *offline*, expressa por

$$\overline{\mathcal{S}_h} = \frac{1}{G} \sum_{i=1}^G \left(\frac{1}{E} \sum_{j=1}^E \mathcal{S}_h(i, j) \right), \quad (6.7)$$

em que G é o número de gerações, E é o número de experimentos e $\mathcal{S}_h(i, j)$ representa o Hipervolume da população na i -ésima geração do experimento j . Essa métrica captura a qualidade de \mathcal{F}_1 em todos os estados temporais do processo, sendo adequada para refletir o desempenho de um algoritmo em um MOP dinâmico. Nos MOPs implementados, escolheu-se (11, 11) como o ponto de referência para a computação de $\mathcal{S}_h(i, j)$.

Um quadro descritivo das métricas utilizadas nesse Capítulo é apresentado na Tabela 6.2.

6.2 Inserção de Imigrantes na Otimização Multiobjetivo

Os experimentos relatados nessa seção descrevem a aplicação de geradores de diversidade por inserção de imigrantes – sem o uso das paisagens de não-dominância – em ambientes estáticos e dinâmicos, em que o efeito da variação das taxas de imigração (α) é estudado. O objetivo do estudo é caracterizar a repercussão do incremento de α na métrica de Hipervolume *offline* (equação (6.7)). Para tanto, comparam-se entre si os algoritmos NSGA2, NSGA2(R) e NSGA2(E).

Tabela 6.2 Quadro comparativo dos indicadores considerados para discussão dos resultados.

Indicador	Símbolo	Finalidade	Equação
Hipervolume	\mathcal{I}_h	Mensura o volume da região de dominância induzida pelas soluções em \mathcal{F}_1 .	3.9
Diferença para Hipervolume ótimo	$\Delta\mathcal{I}_h$	Mensura a diferença entre o Hipervolume observado e o Hipervolume calculado sobre a distribuição ótima de soluções sobre a PF .	–
Hipervolume <i>offline</i>	$\overline{\mathcal{I}_h}$	Expressa a média de Hipervolume tomada sobre todas as gerações e sobre todos os experimentos.	6.7
Distância Geracional	\mathcal{I}_{GD}	Mensura a proximidade de \mathcal{F}_1 para a PF no espaço de objetivos.	3.6
Proporção de soluções não-dominadas	$ \mathcal{F}_1 _{\mathbb{R}}$	Representa a proporção $\frac{ \mathcal{F}_1 }{ \Psi }$, em que $N = \Psi $ é o tamanho da população.	–
Momento de Inércia	I_M	Mensura a (média de) diversidade genética em Ψ , no espaço de decisão.	4.2
Momento de Inércia Relativo	$I_{M(R)}$	Corresponde à razão I_M/I_M^{\max} , em que I_M^{\max} é o maior valor médio observado dentre todos os tratamentos experimentais considerados.	–
Entropia Normalizada das Frentes	H_N	Mensura a diversidade de Frentes, i.e., a entropia normalizada da distribuição das soluções entre as Frentes identificadas em Ψ .	6.4

6.2.1 Efeitos do Incremento de α em Ambientes Estáticos

Com a avaliação dos efeitos da inserção de imigrantes em MOPs de ambientes estáticos, pretende-se caracterizar a repercussão negativa do aumento da diversidade genética no Hipervolume máximo obtido ao longo das gerações. Na otimização de um único objetivo, já fora experimentalmente observado – e.g. Yang (2008) – que o aumento da taxa de imigração acarreta na redução da capacidade de AEs mono-objetivo em manter níveis médios elevados de aptidão da população em fases estacionárias de problemas dinâmicos. O estudo a seguir, portanto, busca confirmar essa tendência em MOPs estáticos, invariantes no tempo.

Primeiro, com o incremento de α , observa-se da Figura 6.1 para o problema ZDT2 (i) que a diferença para o Hipervolume ótimo aumenta, com tendência exponencial (parte (a) da Figura); (ii) que a distância geracional para a PF^* aumenta, com tendência exponencial (parte (b) da Figura); (iii) que a diversidade genética relativa aumenta (parte (c) da Figura); e (iv) que a proporção do número de soluções não-dominadas mantidas

6.2. INSERÇÃO DE IMIGRANTES NA OTIMIZAÇÃO MULTIOBJETIVO

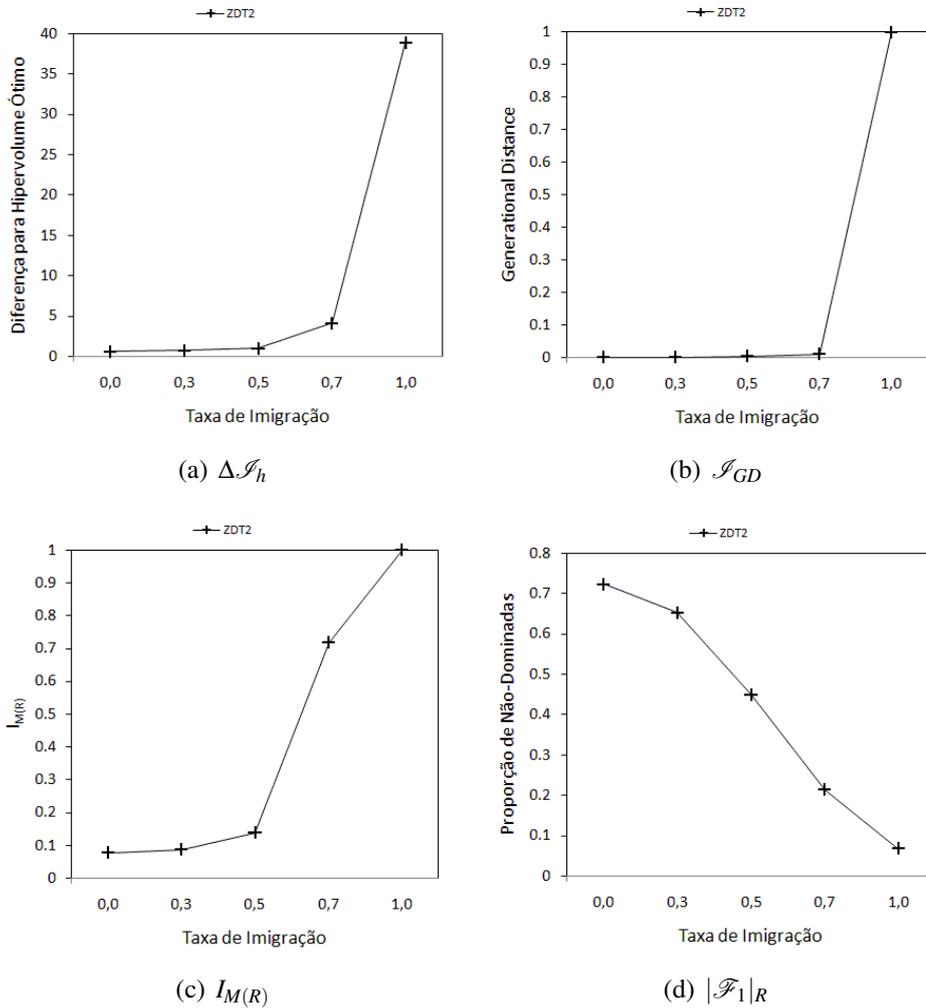


Figura 6.1 Resultados para os experimentos com o NSGA(R) no problema ZDT2.

na população ao longo das 200 gerações diminui (parte (d) da Figura). Ressalta-se que os eventos (i) a (iv) ocorrem da mesma forma para os demais problemas (ZDT1, ZDT3, ZDT4 e ZDT6) e para os experimentos com o NSGA2(E). Esse resultado é explicado a seguir, a partir da visualização do efeito do incremento de α sobre os níveis médios de diversidade genética da população.

Com a visualização da parte (b) da Figura 6.2, percebe-se que os níveis médios de diversidade genética (I_M) tomados a cada geração com o NSGA2(R) *após* a inserção dos imigrantes na população e *antes* da seleção de sobreviventes – I_M nesse caso foi computada sobre o conjunto de $100 \text{ indivíduos} \cup \alpha \times 100 \text{ imigrantes}$ – apresentam leve tendência de crescimento para $\alpha = 0,5$, $\alpha = 0,7$ e, por volta do período compreendido

entre a 15ª e a 200ª geração, também para $\alpha = 0,3$. Para $\alpha = 0,0$, percebe-se que não há meios de evitar a perda de diversidade pelos mecanismos convencionais de seleção por *crowding distance* e pelos operadores de variação implementados no NSGA2. Já para $\alpha = 1,0$, percebe-se que os níveis médios de I_M oscilam em torno de um nível constante ao longo do processo evolucionário. Os pontos de inflexão visualizados, para $\alpha = 0,3$, por volta da 15ª geração e, para $\alpha = 0,5$, por volta da 20ª geração, podem ser explicados pelo afastamento progressivo da subpopulação de trabalho – a qual caminha em direção à PF – da subpopulação de imigrantes, a qual permanece sendo gerada, com maior probabilidade, em torno de uma região fixa, de menor proximidade em relação à PF . Isso porque, até a geração 15 (ou 20, dependendo de α), muitos imigrantes são ainda aproveitados e acabam sobrevivendo na população de trabalho. A partir do ponto de inflexão, entretanto, a probabilidade de que um imigrante seja aproveitado na subpopulação de trabalho diminui significativamente, de modo que as duas subpopulações começam progressivamente a se afastar, o que explica a tendência de aumento de I_M quando medida nesse ponto. O decréscimo das probabilidades de sobrevivência das soluções atípicas inseridas na população a partir do ponto de inflexão não apenas ilustra a dificuldade da estratégia de geração imigrantes aleatórios criar soluções adequadas ao longo de todo o processo evolucionário, mas também explica o porquê do NSGA2(R) apresentar uma tendência exponencial de degradação da qualidade média da população (em termos de Hipervolume) com o incremento de α em períodos estacionários do problema de otimização, já que, no final do processo evolucionário, quando as duas subpopulações encontram-se completamente isoladas, observa-se que o tamanho da subpopulação de trabalho efetivamente utilizada para a computação do Hipervolume (composta de soluções não-dominadas) encontra-se limitado em aproximadamente $\alpha \times 100\%$. Ressalta-se que, nos experimentos com o NSGA2(E), apesar da subpopulação de imigrantes ser gerada a partir de mutações da elite, não se observou ganhos de Hipervolume com significância estatística para cada valor fixo de α em se comparando os resultados obtidos com aqueles relatados para o NSGA2(R).

No que concerne à média dos níveis de diversidade de Frentes (H_N), tomada a cada geração para as populações evoluídas com o NSGA2(R), observa-se na parte (a) da Figura 6.2 que, com $\alpha = 0,0$, o NSGA2 apresenta acentuada tendência de queda em H_N , obtendo, por volta da 90ª geração, $H_N = 0$, indicando que $|\mathcal{F}_1| = 100$, ou seja, que todos os indivíduos na população são mutuamente não-dominados. A perda de Frentes é atenuada com o incremento de α , sendo H_N praticamente constante quando $\alpha = 1$, i.e., quando a população é reinicializada a cada geração. Cabe ressaltar que se observou o

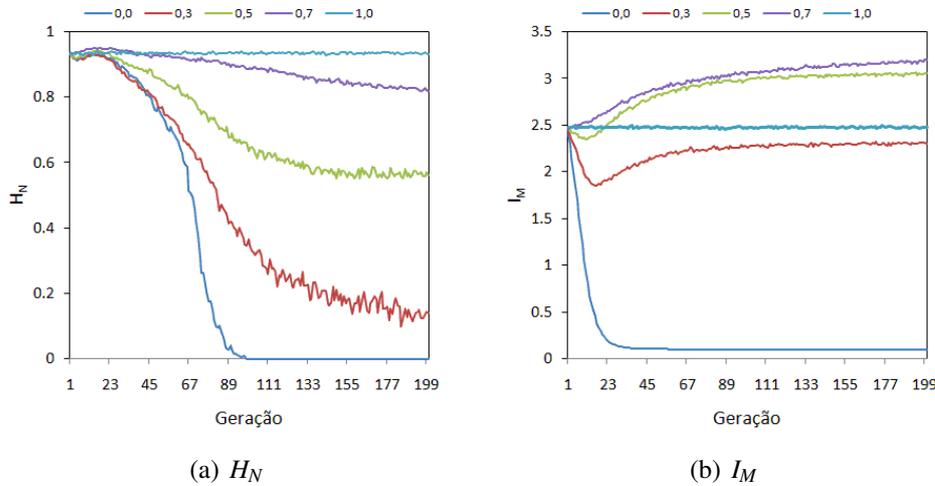


Figura 6.2 Resultados obtidos com o NSGA(R) no ZDT2: (a) entropia normalizada; (b) diversidade medida após a inserção de imigrantes na população e antes da seleção de sobreviventes.

mesmo comportamento nos experimentos com os demais MOPs do conjunto ZDT e nos experimentos realizados com o NSGA2(E).

Em MOPs dinâmicos, a principal consequência da perda de Frentes é que, com a mudança do ambiente – conforme discutido no Capítulo 3, página 43 – soluções antes não-dominadas podem figurar como soluções inadequadas nos novos estados temporais do processo de otimização. Dessa forma, com a inserção de imigrantes, mantêm-se níveis de diversidade de Frentes maiores, os quais podem vir a se mostrarem importantes ante mudanças temporais nas funções objetivo.

6.2.2 Efeitos do Incremento de α em Ambientes Dinâmicos

Para os experimentos em ambientes dinâmicos, percebe-se que o aumento da taxa de imigração leva a maiores valores médios de Hipervolume *offline* no NSGA2(E) e no NSGA2(R), quando se comparam os resultados obtidos com o NSGA2 ($\alpha = 0$). Esse comportamento é ilustrado por meio dos resultados para o FDA1, na Figura 6.3. A única exceção é o caso $\alpha = 1,0$ com o NSGA2(R), em que se observa a degradação da qualidade média das populações evoluídas ao longo de todo o processo evolucionário. Isso porque, com o NSGA2(R), $\alpha = 1,0$ implica que toda a população é substituída a cada geração por imigrantes aleatórios, equivalendo-se a uma estratégia de busca aleatória. Daí resulta também a tendência de queda para taxas de imigração maiores que $\alpha = 0,3$ com o NSGA2(R). O efeito do incremento de α nos experimentos com o NSGA2(E), por

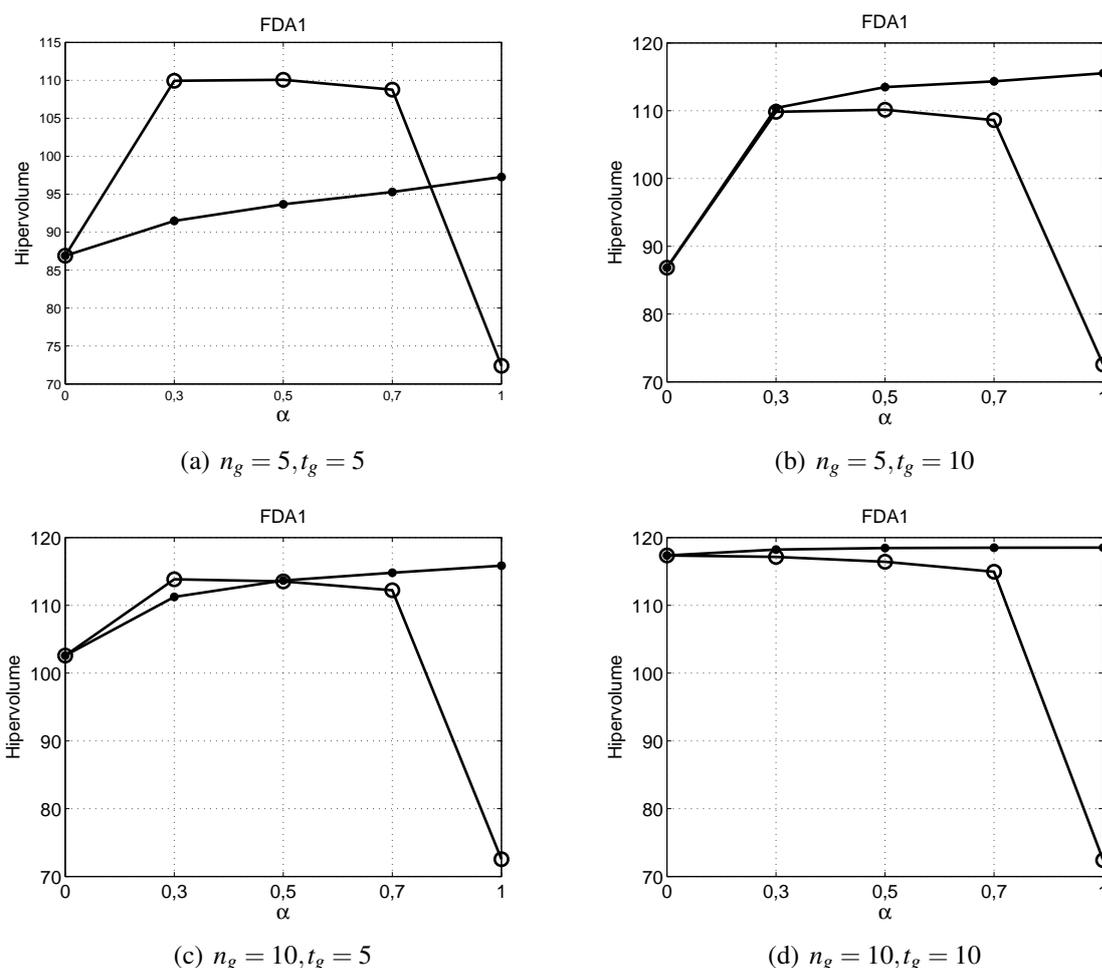


Figura 6.3 Resultados obtidos com o NSGA2(R) (\circ) e NSGA2(E) (\bullet) no FDA1 em função de α .

outro lado, apresenta tendência de crescimento, tendo se obtido, inclusive, os melhores resultados em termos de $\overline{\mathcal{F}}_h$ para $\alpha = 1, 0$. A diferença de comportamento em relação ao incremento de α no NSGA2(R) e no NSGA2(E) deve-se, a princípio, às maiores chances de sobrevivência dos imigrantes correlacionados (inseridos no NSGA2(E)) em relação aquelas dos imigrantes não-correlacionados (inseridos no NSGA2(R)). Dessa forma, para $\alpha = 1, 0$ no NSGA2(E), apesar de toda a população ser substituída pelos imigrantes gerados a partir de mutações da elite da população anterior, resulta que, por serem inseridos em regiões próximas às das soluções contidas em \mathcal{F}_1 , parte dos novos imigrantes potencialmente promove o caminhar à $PF(t)$ e, portanto, conclui-se que o NSGA2(E) tende a ser melhor com respeito ao NSGA2(R) à medida que α cresce.

Também se percebe das partes da Figura 6.3, que os valores de Hipervolume *offline* medidos das populações evoluídas com o NSGA2(E) ultrapassam aqueles obtidos com o

NSGA2(R) para diferentes taxas nas partes (b), (c) e (d) da Figura. Quando se consideram os 12 cenários possíveis de MOPs dinâmicos – com as combinações de parâmetros de severidade, n_g , e periodicidade, t_g –, constata-se que esse evento ocorre em sete dos casos estudados. Nos cinco cenários restantes, observa-se um comportamento semelhante ao da Figura 6.3 (a), em que, apesar de apresentar tendência de crescimento, o Hipervolume *offline* observado para o NSGA2(E) não ultrapassa aquele observado para o NSGA2(R) com o incremento de α . Ressalta-se que a discussão realizada nessa Seção aplica-se também para os resultados obtidos da aplicação do NSGA2(R) e NSGA2(E) nos demais MOPs dinâmicos estudados (FDA2 e dMOP1).

6.3 Seleção de Imigrantes por PNDs

Os objetivos do estudo com a incorporação dos geradores NDL-DG_E e NDL-DG_R ao NSGA2 são (1) avaliar as diferenças do Hipervolume *offline* (equação (6.7)) em relação às diferentes implementações do NSGA2 (ver quadro na Tabela 6.1); e (2) visualizar as correlações estatísticas médias entre as diversidades genética e de Frentes e Hipervolume, a cada geração. Para tanto, comparam-se entre si os algoritmos NSGA2(E_{N_{NDL}}), NSGA2(R_{N_{NDL}}), NSGA2(E), e NSGA2(R) executados sempre a uma taxa de imigração fixa $\alpha = 0,3$ nos MOPs dinâmicos FDA1, FDA2 e dMOP1. Nesse estudo, a escolha de fixar α no menor valor dentre as taxas de imigração estudadas (0,3), se deu, principalmente, por ser essa uma taxa da qual se espera uma menor influência da variação de K_{NDL} , objetivando-se dificultar ao máximo a observação de diferenças significativas nos resultados com o incremento de K_{NDL} . Sugere-se estudar outras combinações entre os valores de α e K_{NDL} em trabalhos futuros.

6.3.1 Repercussão do Incremento de K_{NDL} no Hipervolume *Offline*

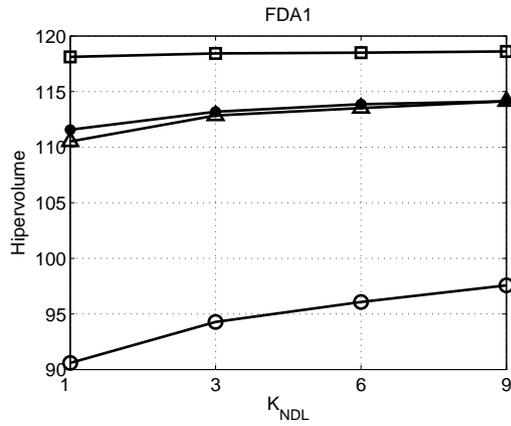
A partir da Tabela 6.3 – e com o auxílio da Figura 6.4 –, a primeira observação dos resultados obtidos em termos de Hipervolume *offline* é a de que a incorporação das NDLs aos geradores de diversidade contribui em $\approx 79\%$ (19/24) dos ambientes dinâmicos testados para o aumento da qualidade média da população ao longo de todo o processo evolucionário com destacada significância estatística em relação à ausência das NDLs no mecanismo de inserção de imigrantes. Nos demais casos ($\approx 21\%$ (5/24)), a incorporação das NDLs é estatisticamente equivalente à sua ausência. Na Tabela 6.3 – especialmente

nas entradas correspondentes ao gerador NDL-DG_E – os baixos *p-values*³ obtidos são um indicativo da capacidade do NDL-DG de direcionar a busca evolucionária nas regiões promissoras em MOPs dinâmicos. Ao se considerarem separadamente os resultados da aplicação dos algoritmos NSGA2(E_{NDL}) e NSGA2(R_{NDL}), percebe-se que o primeiro é consistente nos cenários testados, ao obter resultados significativamente superiores àqueles obtidos com o NSGA2(E) em 100% (12/12) dos cenários de MOPs dinâmicos estudados. Já o segundo obteve ganhos com significância estatística em relação ao NSGA2(R) em $\approx 58\%$ (7/12) dos cenários. Dentre os MOPs dinâmicos estudados, o problema dMOP1 demonstrou ser aquele em que, aparentemente, o emprego do NDL-DG apresenta maiores dificuldades em obter ganhos de $\overline{\mathcal{J}}_h$, talvez devido ao aumento da complexidade da dinâmica temporal com a introdução da troca periódica da variável a partir da qual o valor do primeiro objetivo é computado. Também há a possibilidade de que, como discutido no Capítulo anterior (página 93), o fato do primeiro objetivo do dMOP1 ser definido em termos de uma única variável – enquanto que se define o segundo objetivo em termos das 10 variáveis que compõem o problema – produz um viés indesejado na busca, em que se torna mais fácil obter soluções não-dominadas pela alteração de um único gene, influenciando negativamente, assim, a capacidade de exploração dos algoritmos os quais empregam o NDL-DG. Finalmente, ainda se levando em consideração a Tabela 6.3 e a Figura 6.4, não há padrões evidentes da significância estatística dos ganhos de $\overline{\mathcal{J}}_h$ sobre a ausência das NDLs a partir da variação dos parâmetros de severidade (n_g) e periodicidade (t_g). Por exemplo, em cada um dos três MOPs dinâmicos, observa-se que o NSGA2(E_{NDL}) obtém os maiores ganhos de $\overline{\mathcal{J}}_h$ com significância estatística em relação aos valores obtidos com o NSGA2(E) em três diferentes combinações de n_g e t_g – FDA1(5, 10), FDA2(10, 5) e dMOP1(10, 10). Esse fato pode ser interpretado como (1) um indicativo da consistência dos resultados obtidos com o NDL-DG sobre diferentes dinâmicas temporais e/ou (2) que diferentes parametrizações de n_g e t_g produzem efeitos distintos de mudanças no ambiente os quais dependem da forma como os MOPs dinâmicos são projetados. Como observado em Goh e Tan (2009), dada a forma como o conjunto FDA é projetado, n_g não influencia apenas a severidade, mas também repercute no período das funções seno empregadas na dinâmica temporal daqueles problemas, o que suporta a interpretação (2).

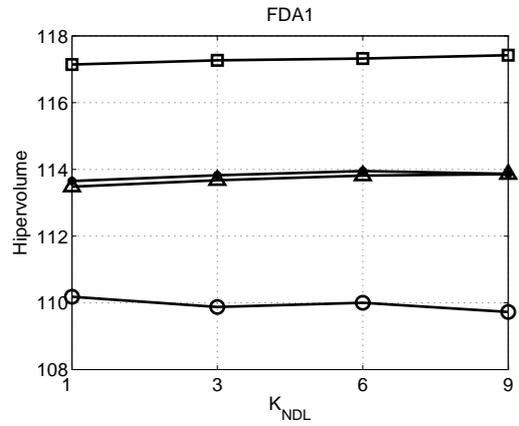
Um segundo fato chama a atenção em alguns cenários experimentais com a incorporação do NDL-DG ao NSGA2: apesar do NSGA2(E) produzir populações com

³O *p-value* corresponde à probabilidade de que as diferenças entre as médias observadas de $\overline{\mathcal{J}}_h$ possam ser explicadas unicamente por fatores aleatórios inerentes à natureza estocástica das simulações.

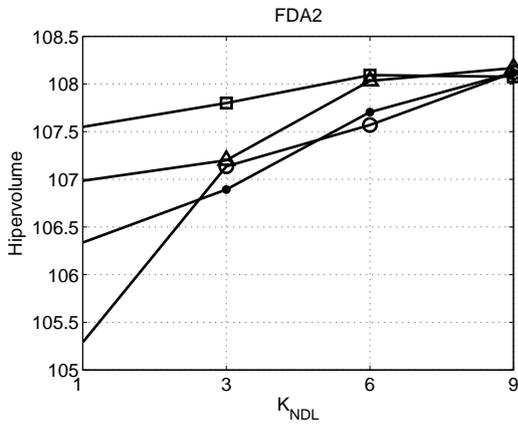
6.3. SELEÇÃO DE IMIGRANTES POR PNDS



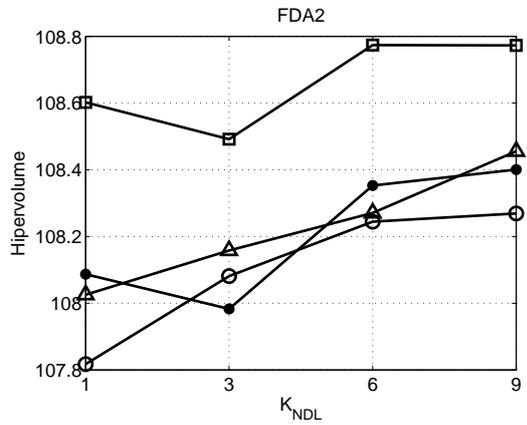
(a) NSGA2(E_NDL) no FDA1



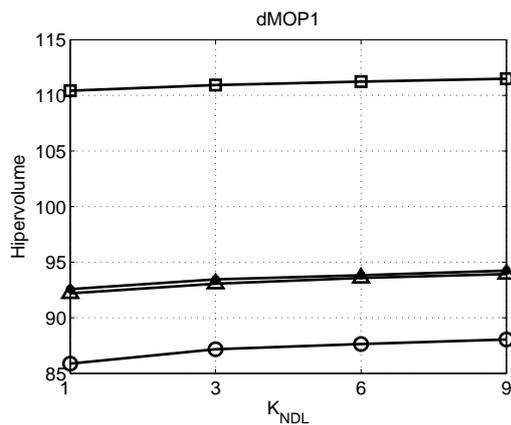
(b) NSGA2(R_NDL) no FDA1



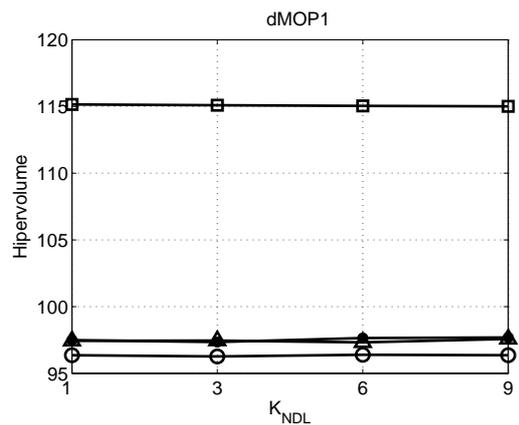
(c) NSGA2(E_NDL) no FDA2



(d) NSGA2(R_NDL) no FDA2



(e) NSGA2(E_NDL) no dMOP1



(f) NSGA2(R_NDL) no dMOP1

Figura 6.4 Variação no Hipervolume (\mathcal{I}_H) com o incremento de K_{NDL} em MOPs dinâmicos: $n_g = 5, t_g = 5$ (○); $n_g = 5, t_g = 10$ (△); $n_g = 10, t_g = 5$ (●) e $n_g = 10, t_g = 10$ (□).

Tabela 6.3 p -values computados para os testes de significância de $\overline{\mathcal{F}}_h$ entre a melhor configuração com NDLS (K_{NDL} entre parênteses) e sem o uso de NDLS.

Parâmetros	FDA1		FDA2		dMOP1	
	NDL-DG _E	NDL-DG _R	NDL-DG _E	NDL-DG _R	NDL-DG _E	NDL-DG _R
$n_g = 5, t_g = 5$	(9) $\approx 10^{-60}$	(6) 0,13	(9) 0,0	(9) $\approx 10^{-15}$	(9) $\approx 10^{-05}$	(3) 0,93
$n_g = 5, t_g = 10$	(9) $\approx 10^{-128}$	(9) $\approx 10^{-04}$	(9) $\approx 10^{-75}$	(9) $\approx 10^{-15}$	(9) $\approx 10^{-04}$	(9) 0,77
$n_g = 10, t_g = 5$	(9) $\approx 10^{-76}$	(6) $\approx 10^{-03}$	(9) $\approx 10^{-146}$	(9) $\approx 10^{-09}$	(9) $\approx 10^{-04}$	(9) 0,67
$n_g = 10, t_g = 10$	(9) $\approx 10^{-11}$	(9) $\approx 10^{-03}$	(6) $\approx 10^{-15}$	(9) $\approx 10^{-03}$	(9) $\approx 10^{-72}$	(3) 0,06

Hipervolume *offline* significativamente menor que aquelas evoluídas com o NSGA2(R), a partir do incremento de K_{NDL} , os valores de Hipervolume das populações evoluídas com o NSGA2(E_{NDL}) ultrapassam, na média, aqueles obtidos com o NSGA2(R) e, em alguns casos, com o NSGA2(R_{NDL}) utilizando os melhores valores de K_{NDL} . Isso fora observado em $\approx 41\%$ (5/12) dos cenários de otimização dinâmica estudados: FDA1(5, 10), FDA1(10, 5), FDA2(5, 5), FDA2(5, 10) e FDA2(10, 5). Por exemplo, no FDA1(5, 10), o NSGA2(E) produz populações em que $\overline{\mathcal{F}}_h \approx 110,5$, enquanto que as populações evoluídas com o NSGA2(R) resultam em $\overline{\mathcal{F}}_h \approx 113,5$. Com o incremento de K_{NDL} , entretanto, já com $K_{\text{NDL}} = 6$ tem-se que o valor de $\overline{\mathcal{F}}_h$ ($\approx 113,9$) para as populações otimizadas com o NSGA2(E_{NDL}) é maior – com significância estatística – que aquele obtido com o NSGA2(R). Para $K_{\text{NDL}} = 9$, esse valor chega a $\approx 114,13$, ultrapassando significativamente o melhor resultado obtido com o NSGA2(R_{NDL}) nesse cenário, mesmo tendo gerado níveis de diversidade genética menores ($I_M \approx 70$ contra $I_M \approx 210$ no NSGA2(R)). Essa observação não apenas é indicativa do alto nível de exploração em regiões promissoras induzido pela combinação das NDLS com a estratégia de imigrantes correlacionados, mas também abre uma nova perspectiva para a otimização dinâmica em geral: a hipótese de que explorar regiões promissoras – ainda que com a manutenção de níveis de diversidade significativamente mais baixos – permite, em casos a ainda serem identificados em trabalhos futuros, uma melhor adaptação da população do que a simples manutenção de altos níveis de diversidade ao longo de todo o processo evolucionário.

Das Tabelas 6.4 a 6.7, analisando-se os MOPs dinâmicos separadamente e ao se considerarem as quatro combinações de valores de n_g e t_g , têm-se:

- No FDA1, o NSGA2(E_{NDL}) obtém ganhos significativos de Hipervolume em, no mínimo, 94% das 200 gerações executadas em relação aos valores obtidos com o NSGA2(E), apresentando resultados equivalentes (coluna \equiv) no restante das gerações, enquanto que o NSGA2(R_{NDL}) obtém o mesmo quadro em, no mínimo, 25% das gerações em relação aos valores obtidos com o NSGA2(R), apresentando

6.3. SELEÇÃO DE IMIGRANTES POR PNDS

Tabela 6.4 Percentual do número de gerações em que as NDLs apresentam ganhos (+) com significância estatística para $K_{\text{NDL}} > 1$ em relação à ausência de NDLs ($n_g = 5, t_g = 5$)

Algoritmo	Indicador	FDA1			FDA2			dMOP1		
		+	≡	-	+	≡	-	+	≡	-
NSGA2(E_{NDL})	\mathcal{S}_H	94,0	6,0	0,0	98,5	1,5	0,0	83,9	15,6	0,5
	I_M	96,0	4,0	0,0	76,9	23,1	0,0	59,8	39,7	0,5
	H_N	1,0	82,4	16,6	0,5	90,5	9,0	6,0	70,4	23,6
NSGA2(R_{NDL})	\mathcal{S}_H	32,2	42,7	25,1	24,1	75,9	0,0	25,1	48,2	26,7
	I_M	3,0	72,4	24,6	4,0	95,5	0,5	24,1	55,8	20,1
	H_N	5,5	91,5	3,0	2,5	97,5	0,0	22,1	74,9	3,0

Tabela 6.5 Percentual do número de gerações em que as NDLs apresentam ganhos (+) com significância estatística para $K_{\text{NDL}} > 1$ em relação à ausência de NDLs ($n_g = 5, t_g = 10$).

Algoritmo	Indicador	FDA1			FDA2			dMOP1		
		+	≡	-	+	≡	-	+	≡	-
NSGA2(E_{NDL})	\mathcal{S}_H	95,5	4,5	0,0	43,7	56,3	0,0	81,9	17,6	0,5
	I_M	84,4	15,6	0,0	67,8	30,7	1,5	67,3	23,6	9,1
	H_N	4,5	69,9	25,6	0,0	97,5	2,5	4,5	65,8	29,7
NSGA2(R_{NDL})	\mathcal{S}_H	31,7	68,3	0,0	29,1	70,9	0,0	19,1	56,3	24,6
	I_M	16,6	81,4	2,0	2,5	97,5	0,0	34,2	62,8	3,0
	H_N	7,5	88,5	4,0	4,5	93,0	2,5	22,6	72,9	4,5

Tabela 6.6 Percentual do número de gerações em que as NDLs apresentam ganhos (+) com significância estatística para $K_{\text{NDL}} > 1$ em relação à ausência de NDLs ($n_g = 10, t_g = 5$).

Algoritmo	Indicador	FDA1			FDA2			dMOP1		
		+	≡	-	+	≡	-	+	≡	-
NSGA2(E_{NDL})	\mathcal{S}_H	95,0	5,0	0,0	86,9	13,1	0,0	83,4	15,1	1,5
	I_M	71,9	28,1	0,0	49,2	50,8	0,0	63,8	27,6	8,6
	H_N	4,5	79,4	16,1	0,0	90,9	9,1	7,5	60,3	32,2
NSGA2(R_{NDL})	\mathcal{S}_H	31,7	53,8	14,6	36,2	63,8	0,0	17,1	58,8	24,1
	I_M	11,6	85,4	3,0	17,1	80,9	2,0	27,6	65,8	6,6
	H_N	8,5	88,0	3,5	2,5	96,5	1,0	23,1	68,3	8,6

Tabela 6.7 Percentual do número de gerações em que as NDLs apresentam ganhos (+) com significância estatística para $K_{NDL} > 1$ em relação à ausência de NDLs ($n_g = 10, t_g = 10$).

Algoritmo	Indicador	FDA1			FDA2			dMOP1		
		+	≡	-	+	≡	-	+	≡	-
NSGA2(E _{NDL})	\mathcal{S}_H	97,5	2,5	0,0	35,2	64,8	0,0	90,5	9,0	0,5
	I_M	89,4	10,6	0,0	65,3	34,2	0,5	81,4	18,6	0,0
	H_N	2,0	90,0	8,0	0,0	90,0	10,0	2,5	56,3	41,2
NSGA2(R _{NDL})	\mathcal{S}_H	25,1	74,9	0,0	30,2	69,8	0,0	10,6	64,8	24,6
	I_M	10,6	86,4	3,0	0,0	98,0	2,0	52,8	42,2	5,0
	H_N	2,5	95,0	2,5	4,5	91,5	4,0	35,7	58,8	5,5

resultados equivalentes na maioria do restante das gerações;

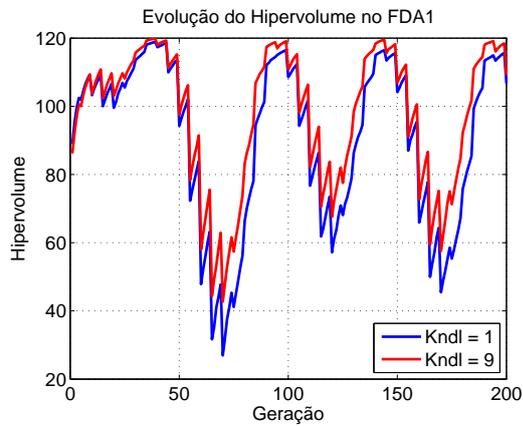
- No FDA2, o NSGA2(E_{NDL}) obtém ganhos significativos de Hipervolume em, no mínimo, 35,2% das 200 gerações executadas em relação aos valores obtidos com o NSGA2(E), apresentando resultados equivalentes no restante das gerações, enquanto que o NSGA2(R_{NDL}) obtém o mesmo quadro em, no mínimo, 24,1% das gerações em relação aos valores obtidos com o NSGA2(R), apresentando resultados equivalentes no restante das gerações;
- No dMOP1, o NSGA2(E_{NDL}) obtém ganhos significativos de Hipervolume em, no mínimo, 81,9% das 200 gerações executadas em relação aos valores obtidos com o NSGA2(E), apresentando resultados equivalentes na maioria do restante das gerações. O NSGA2(R_{NDL}), por outro lado, não apresenta percentuais do número de gerações em que há ganhos significativos de Hipervolume em relação ao NSGA2(R) os quais demonstram a consistência do algoritmo no dMOP1, como se observa nos demais problemas. Isso se reflete no fato dos valores de Hipervolume *offline* obtidos com o NSGA2(R_{NDL}) serem estatisticamente equivalentes aos obtidos com o NSGA2(R) (Tabela 6.3).

As observações anteriores indicam que, quando se analisam as porcentagens do número de gerações em que o Hipervolume das populações evoluídas com o uso do NDL-DG apresenta ganhos com significância estatística em relação àqueles dos mecanismos de inserção de imigrantes sem seleção por NDLs ($K_{NDL} = 1$), a eficiência do NSGA2(E_{NDL}) e do NSGA2(R_{NDL}) nos MOPs dinâmicos estudados também se destaca, em comparação às do NSGA2(E) e do NSGA2(R), respectivamente. Esse é um resultado importante porque a indicação de que um algoritmo A produz soluções mais bem adaptadas em relação àquelas produzidas por um algoritmo B na maioria dos instantes de tempo é

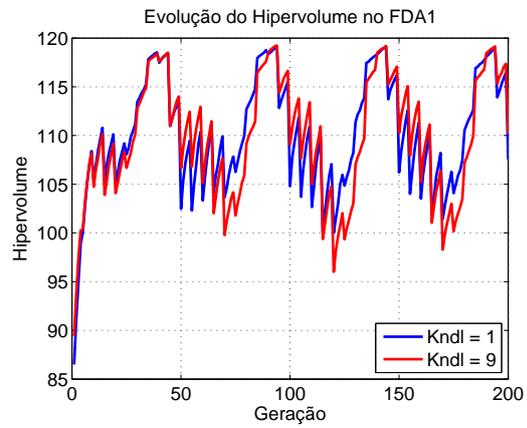
crucial na escolha de qual método utilizar em problemas de otimização dinâmica práticos, em que se desconhece a dinâmica temporal exata das funções objetivo e/ou restrições.

As Figuras 6.5 a 6.8 mostram a evolução temporal da média tomada sobre os 30 experimentos do indicador de Hipervolume computado a cada geração, após as etapas de seleção e recolocação de sobreviventes no NSGA2(E), NSGA2(R) e nas melhores configurações (em termos de Hipervolume *offline*) obtidas com a variação de K_{NDL} no NSGA2(E_{NDL}) e NSGA2(R_{NDL}). Em geral, percebe-se que a variação dos parâmetros do ambiente (n_g e t_g) leva a dinâmicas de mudanças temporais bastante distintas, nos três MOPs dinâmicos estudados. A partir da inspeção visual dos gráficos, analisam-se a seguir as três situações possíveis, imediatamente após mudanças temporais: (1) o Hipervolume médio das populações evoluídas com o uso do NDL-DG (com $K_{\text{NDL}} > 1$) fica acima daquele o qual corresponde às populações evoluídas sem as NDLs; (2) o Hipervolume médio com $K_{\text{NDL}} > 1$ fica abaixo; ou (3) não há diferenças perceptíveis entre os valores médios de Hipervolume de ambas as populações.

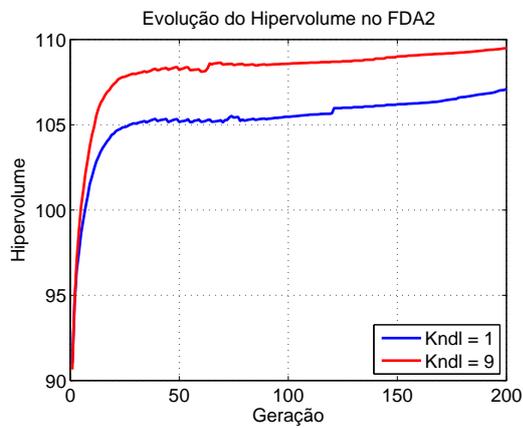
- A situação (1) é predominante nos gráficos, em conformidade com a discussão anterior da análise das Tabelas 6.4 a 6.7, em especial, com o algoritmo NSGA2(E_{NDL}). Esse fato demonstra que, em geral, as populações evoluídas com a inserção de imigrantes gerados pelo NDL-DG mostram-se mais bem adaptadas às mudanças temporais. Uma possível razão é o fato de, em geral, os níveis médios de diversidade genética das populações evoluídas com a presença das paisagens de não-dominância terem estado acima daqueles observados para as populações evoluídas sem as NDLs.
- A situação (2) ocorre, principalmente, em alguns períodos temporais dos experimentos com o NSGA2(R_{NDL}). Embora seja difícil afirmar o *porquê* da situação (2) ocorrer no NSGA2(R_{NDL}), a identificação de *quando* ela ocorre pode ser suficiente para tomadas de decisão de, em se detectando situações como as identificadas, fazer uso de estratégias mais eficientes. Dessa forma, analisando-se os gráficos – especificamente, nas partes (b) e (f) das Figuras 6.5 a 6.8 –, percebe-se que os padrões em que a situação (2) ocorre correspondem, em geral, aos períodos em que as mudanças temporais são menos acentuadas e em que a série de Hipervolume médio apresenta tendência de crescimento. Note-se, todavia, que mesmo com a observação da situação (2) naqueles gráficos, a situação (1) de certa forma compensa as quedas de Hipervolume em outros estados temporais, levando à obtenção de ganhos de Hipervolume *offline* com significância estatística em $\approx 58\%$ (7/12) dos cenários, como já discutido anteriormente;



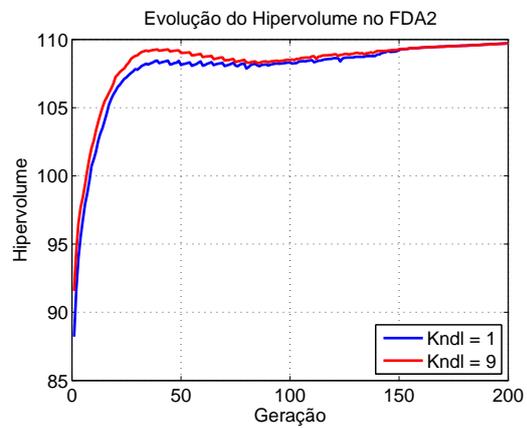
(a) NSGA2(E_{NDL}) no FDA1



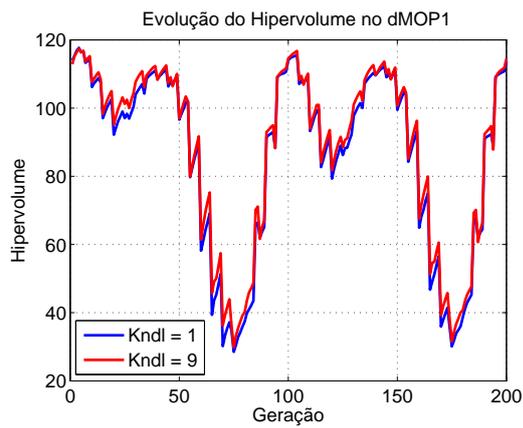
(b) NSGA2(R_{NDL}) no FDA1



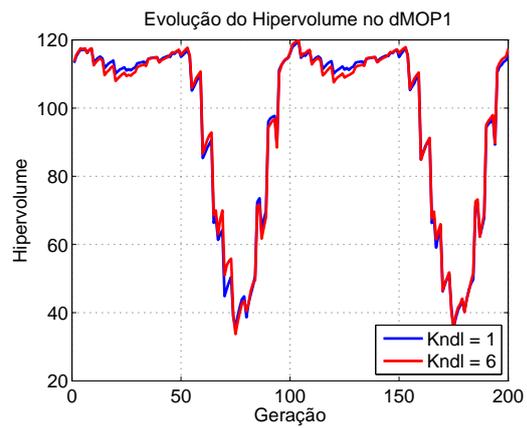
(c) NSGA2(E_{NDL}) no FDA2



(d) NSGA2(R_{NDL}) no FDA2



(e) NSGA2(E_{NDL}) no dMOP1



(f) NSGA2(R_{NDL}) no dMOP1

Figura 6.5 Evolução temporal do indicador de Hipervolume com e sem o NDL-DG em ambientes dinâmicos ($n_g = 5, t_g = 5$).

6.3. SELEÇÃO DE IMIGRANTES POR PNDS

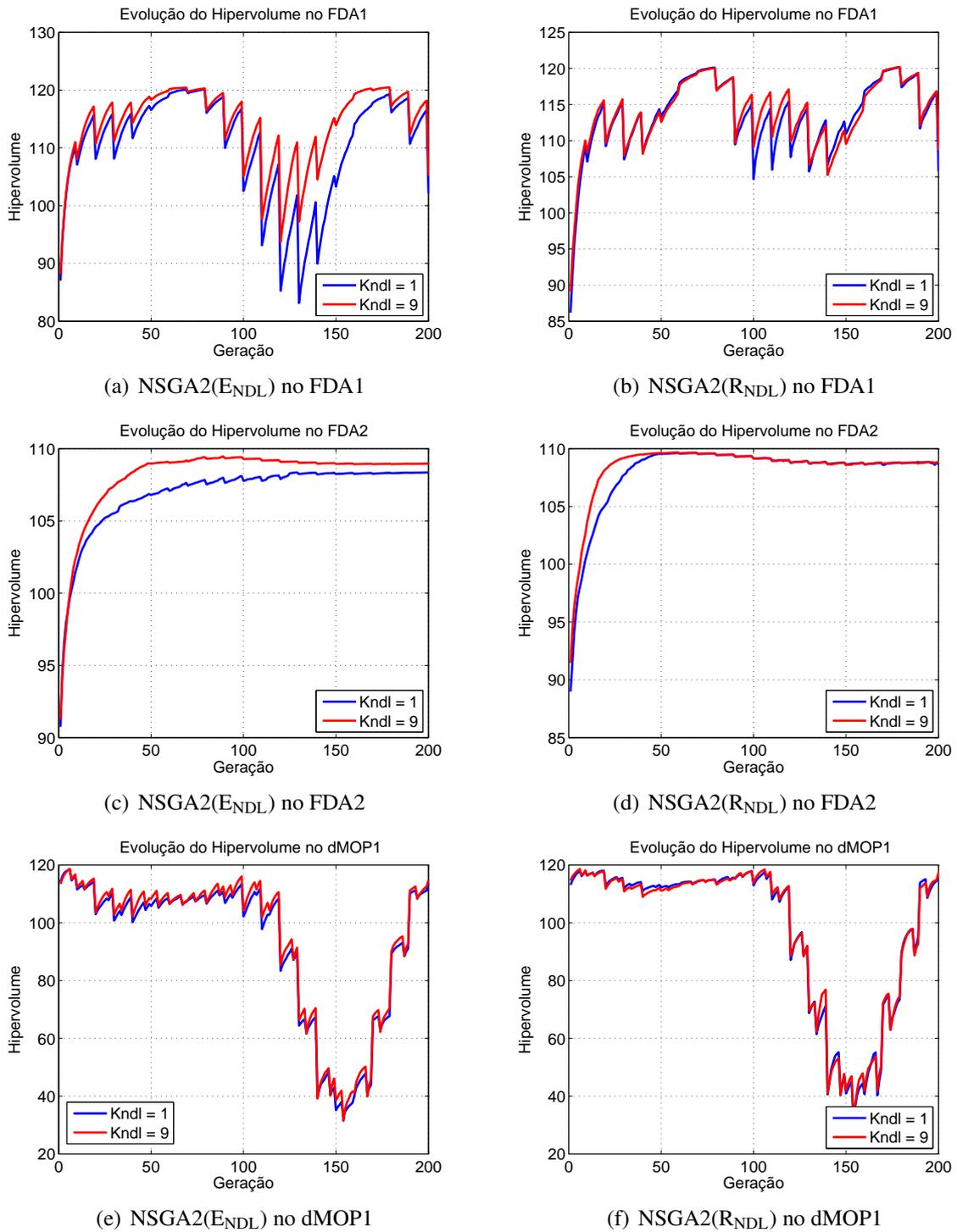


Figura 6.6 Evolução temporal do indicador de Hipervolume com e sem o NDL-DG em ambientes dinâmicos ($n_g = 5, t_g = 10$).

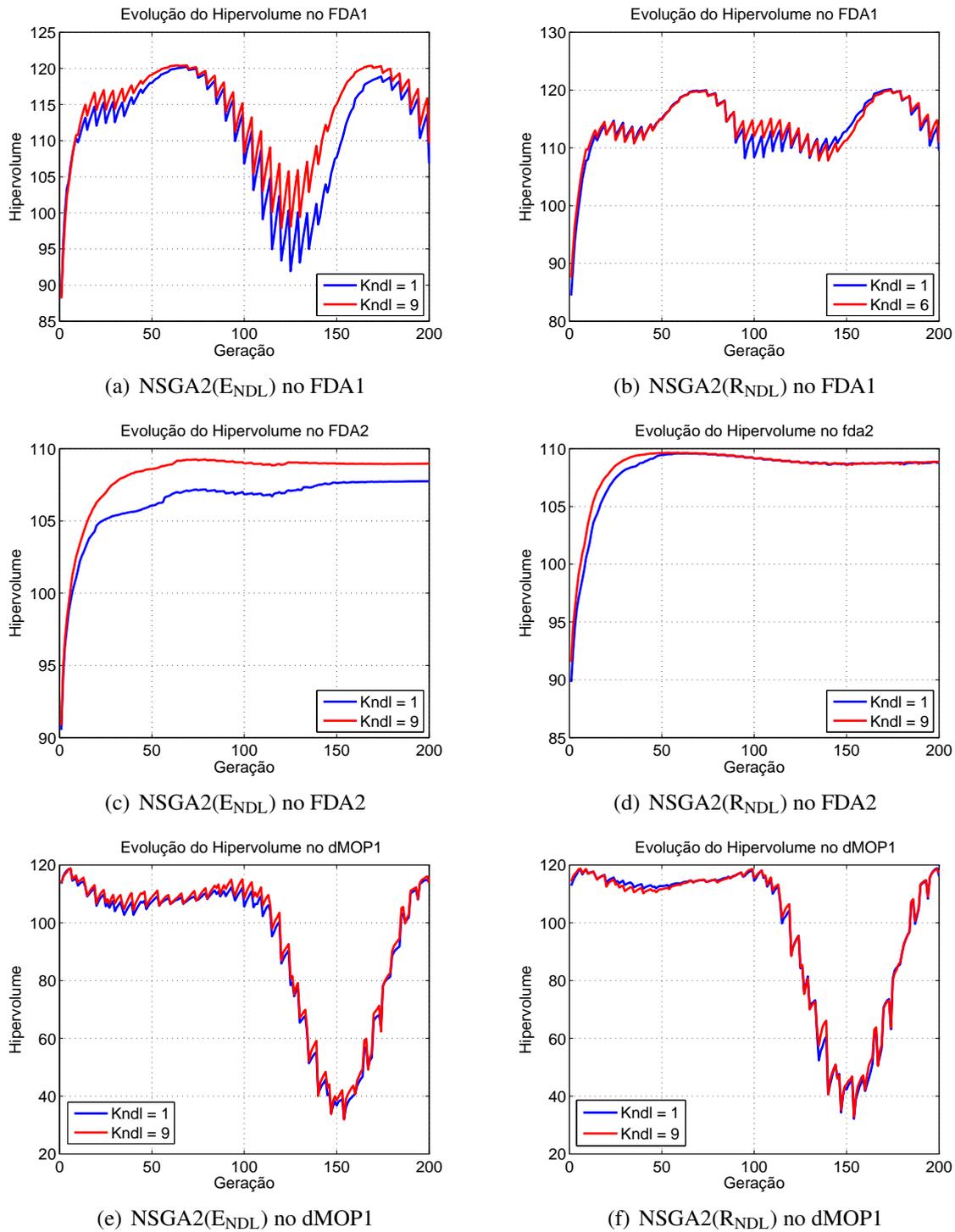


Figura 6.7 Evolução temporal do indicador de Hipervolume com e sem o NDL-DG em ambientes dinâmicos ($n_g = 10, t_g = 5$).

6.3. SELEÇÃO DE IMIGRANTES POR PNDS

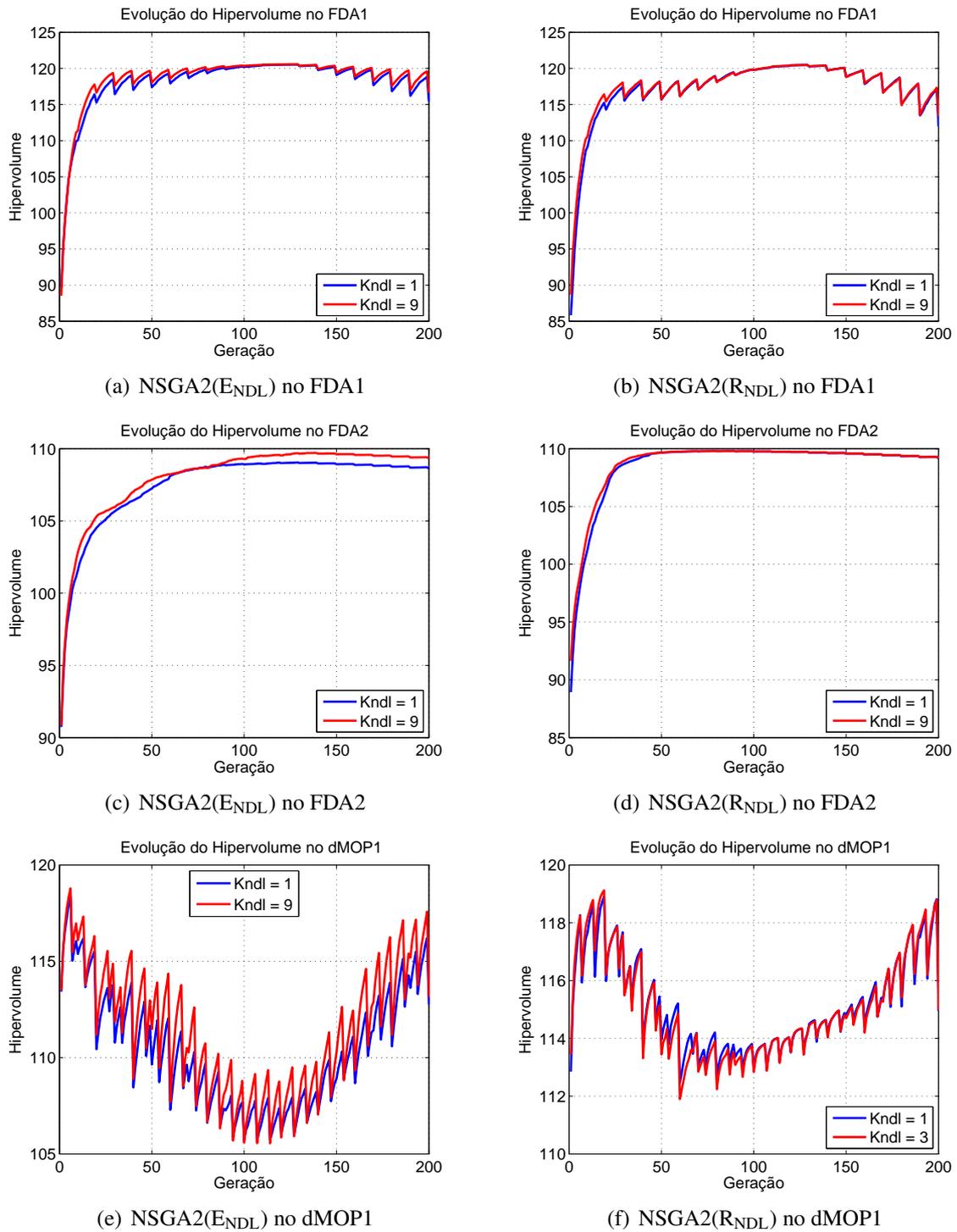


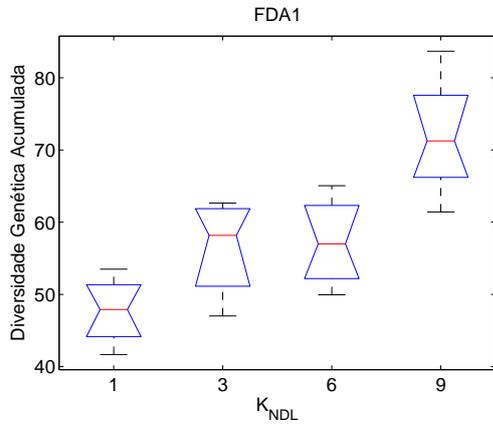
Figura 6.8 Evolução temporal do indicador de Hipervolume com e sem o NDL-DG em ambientes dinâmicos ($n_g = 10, t_g = 10$).

- Observa-se a situação (3) na maioria dos gráficos relativos à evolução de Hipervolume, tanto naqueles obtidos com o NSGA2(E_{NDL}), quanto com o NSGA2(R_{NDL}). Uma situação notável ocorre na evolução das populações com o NSGA2(E_{NDL}): apesar do Hipervolume das populações evoluídas com o NSGA2(E_{NDL}) ser levado a níveis praticamente idênticos àqueles relativos às populações evoluídas com o NSGA2(E) ($K_{NDL} = 1$) após mudanças temporais, a qualidade da exploração induzida pelo uso das NDLs mais do que compensa as quedas abruptas de Hipervolume em gerações subsequentes, levando a níveis de Hipervolume significativamente acima daqueles obtidos com $K_{NDL} = 1$. Esse padrão pode ser visualizado, por exemplo, na Figura 6.8 (e), de forma mais acentuada entre a 75ª a 150ª geração.

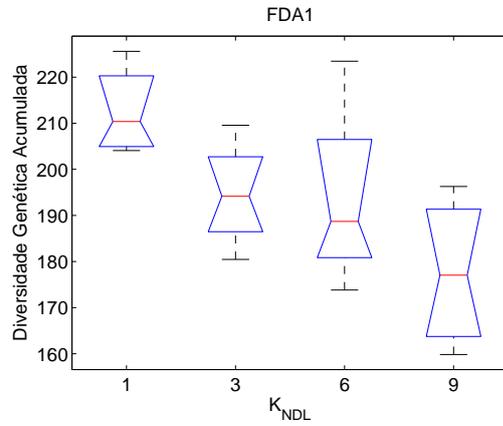
Os níveis medianos de diversidade genética (I_M) acumulada produzida pelos geradores NDL- DG_E e NDL- DG_R ao longo das 200 gerações nos MOPs dinâmicos estudados podem ser visualizados nos *box-plots* das Figuras 6.9 a 6.12. Observa-se que o NDL- DG_E tende a gerar maiores níveis de diversidade genética – quase sempre com significância estatística – com o incremento de K_{NDL} . Não se pode afirmar o mesmo dos resultados obtidos com o NDL- DG_R , apesar do mesmo quadro ter sido observado em 50% (6/12) dos cenários de otimização: no FDA1(10,5), FDA2(5,5), FDA2(5,10), dMOP1(5,10), dMOP1(10,5) e dMOP1(10,10). Uma observação decorrente dos resultados obtidos com o NSGA2(E_{NDL}) e com o NSGA2(R_{NDL}) (Figuras 6.9 a 6.12) é a de que, em geral, a inserção de imigrantes não-correlacionados (NDL- DG_R) é capaz de gerar maiores níveis de diversidade genética quando comparada à inserção de imigrantes correlacionados (NDL- DG_E). A análise a partir das Tabelas 6.4 a 6.7, em termos de diversidade genética (I_M) e diversidade de Frentes (H_N), leva aos seguintes relatos:

- No FDA1, em que o conjunto ótimo de Pareto (Ω^*) muda, mas a PF se mantém inalterada, o NDL- DG_E gera níveis de I_M significativamente maiores em, no mínimo, 71,9% das 200 gerações executadas em relação aos níveis obtidos sem o uso das paisagens (NSGA2(E)), apresentando resultados equivalentes (coluna \equiv) no restante das gerações. Quanto à diversidade de Frentes, H_N , o NDL- DG_E obtém níveis significativamente menores em no mínimo 8,0% das gerações, apresentando resultados equivalentes na maioria do restante das gerações. Já para o NDL- DG_R , com exceção do cenário FDA1(5,5), obtém-se o mesmo quadro para I_M em, no mínimo, 10,1% das gerações em relação aos níveis obtidos nos experimentos com o NSGA2(R), observando-se na maioria do restante das gerações resultados equivalentes. Para o caso FDA1(5,5), observa-se que o NDL- DG_R produz níveis *menores* de I_M com significância estatística em relação àqueles obtidos com o NSGA2(R)

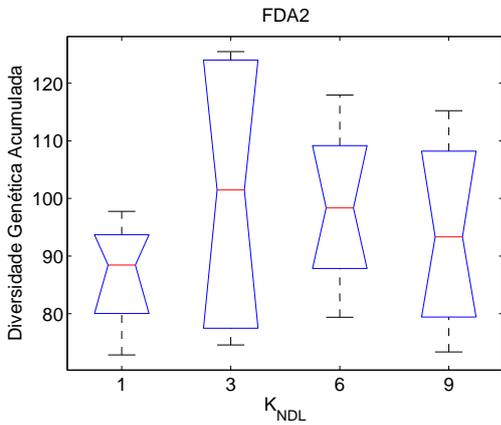
6.3. SELEÇÃO DE IMIGRANTES POR PNDS



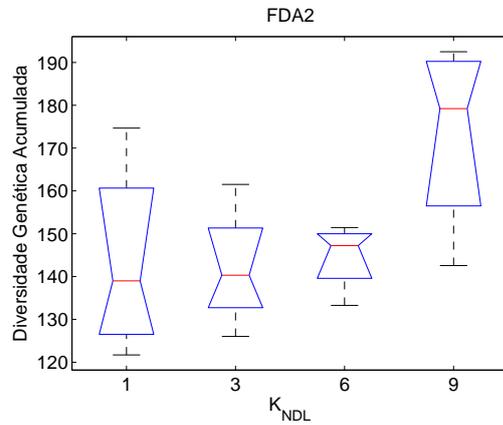
(a) NSGA2(E_{NDL}) no FDA1



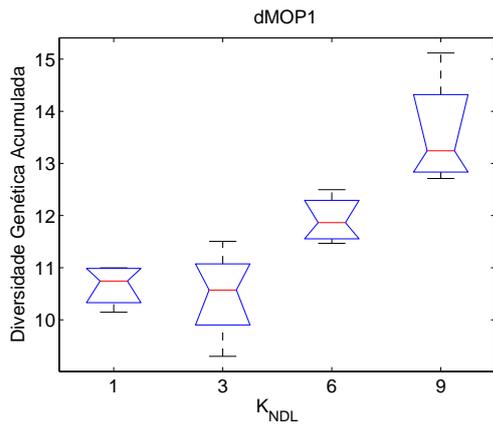
(b) NSGA2(R_{NDL}) no FDA1



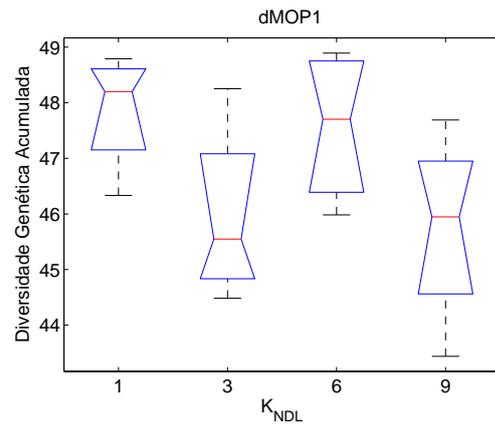
(c) NSGA2(E_{NDL}) no FDA2



(d) NSGA2(R_{NDL}) no FDA2

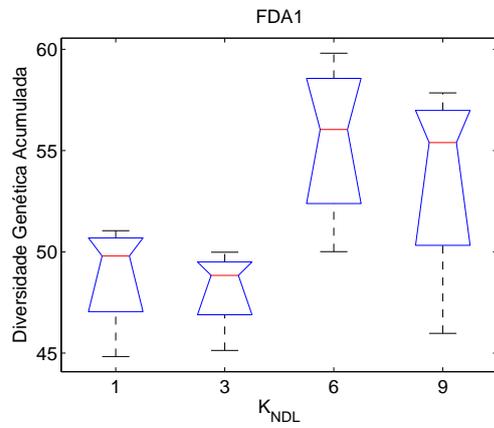


(e) NSGA2(E_{NDL}) no dMOP1

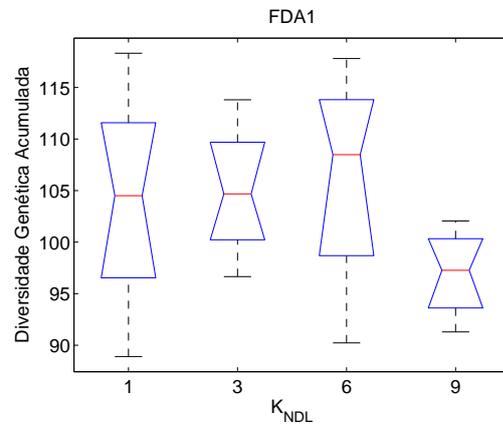


(f) NSGA2(R_{NDL}) no dMOP1

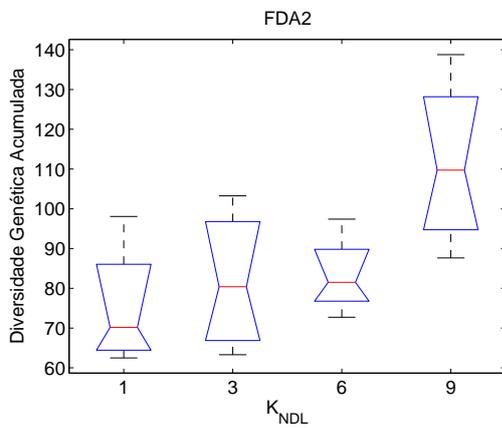
Figura 6.9 Box-plots para a diversidade genética (I_M) com e sem o NDL-DG em ambientes dinâmicos ($n_g = 5, t_g = 5$).



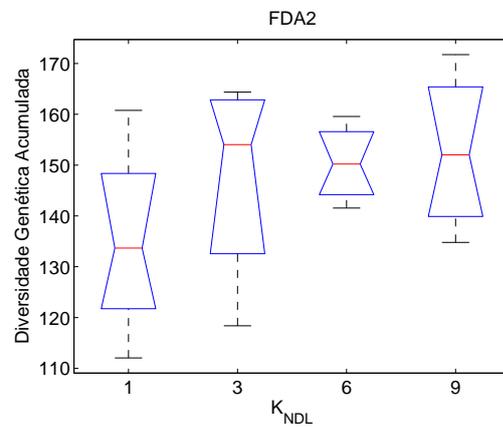
(a) NSGA2(E_{NDL}) no FDA1



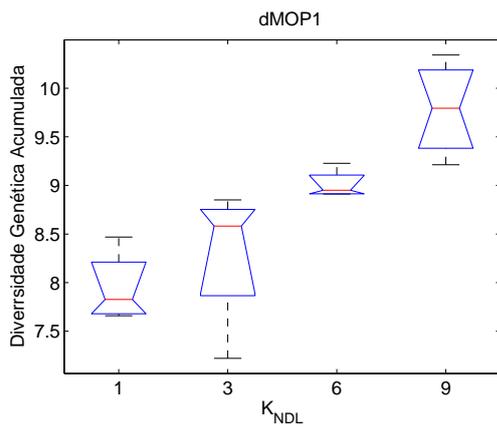
(b) NSGA2(R_{NDL}) no FDA1



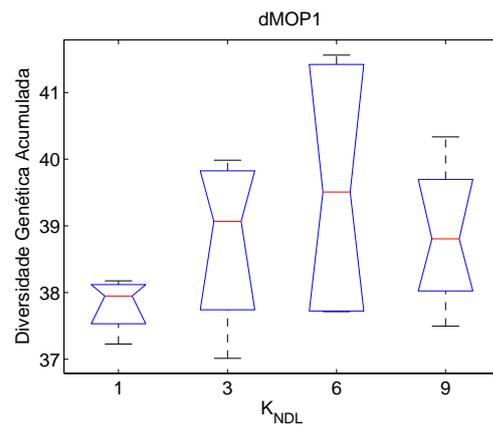
(c) NSGA2(E_{NDL}) no FDA2



(d) NSGA2(R_{NDL}) no FDA2



(e) NSGA2(E_{NDL}) no dMOP1



(f) NSGA2(R_{NDL}) no dMOP1

Figura 6.10 Box-plots para a diversidade genética (I_M) com e sem o NDL-DG em ambientes dinâmicos ($n_g = 5, t_g = 10$).

6.3. SELEÇÃO DE IMIGRANTES POR PNDS

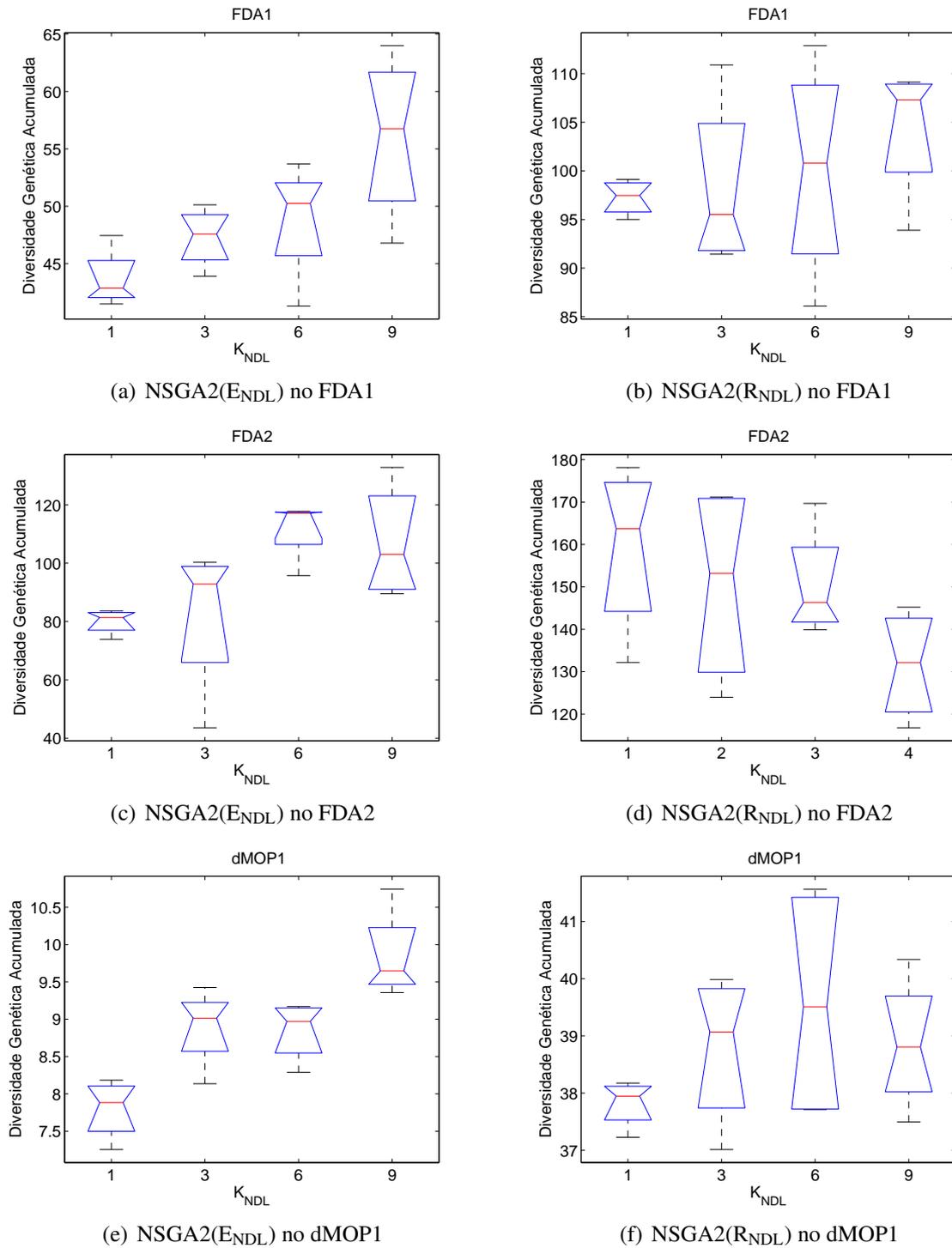


Figura 6.11 Box-plots para a diversidade genética (I_M) com e sem o NDL-DG em ambientes dinâmicos ($n_g = 10, t_g = 5$).

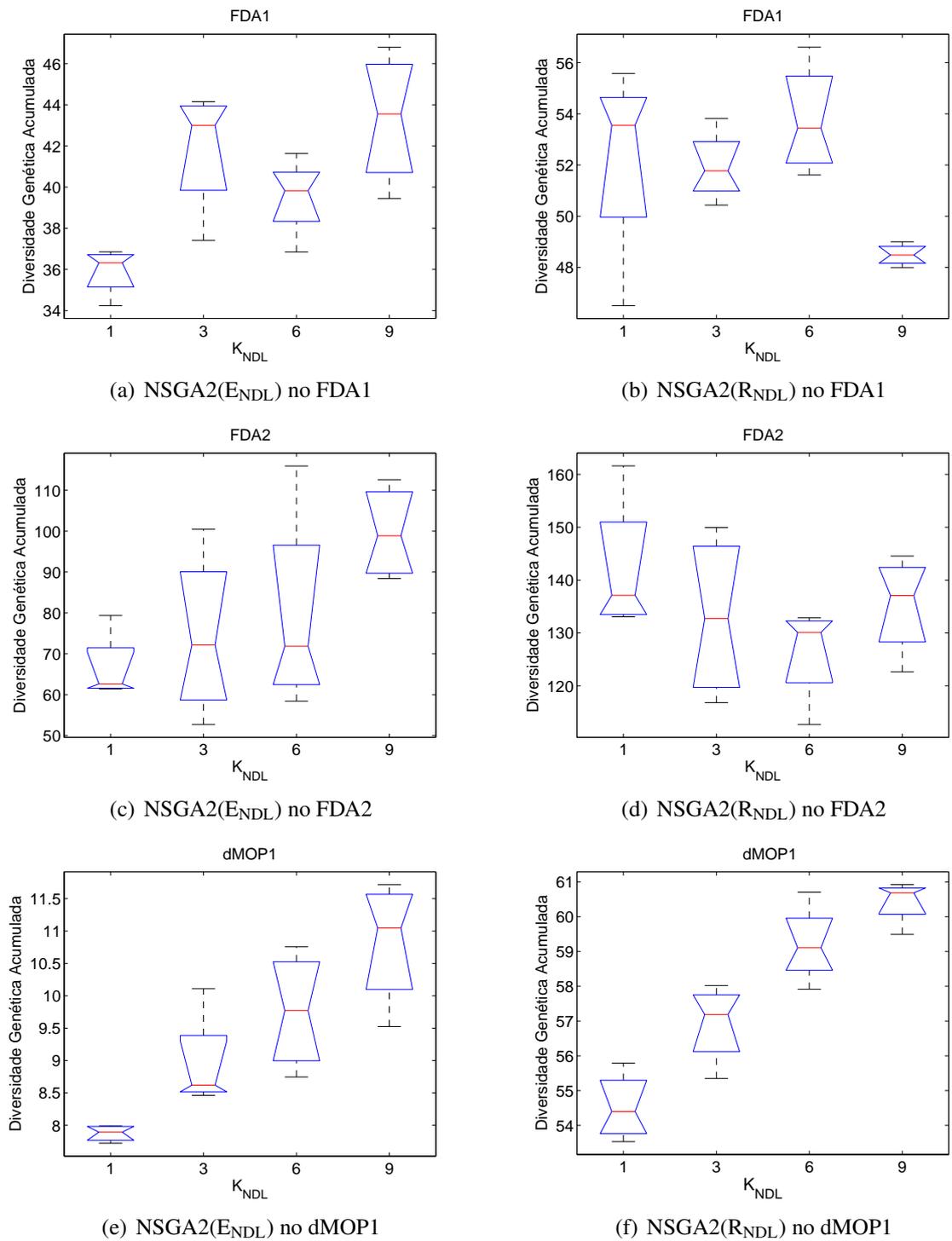


Figura 6.12 Box-plots para a diversidade genética (I_M) com e sem o NDL-DG em ambientes dinâmicos ($n_g = 10, t_g = 10$).

em 24,6% das gerações e, mesmo assim, apresenta ganhos significativos em termos de Hipervolume. Em termos de H_N , o NDL-DG_R produz níveis significativamente maiores em, no mínimo, 2,0% das gerações, equivalendo-se ao NSGA2(R) nesse quesito na maioria do restante das gerações;

- No FDA2, em que Ω^* se mantém inalterado e a PF muda com o tempo, o NDL-DG_E gera níveis de I_M significativamente maiores em, no mínimo, 49,2% das 200 gerações executadas em relação aos níveis obtidos com o NSGA2(E), apresentando resultados equivalentes na maioria das gerações restantes. Quanto à H_N , obtém níveis significativamente menores em, no mínimo, 2,5% das gerações, equivalendo-se ao NSGA2(E) nesse quesito na maioria das gerações restantes. Já o NDL-DG_R obtém níveis de diversidade genética estatisticamente equivalentes em no mínimo 80,9% das gerações em relação aos níveis obtidos com o NSGA2(R), apresentando níveis ora significativamente inferiores, ora significativamente superiores no restante das gerações. Em termos de H_N , o NDL-DG_R produz níveis significativamente maiores em, no mínimo, 2,5% das gerações, equivalendo-se ao NSGA2(R) nesse quesito na maioria do restante das gerações;
- No dMOP1, em que (Ω^*) muda, a PF se mantém inalterada e a variável de decisão que controla o primeiro objetivo muda, o NDL-DG_E gera níveis de I_M significativamente maiores em, no mínimo, 59,8% das 200 gerações executadas em relação aos níveis obtidos com o NSGA2(E), apresentando resultados equivalentes na maioria das gerações restantes. Quanto à H_N , obtém níveis significativamente menores em, no mínimo, 23,6% das gerações, equivalendo-se ao NSGA2(E) nesse quesito na maioria das gerações restantes. Já o NDL-DG_R obtém níveis de I_M estatisticamente equivalentes em no mínimo 42,2% das gerações em relação aos níveis obtidos com o NSGA2(R), apresentando níveis significativamente maiores na maioria das gerações restantes. Em termos de H_N , o NDL-DG_R produz níveis significativamente maiores em, no mínimo, 21,0% das gerações, equivalendo-se ao NSGA2(R) nesse quesito na maioria das gerações restantes.

A primeira observação relevante quanto à relação entre as diversidades genética e de Frentes, descrita nos relatos anteriores, é a de que, ao contrário do que se poderia esperar, o aumento em I_M induzido pelo NDL-DG não acarreta, em geral, no incremento em H_N . De fato, como exemplifica a Figura 6.13, mesmo tendo gerado níveis de diversidade genética maiores ou iguais àqueles produzidos sem o uso de NDLs (partes (b) e (d) da Figura), tanto o NDL-DG_E quanto o NDL-DG_R obtêm por vezes menores níveis

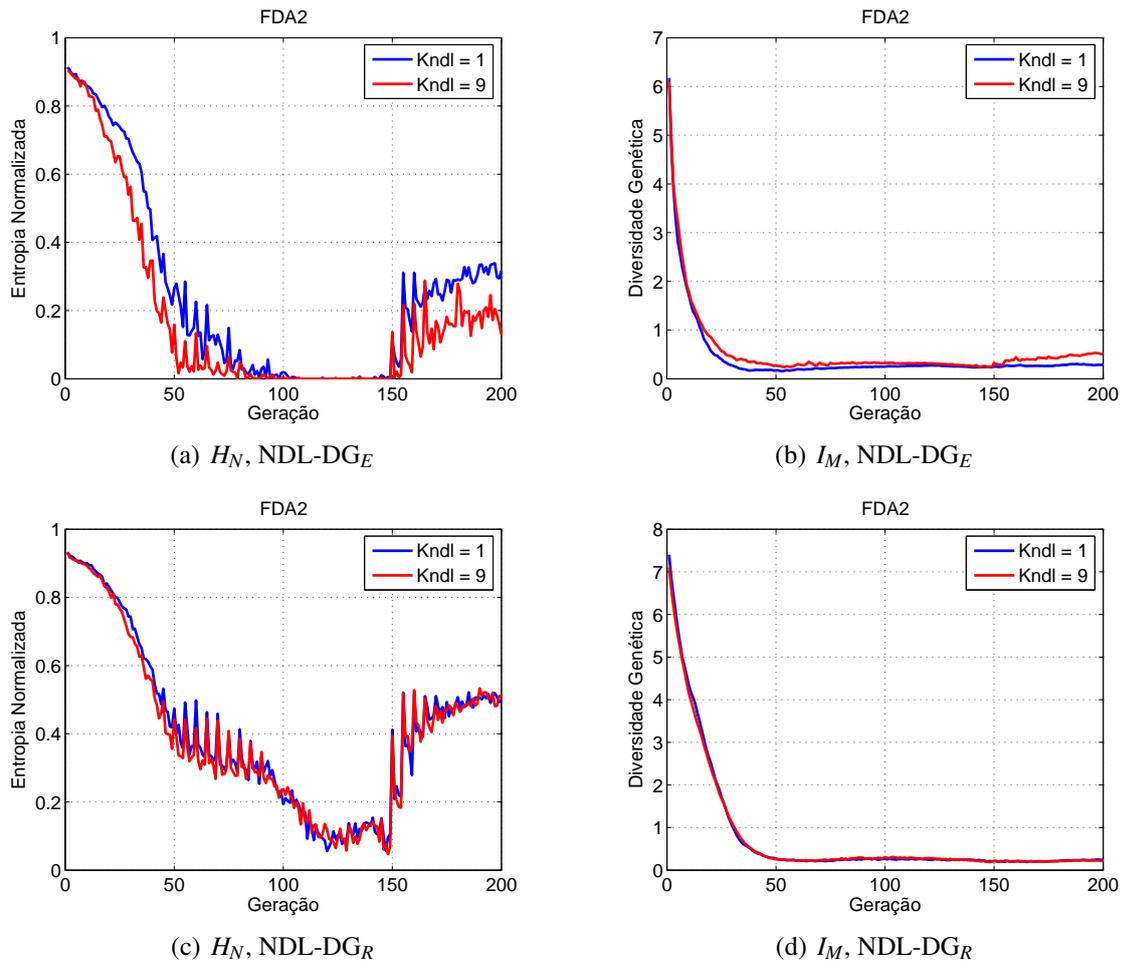


Figura 6.13 Evolução da Entropia Normalizada (H_N) e da Diversidade Genética (I_M) com e sem o NDL-DG no FDA2 ($n_g = 5, t_g = 5$).

de diversidade de frente (partes (a) e (c) da Figura). Especificamente na parte (a) da Figura, nota-se, entre as gerações 100 e 150, que H_N atinge o nível zero, indicando que, na média, e entre as gerações compreendidas naquele intervalo temporal, praticamente todas as soluções mantidas nas populações evoluídas com o NSGA2(E_{NDL}) são não-dominadas. Esse efeito é observado com maior frequência no FDA2, o qual, por ser um MOP dinâmico de Tipo 1, apresenta a evolução de Hipervolume mais parecida com aquelas obtidas nos experimentos com ambientes estáticos. Esse resultado não apenas sugere que os geradores propostos são capazes de manter níveis adequados de diversidade genética nos problemas estudados, mas também que os geradores baseados em NDLS intensificam – dado os valores mais baixos de H_N – a exploração do espaço de busca na proximidade do conjunto de não-dominadas, \mathcal{F}_1 (ver discussão sobre as propriedades de H_N na página 100). Ademais – principalmente quanto aos níveis de I_M consistentemente maiores e quanto aos níveis de H_N consistentemente menores nos experimentos com o NDL-DG $_E$ –, os resultados das Tabelas 6.4 a 6.7 sugerem a existência de padrões de correlação estatística entre os dois tipos de diversidade aferidas (genética e de Frentes) e o indicador de Hipervolume *offline* dos conjuntos de não-dominadas evoluídos com o NSGA2(E_{NDL}) e o NSGA2(R_{NDL}).

Motivadas por essas observações, apresentam-se a seguir análises das correlações estatísticas entre (i) I_M e $\overline{\mathcal{S}}_h$; e (ii) H_N e $\overline{\mathcal{S}}_h$. Note-se que os valores apresentados nas próximas seções são as médias dos coeficientes de Pearson computados entre os níveis de diversidade e Hipervolume entre duas gerações subseqüentes, obtidos a partir das populações evoluídas com os algoritmos estudados ao longo de todo o processo temporal (200 gerações). O objetivo do estudo é a visualização do comportamento das interações entre diversidade e Hipervolume com o incremento da pressão de seleção dos imigrantes com base nas probabilidades de não-dominância (controlada por K_{NDL}).

6.3.2 Correlação Entre Diversidade Genética e Hipervolume

A primeira observação da média dos valores de correlação de Pearson (Figura 6.14) é a de que, em geral, na ausência das NDLS ($K_{\text{NDL}} = 1$), a diversidade genética correlaciona-se positivamente com o Hipervolume na dinâmica de aproximação da PF com maior intensidade nas populações evoluídas com o NSGA2(E) do que naquelas evoluídas pelo NSGA2(R), com exceção do problema FDA2, no qual os níveis médios de correlação são, em sua maioria, negativos no NSGA2(R) e positivos no NSGA2(E) para cada combinação de n_g e t_g . Observa-se ainda que, com o incremento de K_{NDL} , os níveis de correlação entre I_M e \mathcal{S}_h tendem a aumentar no FDA2 e no dMOP1 (embora fiquem

praticamente constantes na parte (c) da Figura), e apresentam tendência de queda no FDA1, principalmente para o NSGA2(E_NDL). O fato da correlação entre diversidade genética e Hipervolume diminuir nos experimentos com o NSGA2(E_NDL) no FDA1, somado aos fatos de que, com o incremento de K_{NDL} , (1) \mathcal{S}_h aumenta significativamente; e (2) I_M também aumenta significativamente, pode ser um indicativo de que a verdadeira razão para os ganhos de \mathcal{S}_h seja a de uma exploração mais eficiente de regiões promissoras, devido à influência das NDLs, reduzindo, em contrapartida, a correlação dos níveis absolutos de I_M na qualidade do conjunto de não-dominadas. Já nos demais problemas (FDA2 e dMOP1), o fato da correlação positiva entre I_M e \mathcal{S}_h tender a aumentar com o incremento de K_{NDL} pode significar (especialmente no dMOP1 para o qual as correlações atingiram os maiores níveis absolutos), por outro lado, que a manutenção de níveis mínimos de diversidade genética é mais importante naqueles problemas e, em conjunto com uma exploração mais eficiente de regiões promissoras, repercute positivamente para o aumento da qualidade média do conjunto de não-dominadas.

6.3.3 Correlação Entre Diversidade de Frentes e Hipervolume

Por fim, busca-se visualizar o efeito do incremento de K_{NDL} nas interações entre H_N e \mathcal{S}_h . Assim, da Figura 6.15, observa-se que H_N correlaciona-se negativamente com \mathcal{S}_h em todos os cenários de otimização dinâmica estudados, indicando que, na média, principalmente para $K_{\text{NDL}} = 1$, a manutenção de menores valores de H_N são preferíveis para o aumento da qualidade do conjunto de não-dominadas ao longo do processo evolucionário. Como discutido na página 100, $H_N = 0$ representa um nível máximo de exploração de soluções não-dominadas e $H_N = 1$ representa a divisão equânime dos indivíduos da população entre todas as Frentes existentes. Portanto, menores valores de H_N indicam um aumento do nível de exploração em torno das primeiras Frentes da população. Com o incremento de K_{NDL} , percebe-se, em praticamente todos os cenários de otimização, que as correlações negativas tendem, em geral, a se aproximar de zero, indicando perda de correlação entre H_N e \mathcal{S}_h . Novamente, uma hipótese para esse resultado é que o aumento da exploração em regiões promissoras induzido pelo incremento da influência das NDLs no processo de inserção de imigrantes contribui para a redução da importância de fatores como os níveis de H_N na obtenção de conjuntos $|\mathcal{F}_1|$ de maior qualidade média.

6.3. SELEÇÃO DE IMIGRANTES POR PNDS

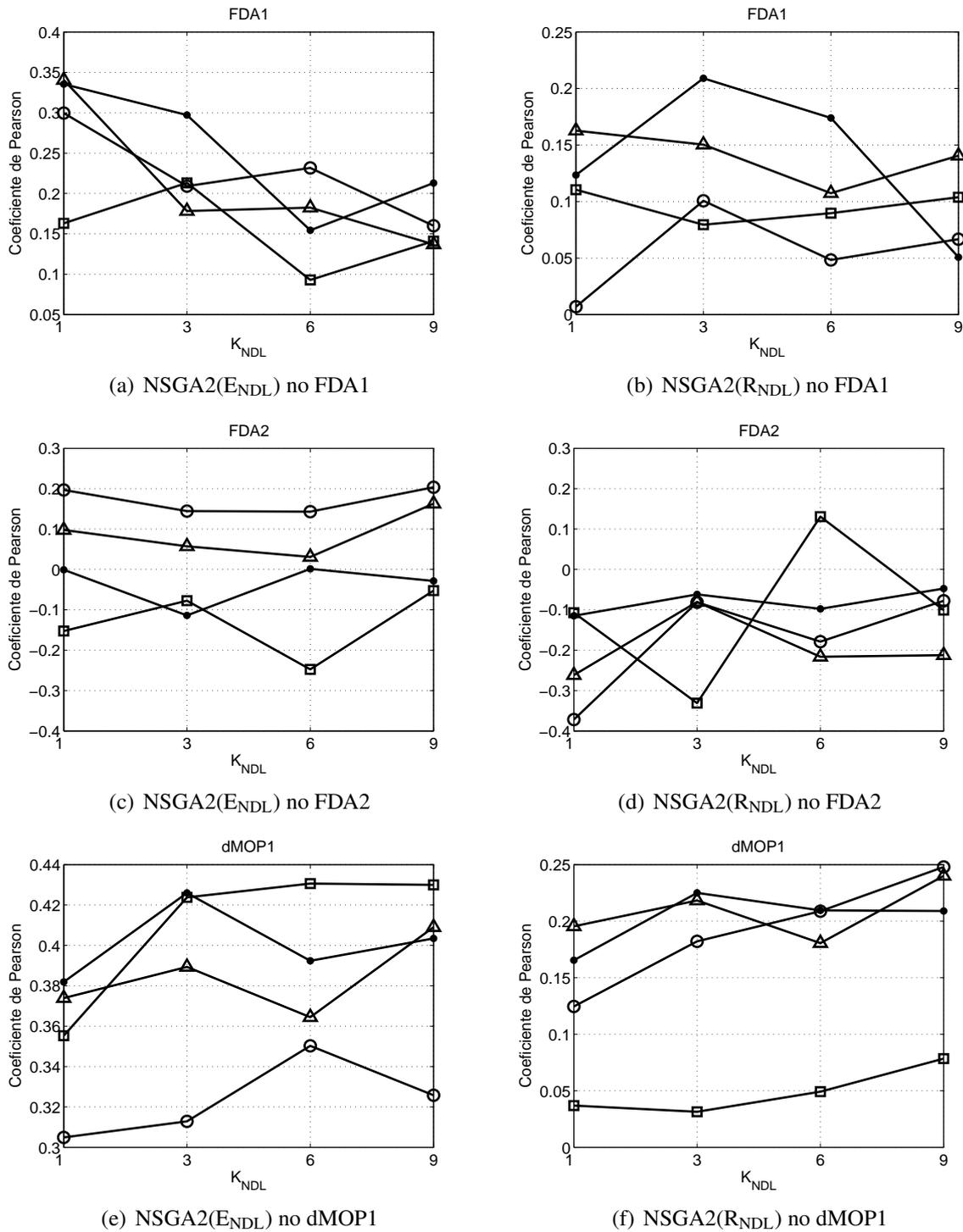


Figura 6.14 Variação da correlação de Pearson entre I_M e J_H com o incremento de K_{NDL} em MOPs dinâmicos: $n_g = 5, t_g = 5$ (\circ); $n_g = 5, t_g = 10$ (\triangle); $n_g = 10, t_g = 5$ (\bullet) e $n_g = 10, t_g = 10$ (\square).

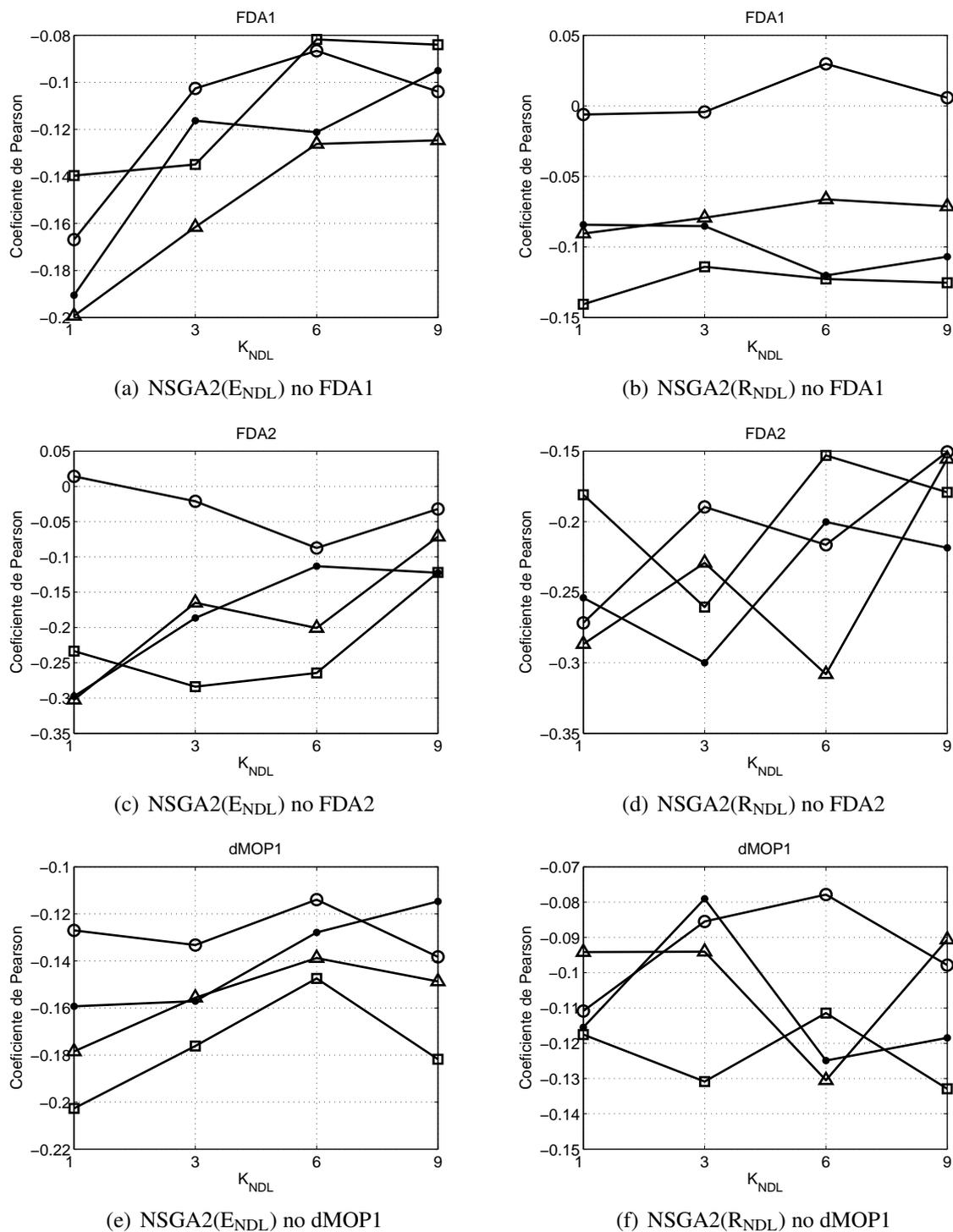


Figura 6.15 Variação da correlação de Pearson entre H_N e \mathcal{I}_H com o incremento de K_{NDL} em MOPs dinâmicos: $n_g = 5, t_g = 5$ (\circ); $n_g = 5, t_g = 10$ (\triangle); $n_g = 10, t_g = 5$ (\bullet) e $n_g = 10, t_g = 10$ (\square).

6.4 Discussão Final e Síntese

Esse Capítulo propôs a incorporação dos geradores de diversidade por inserção de imigrantes ao algoritmo NSGA2, aplicando-o a problemas de otimização multiobjetivo (MOPs) estáticos e dinâmicos, e apresentou um estudo sistemático preliminar dos efeitos de variação paramétrica inerentes aos modelos considerados. Por meio da variação da taxa de imigração (α), mostraram-se os efeitos do incremento do número de imigrantes inseridos na população na qualidade média do conjunto de soluções não-dominadas, \mathcal{F}_1 , medida em termos de Hipervolume *offline* ($\overline{\mathcal{I}}_h$). Desse estudo, conclui-se:

- Ambientes estáticos – apesar de, na prática, a inserção de imigrantes raramente ser utilizada em MOPs estáticos, o estudo apresentado nesse capítulo com o conjunto ZDT sugere que, em fases estacionárias de longa duração de um problema dinâmico, a taxa de imigração deve ser mantida próxima de zero, sob pena de degradar a qualidade de \mathcal{F}_1 ;
- Ambientes dinâmicos – dentre o conjunto de valores testados, na aplicação da estratégia de imigrantes não-correlacionados, NSGA2(R), o incremento da taxa de imigração atinge o nível crítico em $\alpha = 0,3$, a partir do qual $\overline{\mathcal{I}}_h$ apresenta tendência de queda, recomendando-se, portanto, a utilização de valores menores que 0,5 para aquele parâmetro. Já na aplicação da estratégia de imigrantes correlacionados, NSGA2(E), recomenda-se o uso de taxas de imigração elevadas, de modo a gerar diversidade e promover o caminharmento à Frente de Pareto (PF);

Em seguida, avaliaram-se os efeitos da incorporação dos geradores NDL-DG, os quais selecionam os imigrantes inseridos na população por probabilidade de não-dominância, no algoritmo NSGA2. Em termos de Hipervolume *offline*, e em relação à ausência da seleção de imigrantes por NDLs, observou-se que o uso do NDL-DG leva à obtenção de soluções não-dominadas de maior qualidade média com significância estatística em $\approx 79\%$ das 24 combinações entre parâmetros de severidade (n_g) e periodicidade (t_g), MOPs dinâmicos, e estratégias de geração de imigrantes estudadas. Além disso, observou-se, com o emprego do NDL-DG, que o número de gerações em que os níveis de \mathcal{I}_h são estatisticamente maiores do que aqueles observados sem o uso de NDLs depende das características temporais do MOP dinâmico. Mesmo assim, em todos os três MOPs dinâmicos estudados, o número de gerações em que isso acontece foi expressivo, sendo, em média, de $\approx 82\%$ nos experimentos com imigrantes correlacionados e de $\approx 26\%$ com imigrantes não-correlacionados. O fato dos ganhos obtidos com o NSGA2(R_{NDL}) não

terem sido tão expressivos quanto aqueles alcançados com o NSGA2(E_{NDL}) pode ser devido aos diferentes efeitos da sinergia entre as regiões no espaço de objetivos onde os imigrantes são gerados e às probabilidades de não-dominância estimadas naquele espaço, merecendo estudos mais aprofundados no futuro.

Do estudo em MOPs dinâmicos dos efeitos do incremento da pressão de seleção de imigrantes por probabilidade de não-dominância, controlada por K_{NDL} , conclui-se:

- Imigrantes correlacionados – aumentar a pressão de seleção dos imigrantes por NDLS leva, em 100% dos casos estudados, a aumentos em $\overline{\mathcal{F}_h}$ com significância estatística. Portanto, recomenda-se o incremento de K_{NDL} em conformidade com o conjunto de valores testados nesse estudo.
- Imigrantes não-correlacionados – incrementar K_{NDL} não leva necessariamente a ganhos em $\overline{\mathcal{F}_h}$ com significância estatística em todos os cenários estudados – $\approx 58\%$ (7/12). A partir da análise dos resultados, recomenda-se o uso de NDLS em MOPs dinâmicos em que não há desbalanceamento do número de variáveis mapeadas por cada função objetivo.

Quanto aos níveis de diversidade genética (I_M) e diversidade de Frentes (H_N) produzidos para cada um dos quatro valores de K_{NDL} estudados, observou-se, dos experimentos com o NSGA2(E_{NDL}), que maiores valores de K_{NDL} geralmente levam a maiores níveis de I_M e a menores níveis de H_N . Já nos resultados obtidos com o NSGA2(R_{NDL}), observou-se quanto à I_M que, em alguns cenários de otimização, com o incremento de K_{NDL} , o efeito contrário é produzido: a variação de I_M apresenta tendência de queda. Esse resultado motivou o estudo da correlação entre (i) I_M e $\overline{\mathcal{F}_h}$; e entre (ii) H_N e $\overline{\mathcal{F}_h}$. Desse estudo, conclui-se que, em geral, o incremento de K_{NDL} reduz a importância relativa dos níveis de diversidade, tanto genética quanto de Frentes, na obtenção de conjuntos de soluções não-dominadas de maior qualidade média nos MOPs dinâmicos estudados, embora alguns casos em que o contrário ocorreu tenham sido observados.

Portanto, pelo exposto, atinge-se o objetivo principal da pesquisa, a qual buscou a identificação de métodos de geração de diversidade os quais fossem capazes não apenas de reverter dos efeitos negativos do incremento dos níveis de dispersão populacional na dinâmica de aproximação à PF , mas também que promovessem a exploração mais eficiente de regiões promissoras, levando à obtenção de conjuntos de soluções não-dominadas de maior qualidade ao longo de um processo de otimização dinâmico e multiobjetivo.

7

Conclusão e Perspectivas de Investigação

O futuro influencia o presente tanto quanto o passado.

– Friedrich Nietzsche

A pesquisa relatada nessa dissertação abordou aspectos inerentes ao projeto de Algoritmos Evolucionários Multiobjetivos (MOEAs), em especial a geração de diversidade – bem como a influência dessa – no caminhamento em direção à superfície de compromisso onde as soluções ótimas de um Problema de Otimização Multiobjetivo (MOP) são encontradas, conhecida como Frente de Pareto (*PF*). Por meio de experimentos, avaliou-se a incorporação de geradores de diversidade baseados na inserção de soluções atípicas (*imigrantes*) no algoritmo NSGA2 aplicado a MOPs estáticos e dinâmicos. Propôs-se, então, a incorporação de critérios de seleção de imigrantes por probabilidades estimadas de não-dominância aos geradores de diversidade para o problema de atenuar os efeitos negativos do aumento dos níveis de dispersão em populações de soluções candidatas evoluídas por MOEAs na qualidade média do conjunto de soluções não-dominadas, \mathcal{F}_1 . Os estudos de variação paramétrica das diferentes versões do modelo proposto, NDL-DG, incorporadas ao NSGA2, mostraram que, não apenas a solução proposta atende aos requisitos do problema descrito, mas também provê a flexibilidade necessária para ser melhorada ainda mais e aplicada a outros problemas em aberto na otimização multiobjetivo evolucionária. Cabe ressaltar que, apesar dos estudos experimentais relatados no Capítulo 6 terem considerado o caso específico da aplicação do algoritmo NSGA2, espera-se que os métodos de geração de diversidade dirigida desenvolvidos nessa dissertação sejam potencialmente aplicáveis para qualquer MOEA.

Uma síntese dos principais resultados obtidos é descrita na próxima Seção, enquanto as sugestões de trabalhos futuros e as publicações decorrentes dessa dissertação encontram-se nas Seções finais.

7.1 Principais Contribuições

Como principais contribuições originadas da pesquisa, enumeram-se:

1. A caracterização da correlação estatística entre a diversidade genética e a aptidão – medida em termos do indicador de Hipervolume – dos indivíduos nas populações de soluções candidatas em MOEAs (Capítulos 4 e 6);
2. A extensão dos modelos de inserção de imigrantes utilizados em AEs mono-objetivo para MOEAs (Capítulos 5 e 6);
3. A aplicação de generalizações de estatísticas de ordem multivariada para caracterizar probabilisticamente as relações de dominância entre as soluções candidatas em MOEAs (Capítulo 5);
4. A concepção e formalização do conceito de *Paisagens de Não-Dominância* (NDLs) para guiar decisões de inserção de imigrantes em populações (Capítulo 5);
5. A obtenção do Teorema 1, o qual sugere o uso das NDLs como um tipo de gradiente estocástico (Capítulo 5).
6. A incorporação das NDLs nos modelos estendidos de geração de diversidade por inserção de imigrantes em MOEAs para dirigir a exploração de regiões promissoras em MOEAs (Capítulo 5).

Dentre as contribuições decorrentes da linha de investigação principal, listam-se:

1. Uma taxonomia alternativa para a contextualização das estratégias de inserção de imigrantes em termos de dependências estatísticas dos modelos probabilísticos utilizados para a geração de soluções atípicas (Capítulo 4);
2. A modelagem probabilística elementar de um regime de imigração assimétrico, por meio de variáveis aleatórias contínuas (Capítulo 4);
3. A determinação do valor esperado de diversidade genética no modelo de imigração concebido e a caracterização de uma condição suficiente para que tais valores sejam idênticos em dois regimes com taxas de imigração complementares (Capítulo 4);
4. A proposta do uso da entropia normalizada para caracterizar um efeito denominado nesse trabalho como *perda de frentes*, relevante para o entendimento da dinâmica do caminhar à *PF* em MOEAs (Capítulo 6);

5. A formulação de uma nova classe de problemas dinâmicos multiobjetivo, adaptado a partir de problemas já existentes (Capítulo 6).

7.2 Limitações do Trabalho

Apesar do modelo de geração de diversidade proposto – NDL-DG – mostrar-se promissor em evoluir conjuntos \mathcal{F}_1 de maior qualidade média em MOPs dinâmicos, a escolha particular de como incorporar as NDLs aos geradores de imigrantes incorre na seguinte limitação: por se ter optado, dentre um conjunto de $K_{\text{NDL}} > 1$ imigrantes gerados, pela inserção do imigrante com a maior probabilidade de não-dominância, é possível que o forte viés induzido por esse critério leve a problemas quanto à manutenção de diversidade em certos estágios do processo de otimização, podendo essa hipótese, inclusive, explicar o motivo do NSGA2(R_{NDL}) não ter apresentado resultados tão contundentes quanto aqueles obtidos com o NSGA2(E_{NDL}). Logo, é possível que estratégias de seleção *proporcionais* às probabilidades de não-dominância realizem um melhor balanço entre os efeitos temporais positivos e negativos da escolha de imigrantes baseadas em NDLs.

Em consequência dessa dissertação ter se concentrado na prova de princípio de que a geração de diversidade em concomitância à atenuação dos efeitos negativos da elevação dos níveis de dispersão populacional no caminhamento à *PF* pode ser alcançada pelo uso de NDLs, a obtenção de um modelo com baixo custo computacional tornou-se um objetivo secundário na pesquisa. Dessa forma, cabe ressaltar que o custo computacional adicional em passar de $K_{\text{NDL}} = 1$ (ausência de NDLs) para $K_{\text{NDL}} = 2$ é de estimar a *pdf* da população no espaço de objetivos e computar a probabilidade de não dominância para dois imigrantes produzidos a cada geração. Mesmo assim, o tempo necessário para a execução de 200 gerações em um problema dinâmico com as parametrizações utilizadas nesse trabalho para o maior valor de K_{NDL} estudado, foi de apenas (aproximadamente e em média) dois minutos em um *laptop* de uso pessoal com processador Intel Core2 Duo (utilizando apenas um núcleo de processamento) e dois Gigabytes de memória RAM. Diversas estratégias mais eficientes podem existir para alcançar o mesmo objetivo a partir do uso de probabilidades de não-dominância, não apenas em função das escolhas particulares dos algoritmos de estimação e de estruturas de dados mais eficientes, mas também em termos da incorporação das NDLs no processo de amostragem do espaço de busca, eliminando a necessidade de abordagens do tipo *gerar e testar*, como a que foi proposta nessa pesquisa. Discutem-se esses e outros temas na próxima Seção.

7.3 Sugestões para a Continuidade do Trabalho

De forma geral, faz-se necessário estender o estudo experimental apresentado nessa dissertação para outros tipos de MOPs dinâmicos, que não apenas apresentem dinâmicas temporais diferentes, mas também incorporem restrições sobre as variáveis e um maior número de funções-objetivo. A avaliação do NDL-DG como gerador de diversidade em MOEAs aplicado a MOPs dinâmicos de graus de dificuldade mais elevados do que os MOPs existentes permitirá uma melhor avaliação do papel da diversidade gerada por esses modelos naqueles problemas, além de uma melhor indicação de como os modelos propostos podem ser melhorados. Além disso, a incorporação do NDL-DG em outros MOEAs de estado da arte é altamente recomendada como trabalhos futuros pois permitirá melhor avaliarem-se as interações entre a diversidade gerada por esses modelos e entre os diferentes mecanismos de manutenção de diversidade comumente empregados nos operadores de variação, critérios de seleção de sobreviventes e estratégias de atribuição de aptidão.

7.3.1 Controle de Parâmetros do NDL-DG

Como visto no Capítulo 5, o NDL-DG possui dois parâmetros ajustáveis (além da escolha do modelo de geração de imigrantes), a taxa de imigração, α , a qual define o número de soluções da população substituídas pelos imigrantes gerados, e o número de imigrantes gerados para cada imigrante efetivamente inserido na população, K_{NDL} , o qual permite o ajuste da influência da seleção de imigrantes por probabilidades de não-dominância no processo de geração de diversidade dirigida proposto. A auto-adaptação desses dois parâmetros, além de eliminar a necessidade de ajuste manual, pode se mostrar importante para melhorar ainda mais o desempenho dos MOEAs que empregam o NDL-DG para a geração de diversidade em ambientes dinâmicos.

Yu *et al.* (2010) investigaram a influência das taxas de imigração na performance de algoritmos evolucionários para otimização dinâmica mono-objetivo. A partir das investigações iniciais, sugeriram um mecanismo de auto-adaptação da taxa de imigração de tal forma que um balanço adequado de exploração/exploração fosse obtido. Para a análise dos resultados experimentais, uma nova medida de eficiência foi introduzida, a saber, o “efeito de imigração”, definido como a proporção de imigrantes gerados com aptidão acima da aptidão mediana do conjunto de imigrantes. Com base nessa nova medida, o controle da taxa de imigração foi realizado ao se comparar o valor atual do efeito de imigração com um limiar pré-definido, em que a taxa de migração é aumentada

aditivamente para um efeito positivo, i.e., acima do limiar; diminuída aditivamente para um efeito negativo; e mantida constante, caso contrário. A partir da observação dos resultados, os autores concluem que o mecanismo de controle implementado fornece a conveniência de dispensar o ajuste manual da taxa de imigração fixa. Entretanto, os resultados obtidos com o controle apresentado não foram satisfatórios no sentido de não apresentarem ganhos de desempenho em relação às estratégias de utilização de taxas de imigração constantes ao longo de todo o processo evolucionário. Pelo contrário: em alguns casos a estratégia de controle degradou a performance dos AGs estudados. Esse resultado não exclui, porém, a possibilidade do projeto de um mecanismo auto-adaptativo robusto em relação à taxa de imigração.

Dessa forma, há um longo e promissor caminho a ser explorado quanto à investigação de mecanismos eficientes de auto-adaptação, não apenas da taxa de imigração, mas dos outros parâmetros incorporados nos modelos existentes de geração de diversidade em MOEAs.

A seguir, algumas direções de pesquisa para melhorar o desempenho do NDL-DG quanto à geração de imigrantes que promovam diversidade genética e explorem ao mesmo tempo regiões promissoras em MOPs dinâmicos são apontadas. Sugestões para a extensão da aplicabilidade das NDLs a outros problemas em aberto na otimização evolucionária multiobjetivo também são discutidas.

7.3.2 Modelos Alternativos de Geração de Imigrantes por NDLs

A primeira sugestão para atenuar o viés da seleção de imigrantes como realizada nesse trabalho, é o de considerar uma estratégia que seja capaz de distribuir os imigrantes de forma consoante com o formato das paisagens de não-dominância no espaço de objetivos. Dessa forma, sugere-se a computação das *probabilidades de seleção*

$$p_{\mathbf{y}} = P\{\mathbf{y} \text{ ser inserido}\} = \frac{P\{\mathbf{y} \in \mathcal{F}_1\}}{\sum_{i=1}^{K_{\text{NDL}}} P\{\mathbf{y}_i \in \mathcal{F}_1\}}, \quad (7.1)$$

de modo que o imigrante \mathbf{y} gerado seja o escolhido para integrar a população de trabalho com probabilidade $p_{\mathbf{y}}$. Essa estratégia equivale à chamada *seleção proporcional à aptidão* discutida no Capítulo 3, página 39, em que se consideram as probabilidades de não-dominância como uma estimativa do grau de adaptação (adequabilidade) do imigrante na população de trabalho. Note-se que $\sum_{i=1}^{K_{\text{NDL}}} p_{\mathbf{y}_i} = 1$.

Uma segunda possibilidade consiste em modificar a estratégia de geração de imigrantes, ao se incorporarem diretamente as probabilidades estimadas de não-dominância.

Por exemplo, a partir de um conjunto de soluções aleatórias, digamos, Y , o qual não é modificado ao longo de todas as gerações, pode-se estimar a *pdf* de Y no *espaço de decisão* por *Kernels* multivariados da seguinte forma:

$$\hat{p}(\mathbf{y} \in Y) = \frac{1}{N} \sum_{i=1}^N w(\mathbf{y}_i) \left\{ \prod_{j=1}^m \frac{1}{h_j} \phi \left(\frac{y_j - Y_{ij}}{h_j} \right) \right\},$$

$$w(\mathbf{y}) = P\{\mathbf{y}_i \text{ ser inserido}\}, \quad (7.2)$$

em que os pesos $w_i = w(\mathbf{y}_i)$ são computado a partir da equação (7.1). Dessa forma, modifica-se a estimação da *pdf* de Y a cada geração, dando maior importância aos pontos nas regiões que possuem maiores probabilidades de não-dominância. De posse de $\hat{p}(\mathbf{y} \in Y)$, pode-se utilizar uma estratégia de amostragem direta daquela *pdf* para gerar os imigrantes em conformidade com o formato atual das paisagens de não-dominância.

7.3.3 Outras Aplicações para NDLs

A utilidade das NDLs na otimização evolucionária multiobjetivo não se restringe à inserção de imigrantes em MOEAs. Diversas outras aplicações são possíveis e merecem ser investigadas, dentre elas, destacam-se:

1. A utilização de critérios de seleção por probabilidades de não-dominância nas etapas de seleção para reprodução e de sobreviventes no projeto de novos MOEAs;
2. Análise de convergência de MOEAs.

A possibilidade de utilizar as NDLs no item 2 merece comentários adicionais. Poucos estudos teóricos foram publicados na literatura sobre a convergência de MOEAs em MOPs, quer estáticos ou dinâmicos. Em um estudo recente, Beume *et al.* (2011) derivaram pela primeira vez as taxas de convergência de MOEAs conhecidos, em que se prova uma convergência sublinear para o NSGA2 em termos de proximidade à *PF*. A estimação das probabilidades de não-dominância, por outro lado, abre a possibilidade de modelar probabilisticamente a operação de MOEAs, sobretudo aqueles que empregam a estimação da distribuição da população para gerar novos indivíduos no lugar dos operadores de variação genética tradicionais. Por exemplo, em estratégias de seleção por truncamento baseada em *ranks* de Pareto e desempate por *crowding distance*, como aquela empregada no NSGA2, é possível descrever a *pdf* do conjunto de sobreviventes em uma geração

$t + 1$ como

$$p_X^S(\mathbf{x}, t + 1) = \begin{cases} \frac{p_X(\mathbf{x}, t)}{P\{R(\mathbf{x}) < r\}}, & \text{se } R(\mathbf{x}) < r \\ \frac{p_X(\mathbf{x}, t)}{P\{CD(\mathbf{x}) < cd\}}, & \text{se } R(\mathbf{x}) = r \vee CD(\mathbf{x}) \geq cd \\ 0 & \text{caso contrário.} \end{cases}, \quad (7.3)$$

em que r e cd representam níveis críticos de *rank* de Pareto – para $\mathbf{x} \in \mathcal{F}_1$, tem-se $R(\mathbf{x}) = 1$ – e de *crowding distance* descritos para facilitar o entendimento do modelo. Dessa forma, podem-se aplicar as equações derivadas no Capítulo 5 (e.g. $P\{\mathbf{x} \in \mathcal{F}_1\}, \dots, P\{\mathbf{x} \in \mathcal{F}_N\}$) para computar analiticamente $P\{R(\mathbf{x}) < r\}$ a partir de modelos teóricos de *pdf* simples (e.g. Gaussiana). A equação (7.3), portanto, pode-se mostrar útil para análises teóricas do comportamento dos MOEAs existentes.

7.4 Publicações Decorrentes da Pesquisa

Originaram-se as seguintes publicações dos resultados obtidos da pesquisa relatada nessa dissertação:

- Azevedo, C. R. B. e Araújo, A. F. R. (2011). Correlation Between Diversity and Hypervolume in Evolutionary Multiobjective Optimization. Aceito para publicação no *IEEE Congress on Evolutionary Computation (CEC'11)*, New Orleans, USA.
- Azevedo, C. R. B. e Araújo, A. F. R. (2011). Generalized Immigration Schemes for Dynamic Evolutionary Multiobjective Optimization. Aceito para publicação no *IEEE Congress on Evolutionary Computation (CEC'11)*, New Orleans, USA.

Outros artigos referentes aos resultados obtidos da aplicação das Paisagens de Não-Dominância à geração de diversidade dirigida em MOEAs encontram-se em preparação para a submissão a importantes periódicos como o *IEEE Transactions on Evolutionary Computation*, prevista para o primeiro semestre do ano corrente.

Apêndices



Métodos Clássicos de Otimização Multiobjetivo

Este Apêndice apresenta os principais métodos clássicos (i.e., não-evolucionários) de solução para MOPs. Seguir-se-á a classificação apresentada por Marler e Arora (2004), os quais dividem as abordagens clássicas em três grupos, em função do nível de interferência do tomador de decisão no tocante à introdução de preferências (i.e., a importância relativa entre os objetivos) no método de otimização. De acordo com aqueles autores, os métodos de solução de MOPs podem ser agrupados em:

- Métodos de articulação *a priori* de preferências – o usuário indica a importância relativa das funções-objetivo antes da execução do algoritmo de otimização. Restrições são introduzidas ao problema e combinadas em uma função de utilidade, transformando o MOP em um problema de um único objetivo em que uma única solução ótima global é fornecida como resultado do processo.
- Métodos de articulação *a posteriori* de preferências – o usuário não interfere no processo de otimização e seleciona uma solução a partir de um conjunto de soluções matematicamente equivalentes fornecidas pelo algoritmo. Para esse fim, o algoritmo de otimização é codificado para buscar o conjunto ótimo de Pareto.
- Métodos de articulação progressiva de preferências – as etapas de expressão de preferências pelo tomador de decisão e de otimização se revezam em intervalos regulares durante a execução do algoritmo. Preferências parciais são fornecidas ao processo, o qual gera novas soluções de acordo com o interesse expresso pelo usuário.

Cabe ressaltar que, no contexto da otimização multiobjetivo, AEs são, na maioria dos casos, classificados como métodos de articulação *a posteriori* de preferências, embora existam alguns modelos desenvolvidos no intuito de incorporar as preferências do usuário na otimização evolucionária.

Para cada categoria, descreve-se brevemente um ou mais métodos clássicos correspondentes.

A.1 Articulação *a priori* de Preferências

Os métodos de articulação *a priori* de preferências usualmente incorporam parâmetros no modelo de otimização, os quais correspondem a coeficientes, expoentes e restrições de tal forma a refletir as preferências do tomador de decisão em uma única função de utilidade. No contexto das ciências econômicas, uma *função utilidade* representa o grau de satisfação de um indivíduo ou grupo de indivíduos os quais, na otimização multiobjetivo de MOPs, assumem o papel dos tomadores de decisão. Nesse cenário, a função utilidade representa a importância relativa de cada objetivo.

O método mais utilizado para introduzir as preferências do tomador de decisão é conhecido como *Weighted Sum Method*, o qual define a função utilidade U como

$$U = \sum_{i=1}^m w_i f_i(\mathbf{x}). \quad (\text{A.1})$$

Zadeh (1963) demonstrou que, se os pesos w_i forem positivos, minimizar a Equação (A.1) é uma condição suficiente para obter a otimalidade de Pareto. No entanto, Marler e Arora (2004) apontam três dificuldades para a introdução das preferências do tomador de decisão nesse modelo em relação à escolha dos pesos. Primeiro, afirmam que uma escolha *a priori* dos pesos não garante que a solução final será aceitável, levando à necessidade de realizar múltiplas execuções do algoritmo de otimização com novos valores de pesos. Ainda, os pesos devem ser funções dos objetivos – não constantes – para que a soma ponderada na função utilidade reflita adequadamente as preferências do usuário (Messac, 1996, citado por Marler e Arora, 2004). A segunda dificuldade remete à impossibilidade da obtenção de pontos em porções não-convexas do conjunto ótimo de Pareto (Das e Dennis, 1996, citado por Marler e Arora, 2004). No entanto, Marler e Arora (2004) afirmam que, em se utilizando

$$U = \sum_{i=1}^m [w_i f_i(\mathbf{x})]^p, f(\mathbf{x}) > 0 \forall i \quad (\text{A.2})$$

como função utilidade, se o valor de p for suficientemente grande, o algoritmo de otimização será capaz de encontrar pontos ótimos de Pareto em regiões não-convexas com maior facilidade, embora as chances de convergência para um ponto fora do conjunto ótimo de Pareto aumentem à medida que p tende a infinito. Finalmente, a última dificuldade apontada pelos autores remete a não obtenção de um conjunto diverso e espaçado de pontos ótimos de Pareto à medida que se variam os pesos, impedindo, portanto, uma cobertura consistente da frente de Pareto.

Nos métodos do tipo Função-Objetivo com Fronteiras, um único objetivo f_s (o de maior importância) é otimizado separadamente. Todos os demais objetivos são utilizados em restrições que são adicionadas ao problema original na forma

$$l_i \leq f_i(\mathbf{x}) \leq \varepsilon_i, i = 1, \dots, m, i \neq s, \quad (\text{A.3})$$

em que l_i e ε_i são os limites inferiores e superiores para a função objetivo $f_i(\mathbf{x})$, respectivamente. Na abordagem restrição ε (ε -constraint (Haines *et al.*, 1971, citado por Marler e Arora, 2004), supondo minimização dos objetivos, apenas os limites superiores ε_i são considerados, os quais, quando variados sistematicamente, levam à obtenção de um subconjunto de pontos ótimos de Pareto. A principal dificuldade do método restrição- ε reside na escolha dos valores individuais de ε_i . Carmichael (1980), citado por Marler e Arora (2004), sugere que os limites superiores sejam escolhidos tal que

$$f_i(x_i^*) \leq \varepsilon_i \leq f_s(x^*). \quad (\text{A.4})$$

Dessa forma, em vez de requerer que pesos sejam determinados ou uma ordem de preferência particular nos objetivos seja escolhida, o método ε -constraint se baseia na variação consistente dos limites ε_i para produzir um conjunto de pontos no ótimo de Pareto, embora um usuário possa encontrar dificuldades em externar suas preferências em termos de limites dos objetivos, podendo até mesmo comprometer o processo de obtenção de soluções factíveis.

A.2 Articulação *a posteriori* de Preferências

Em situações em que a função de preferência não pode ser aproximada satisfatoriamente por meio de alguma combinação dos objetivos, o algoritmo de otimização pode ser adaptado para fornecer, como resultado do processamento, um conjunto diverso de soluções que reflitam diferentes atribuições de preferências. Nessa situação, deseja-

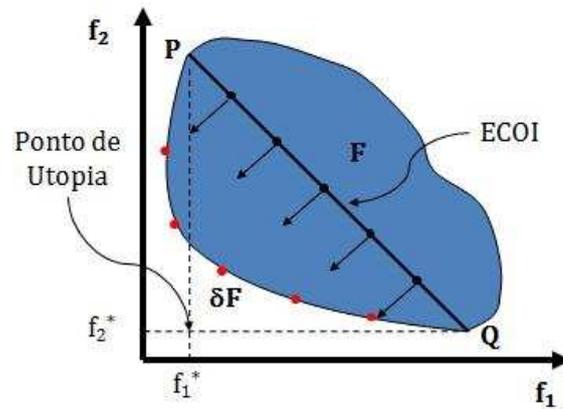


Figura A.1 Operação do método NBI no espaço de objetivos. Os pontos amostrados no envoltório convexo são otimizados na direção do ponto de utopia (ver Definição no capítulo 4). No diagrama, o ECOI é representado pelo seguimento de reta PQ (ver Definição A.1).

se obter uma representação aproximada do conjunto ótimo de Pareto. Os métodos paramétricos com articulação *a priori* de preferências, como o método de soma ponderada, podem ser utilizados para aproximar o conjunto de Pareto mediante a variação sistemática dos parâmetros dos modelos. No entanto, segundo Marler e Arora (2004), tal abordagem pode ser ineficiente, pois, além de não haver garantia de convergência para uma solução ótima por meio de uma atribuição arbitrária de valores paramétricos, o espalhamento das soluções obtidas no conjunto ótimo de Pareto é, em geral, não-uniforme, mesmo que a variação dos pesos seja uniforme. Portanto, algoritmos especialmente projetados para a obtenção simultânea de várias soluções no ótimo de Pareto apresentam-se como uma opção adequada na articulação *a posteriori* de preferências.

No método Normal Boundary Intersection (NBI) (Das e Dennis, 1996), procura-se obter um espalhamento uniforme de pontos sobre o conjunto de Pareto mediante uma seleção uniforme de parâmetros. Ademais, o método NBI opera independentemente da escala relativa entre as diferentes funções-objetivo. A idéia central do algoritmo consiste em resolver sistematicamente vários subproblemas NBI formulados a partir de uma intuição geométrica simples: a de que a interseção entre a normal de um ponto na Envoltória Convexa dos Ótimos Individuais (ECOI) e a fronteira do espaço dos objetivos F , δF , representa, geralmente, pontos ótimos na Frente de Pareto. Assim, ao escolher-se um ponto inicial sobre o ECOI, caminha-se na direção da normal em relação à origem para obter um ponto em δF (Figura A.1). Resolver diversos subproblemas NBI com pontos iniciais espalhados uniformemente no ECOI resulta na obtenção de pontos espalhados uniformemente em δF e esses, por sua vez, serão necessariamente pontos ótimos de

Pareto se F for convexo. Se F não for convexo, então o método NBI pode não obter pontos no conjunto ótimo de Pareto, embora tal situação ocorra apenas em algumas situações especiais. A ECOI é determinada a partir dos pontos \mathbf{x}_i^* , os quais são obtidos através da otimização individual dos objetivos f_i (ver Definição (A.1)).

Definição A.1 (Envoltória Convexa de Ótimos Individuais (ECOI)). Seja $F_i^* = F_i(\mathbf{x}_i^*)$ e Φ uma matriz $n \times n$ cuja i -ésima coluna é o vetor $F_i^* - F^*$. Então o conjunto $\text{ECOI} \subset \mathcal{R}^n$ é composto pelas combinações convexas de F_i^* , i.e., $\{\Phi w : w \in \mathcal{R}^n, \sum_{i=1}^n w_i = 1, w_i \geq 0\}$ (Das e Dennis, 1996).

A principal limitação relatada por Das e Dennis (1996) no método NBI diz respeito à impossibilidade de obtenção de pontos na fronteira de Pareto para MOPs de três ou mais objetivos. Isso ocorre porque, por exemplo, com três objetivos, o ECOI obtido é um plano compreendido pelos vértices do triângulo formado pelos pontos ótimos individuais de cada função objetivo. Em uma situação em que o espaço objetivo F for, por exemplo, uma esfera, o ECOI estará contido na mesma e, claramente, haverá regiões de pontos inacessíveis ao método NBI.

A.3 Articulação Progressiva de Preferências

Outra possibilidade para a articulação de preferências do tomador de decisão consiste na entrada de informações durante o processo de otimização. Essa abordagem é geralmente utilizada quando não se tem conhecimento *a priori* do problema em termos das relações de custo-benefício entre os objetivos. Nesse contexto, pode-se executar um processo de refinamento *on-line* das preferências, à medida em que o algoritmo de otimização encontra novas soluções de compromisso. Após o exame de uma solução pelo usuário, o mesmo pode adaptar suas preferências de acordo com os resultados alcançados. Tal procedimento pode ser compreendido nos seguintes passos (Masud e Ravindran, 2008):

Passo 1 Encontrar uma solução, de preferência factível e eficiente;

Passo 2 Interagir com o tomador de decisão para obter sua reação em relação à solução obtida;

Passo 3 Repetir os Passos 1 e 2 até que uma solução satisfatória seja obtida.

Benayoun *et al.* (1971) propuseram o *Step Method* (STEM), um dos primeiros algoritmos nessa categoria. O STEM introduz parâmetros de relaxamento dos objetivos no

modelo de otimização, os quais são ajustados pelo tomador de decisão interativamente. O processo é repetido até que uma solução de compromisso seja encontrada ou não seja possível prosseguir com o processo de relaxamento. A metodologia geral do STEM é descrita em quatro passos:

Passo 0 *Obter os valores ideais das funções-objetivo sobre a região factível inicial Ω :*

$$f_i^* = \min_{\mathbf{x} \in \Omega} \{f_i(\mathbf{x})\}, \quad (\text{A.5})$$

definir o conjunto de índices de relaxamento (\mathbf{J}), que contém os índices dos objetivos a serem relaxados, como o conjunto vazio: $\mathbf{J} \leftarrow \emptyset$ e fazer $n \leftarrow 0$, em que n é a iteração atual;

Passo 1 *Construir a região de soluções factíveis da $(n + 1)$ -ésima iteração:*

$$\Omega^{(n+1)} = \begin{cases} \Omega^{(n)} \\ f_i(\mathbf{x}) \geq f_i(\mathbf{x}^{(n)}) - \Delta_i, & i \in \mathbf{J}, \\ f_i(\mathbf{x}) \geq f_i(\mathbf{x}^{(n)}), & i \notin \mathbf{J}. \end{cases} \quad (\text{A.6})$$

em que $\mathbf{x}^{(n)}$ é a solução obtida na n -ésima iteração e $f_i(\mathbf{x}^{(n)})$ é o valor da i -ésima função objetivo no ponto $\mathbf{x}^{(n)}$. Ainda, Δ_i representa a quantidade pela qual $f_i(\mathbf{x}^{(n)})$ deve ser relaxada, para $i \in \mathbf{J}$. Ambos Δ_i e \mathbf{J} foram especificados no Passo 4 na n -ésima iteração;

Passo 2 *Resolver o novo problema MOP, reformulado como*

$$\min_{\mathbf{x}, \lambda} \lambda \quad \text{sujeito a} \begin{cases} \lambda \geq \pi_i \{f_i^* - f_i(\mathbf{x})\}, & i \in \mathcal{I}, \\ \lambda \geq 0, \\ \mathbf{x} \in \Omega^{(n+1)}. \end{cases} \quad (\text{A.7})$$

em que π_i é o peso relativo da distância entre f_i^* e $f_i(\mathbf{x})$ e λ é o valor máximo entre os pesos das distâncias. Aos pesos associados com objetivo relaxados, são atribuídos o valor zero de tal forma que o objetivo relaxado deixa de fazer parte do critério de otimização. Uma atribuição binária $\{0, 1\}$ é a forma mais simples de se manter a premissa do método STEM.

Passo 3 *Avaliar a solução atual obtida.* O tomador de decisão compara cada valor $f_i(\mathbf{x}^{(n+1)})$ com aqueles f_i^* calculados no Passo 0 e julga se a solução é satisfatória

ou não. Caso seja, o processo é interrompido e $\mathbf{x}^{(n+1)}$ é a solução final. Do contrário, prossegue-se para o Passo 4.

Passo 4 *Especificar as quantidade de relaxamento.* Os valores das funções-objetivo que são determinados como insatisfatórios devem ser melhorados. A metodologia STEM supõe que isso pode ser feito às expensas de alguns dos valores satisfatórios das funções-objetivo. Assim, o tomador de decisão seleciona um dos objetivos (digamos, f_s) para o relaxamento, fazendo $\mathbf{J} \leftarrow \{s\}$, e especifica um valor para Δ_s . O algoritmo então retorna ao Passo 1. Caso o usuário não esteja disposto a sacrificar quaisquer objetivos, então o algoritmo é interrompido.

Jeong e Kim (2005) apontam as seguintes vantagens dos métodos de articulação progressiva de preferências (também conhecidos como métodos iterativos): (i) a não necessidade de definir preferências antes do processo; (ii) a aprendizagem das relações de custo-benefício durante o processo; e (iii) a solução final obtida possui maiores perspectivas de ser implementada.

Dentre as desvantagens, destacam-se (Jeong e Kim, 2005): (i) a não garantia de obtenção da solução preferível em um número limitado de iterações; (ii) o alto esforço cognitivo requerido para decisões imediatas; e (iii) o esforço computacional extra para resolver novamente o problema a cada iteração.

B

Problemas de Otimização Multiobjetivo

Nesse Apêndice, apresentam-se alguns problemas encontrados na literatura recente de otimização multiobjetivo por computação evolucionária. Alguns desses problemas são utilizados para a validação e comparação entre novas propostas de algoritmos, como o do Caixeiro viajante multiobjetivo. Outros consideram problemas práticos de interesse crescente no âmbito da indústria, como o caso do projeto de redes *multi-cast*.

B.1 Roteamento *Multi-Cast*

Garrozi e Araujo (2006) e Araújo e Garrozi (2010) propuseram uma solução multiobjetivo para o problema de roteamento em redes *multi-cast*, o qual consiste em transmitir dados redundantes, a partir de um ou mais nós fontes, para múltiplos receptores em uma rede, de forma que o custo e o atraso dos caminhos de roteamento que levam aos destinos sejam mínimos. Na abordagem *Multicast Routing Genetic Algorithm* (MulRoGA), dois objetivos conflitantes são definidos: (i) a maximização dos *links* comuns e (ii) a minimização dos caminhos de roteamento. A formulação do problema, proposta inicialmente em Garrozi e Araujo (2006), considera um terceiro objetivo relativo a um termo de recompensa para as soluções factíveis, i.e., aquelas que não violam restrições no problema. A combinação dos três objetivos em uma única função de aptidão é então realizada da seguinte forma:

$$G(\mathbf{x}) = 1 - \frac{\alpha g_1(\mathbf{x})}{\beta g_2(\mathbf{x})} + \gamma g_3(\mathbf{x}), \quad (\text{B.1})$$

em que g_1 deve ser maximizado em termos do número de *links* comuns e g_2 minimizado em relação ao comprimento total da rota. Os autores reportam resultados obtidos com o MulRoGA que são competitivos em comparação com algoritmos tradicionais de

roteamento via caminho mínimo *unicast* e heurísticas eficientes como o resfriamento simulado (Araújo e Garrozi, 2010).

B.2 Agrupamento de Dados

Em problemas de agrupamento, uma limitação no processo de avaliação da qualidade dos grupos obtidos reside na subjetividade com que o processo de avaliação é conduzido (Jain *et al.*, 1999). Por exemplo, a avaliação da solução obtida considerando apenas a minimização do critério da soma de distâncias intra-agrupamento (intracluster) (Equação (B.2)) pode levar ao resultado trivial da atribuição de cada objeto a um único grupo. Da mesma forma, soluções que utilizam um número elevado de grupos deveriam ser penalizadas.

Nesse sentido, Hruschka *et al.* (2009) afirmam que, em problemas de agrupamento, a qualidade da solução deveria ser avaliada por um conjunto diverso de critérios, de modo a reduzir o viés induzido pela otimização de um único critério na análise de validação. Dessa forma, a abordagem multiojetivo pode ser empregada para encontrar agrupamentos satisfatórios sobre múltiplos pontos de vista. Um exemplo de MOEA o qual considera a otimização de ambos, a variabilidade intracluster total e o número de partições, é aquele proposto por Korkmaz *et al.* (2006), em que uma aproximação da PF é obtida para a análise de *tradeoff*.

Um critério de validação comumente usado é o da soma dos erros quadráticos, o qual deve ser minimizado:

$$f(\mathbf{C}_1, \dots, \mathbf{C}_k) = \sum_{j=1}^k \sum_{\mathbf{x} \in \mathbf{C}_j} \|\mathbf{x}_i - \mathbf{z}_j\|, \quad (\text{B.2})$$

em que $\{\mathbf{C}_1, \dots, \mathbf{C}_k\}$ é um conjunto de k partições, \mathbf{x}_i é o i -ésimo objeto de entrada e \mathbf{z}_j é o centróide da partição \mathbf{C}_j . De acordo com Hruschka *et al.* (2009), a distorção média de $f(\mathbf{C}_1, \dots, \mathbf{C}_k)$ sobre todo o conjunto de entrada pode ser considerada uma medida da variabilidade intracluster. Dessa forma, Handl e Knowles (2007) definem um algoritmo evolucionário multiojetivo para minimizar o critério de compactação, dado pela Equação (B.2), e maximizar a conectividade, definida como o grau com que objetos vizinhos são dispostos na mesma partição.

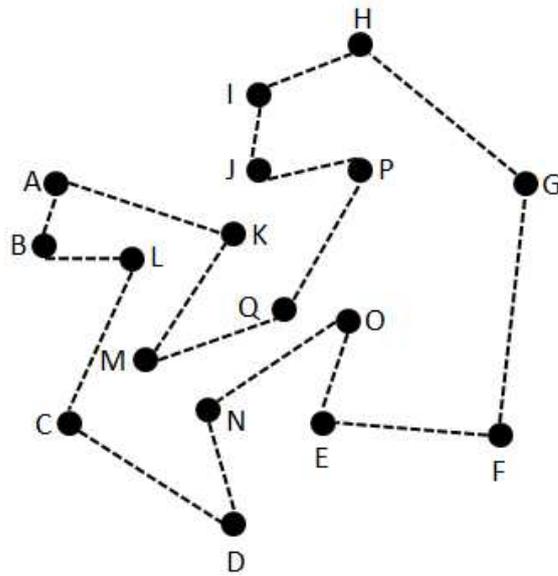


Figura B.1 Representação de uma instância aleatória do TSP.

B.3 Caixeiro Viajante Multiobjetivo

A otimização multiobjetivo também pode ser empregada para resolver com eficiência problemas inicialmente formulados a partir de uma única função objetivo. Trabalhos recentes, e.g. (Jähne *et al.*, 2009; Jensen, 2004; Knowles *et al.*, 2001), demonstram duas abordagens válidas para esse propósito. A primeira diz respeito à adição de novos objetivos ao problema, enquanto a segunda consiste na decomposição do objetivo primário em subobjetivos. Jähne *et al.* (2009) utilizou o Problema do Caixeiro Viajante (TSP, *Traveller Salesman Problem*, do inglês) como estudo de caso. Os resultados iniciais sugerem que as estratégias empregadas aumentam a eficiência da busca evolucionária por soluções no TSP.

O TSP consiste em determinar um ciclo Hamiltoniano¹ de custo mínimo no grafo não-direcionado ponderado cujos vértices representam localidades a serem visitadas e as arestas denotam as rotas disponíveis (Figura B.1). Os pesos associados a cada aresta representam as distâncias a serem percorridas entre duas localidades (cidades) adjacentes. Seja $C = \{c_1, \dots, c_N\}$ um conjunto de N localidades com uma matriz associada M , $N \times N$, construída a partir das distâncias entre cada par de localidades. Dada uma solução candidata descrita pela permutação $\pi = (\pi_1, \dots, \pi_N)$, a qual representa a ordem em que

¹Um ciclo completo composto por todos os vértices em um grafo.

cada cidade é visitada, então a função de custo pode ser descrita como:

$$D(\pi) = \sum_{i=1}^N M(c_{\pi[i]}, c_{\pi[i \oplus 1]}),$$

$$\text{em que } i \oplus 1 = \begin{cases} i+1, & \text{se } i < N, \\ 1, & \text{se } i = N. \end{cases} \quad (\text{B.3})$$

A transformação do TSP descrito na Equação (B.3) em um problema multiobjetivo pode ser feita de várias formas. Por exemplo, Jähne *et al.* (2009) definem o método *Multi-Objectivization via Segmentation* (MOS) segundo o princípio de decomposição dos intracomponentes de uma instância do TSP em dois conjuntos com diferentes características. A função objetivo, após esse passo, é então segmentada em dois novos objetivos atuando separadamente em cada conjunto. Uma das técnicas utilizadas para a decomposição dos intracomponentes considera o valor esperado das distâncias entre cidades. Dessa forma, dois conjuntos de cidades são definidos, a saber,

$$C_{\uparrow} = \{c_i \in C : \mu_i^s > \widehat{\mu}^s\}$$

$$C_{\downarrow} = \{c_i \in C : \mu_i^s \leq \widehat{\mu}^s\} \quad (\text{B.4})$$

$$\text{com } \mu_i^s = \frac{\sum_{s=1}^S d_i^s}{S} \text{ e } \widehat{\mu}^s = \frac{\sum_{i=1}^N \mu_i^s}{N}$$

em que S é o número de soluções candidatas envolvidas na decomposição dos conjuntos e d_i^s é a soma das distâncias entre a cidade c_i e as duas localidades vizinhas na solução candidata s . Ademais, μ_i^s é a distância média em relação às duas cidades vizinhas correspondentes sobre todas as soluções candidatas contidas em S e $\widehat{\mu}^s$ é a média das distâncias médias tomadas em relação a todas as cidades. Os autores argumentam que tal decomposição faz com que a busca possa se concentrar em partes cruciais do problema. Além disso, a variação das soluções sobre as cidades que possuem distâncias médias maiores contribui com maior eficiência para a introdução de diversidade e para escapar de mínimos locais. Os resultados obtidos com a abordagem MOS superaram qualitativamente aqueles obtidos com a aplicação de algoritmos evolucionários de um único objetivo para o TSP.

B.4 Problemas ZDT para MOPs Contínuos

Uma das principais ferramentas disponíveis para a avaliação da eficiência de novos algoritmos de otimização é a utilização de problemas *benchmarks*. Esses problemas são especialmente projetados para observar o comportamento dos algoritmos em diferentes

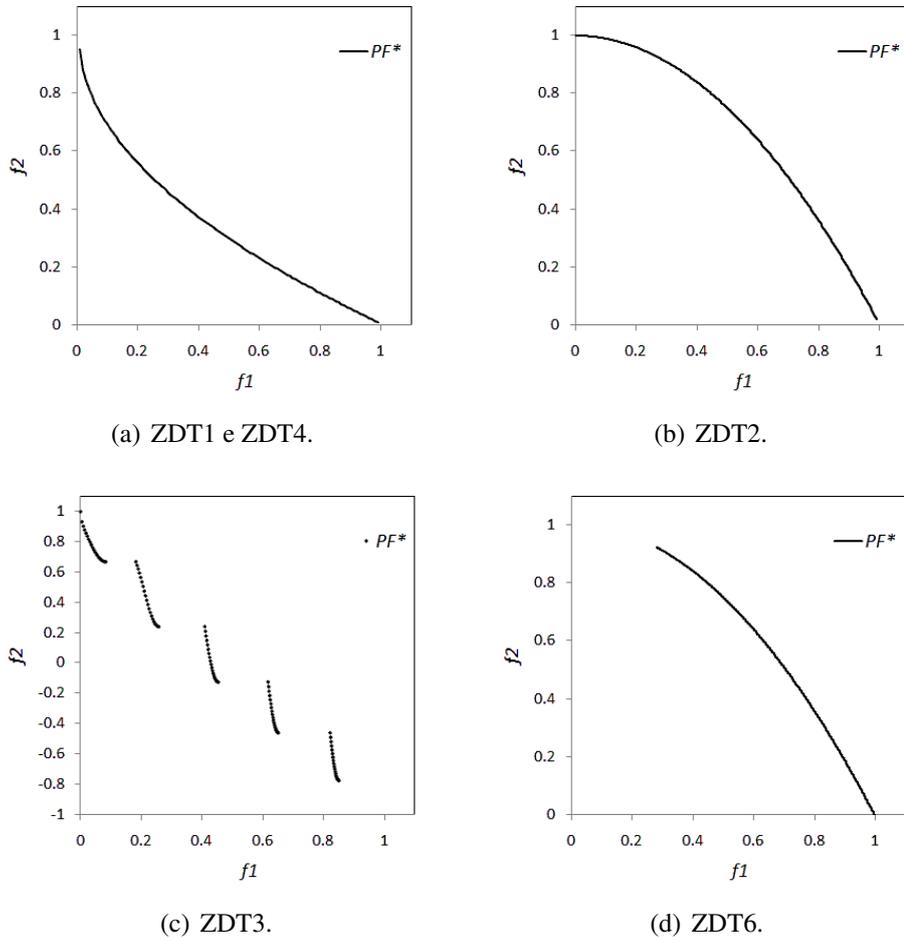


Figura B.2 Frentes de Pareto no conjunto de problemas de otimização ZDT.

contextos de complexidade e níveis de dificuldade.

Na otimização multiobjetivo, o conjunto de problemas *benchmark* ZDT (Zitzler *et al.*, 2002) é um dos mais utilizados por representarem as principais características encontradas em problemas do mundo real. Os problemas ZDT possuem dois objetivos e seguem a forma geral

$$\begin{aligned} \min_{\mathbf{x} \in \Omega} \mathbf{f}(\mathbf{x}) &= (f_1(x_1), f_2(\mathbf{x})) \\ \text{sujeito a } f_2(\mathbf{x}) &= g(x_2, \dots, x_n)h(f_1(x_1), g(x_2, \dots, x_n)) \end{aligned} \quad (\text{B.5})$$

em que f_1 é uma função da primeira variável, g é uma função das $n - 1$ variáveis restantes e h é uma função composta por f_1 e g . Os seis problemas ZDT (Figura B.2) são descritos em termos das funções f_1 , g e h :

- O problema ZDT1 possui uma frente ótima de Pareto convexa:

$$\begin{aligned}
 f_1(x_1) &= x_1 \\
 g(x_2, \dots, x_n) &= 1 + 9 \sum_{i=2}^n x_i / (n-1) \\
 h(\mathbf{x}) &= 1 - \sqrt{f_1(x_1)/g(\mathbf{x})}
 \end{aligned} \tag{B.6}$$

em que $\mathbf{x} \in [0, 1]^n$. A frente ótima de Pareto $PF_{ZDT1} = \mathbf{f}(\Omega^*)$ é obtida fazendo $g(\mathbf{x}) = 1$, i.e., $\Omega^* = \{\mathbf{x} \in \Omega : f_2(\mathbf{x}) = 1 - \sqrt{x_1}\}$.

- O problema ZDT2 possui uma frente ótima de Pareto côncava:

$$\begin{aligned}
 f_1(x_1) &= x_1 \\
 g(x_2, \dots, x_n) &= 1 + 9 \sum_{i=2}^n x_i / (n-1) \\
 h(\mathbf{x}) &= 1 - (f_1(x_1)/g(\mathbf{x}))^2
 \end{aligned} \tag{B.7}$$

em que $\mathbf{x} \in [0, 1]^n$. A frente de Pareto $PF_{ZDT2} = \mathbf{f}(\Omega^*)$ é obtida fazendo $g(\mathbf{x}) = 1$, i.e., $\Omega^* = \{\mathbf{x} \in \Omega : f_2(\mathbf{x}) = 1 - x_1^2\}$.

- O problema ZDT3 apresenta descontinuidades e possui uma frente ótima de Pareto convexa não-contígua dividida em cinco partes:

$$\begin{aligned}
 f_1(x_1) &= x_1 \\
 g(x_2, \dots, x_n) &= 1 + 9 \sum_{i=2}^n x_i / (n-1) \\
 h(\mathbf{x}) &= 1 - \sqrt{f_1(x_1)/g(\mathbf{x})} - (f_1(x_1)/g(\mathbf{x})) \text{sen}(10\pi f_1(x_1))
 \end{aligned} \tag{B.8}$$

em que $\mathbf{x} \in [0, 1]^n$. A frente ótima de Pareto $PF_{ZDT3} = \mathbf{f}(\Omega^*)$ é obtida fazendo $g(\mathbf{x}) = 1$, i.e., $\Omega^* = \{\mathbf{x} \in \Omega : f_2(\mathbf{x}) = 1 - \sqrt{x_1} - x_1 \text{sen}(10\pi x_1)\}$. Note-se, no entanto, que não há descontinuidades no espaço de decisão.

- O problema ZDT4 apresenta uma paisagem de aptidão multimodal, possuindo 21^9

frentes ótimas de Pareto locais para $n = 10$:

$$\begin{aligned}
 f_1(x_1) &= x_1 \\
 g(x_2, \dots, x_n) &= 1 + 10(n-1) + \sum_{i=2}^n (x_i^2 - \cos(4\pi x_i)) \\
 h(\mathbf{x}) &= 1 - \sqrt{f_1(x_1)/g(\mathbf{x})}
 \end{aligned} \tag{B.9}$$

em que $x_1 \in [0, 1]$ e $x_2, \dots, x_n \in [-5, 5]$. A frente ótima de Pareto global é obtida fazendo $g(\mathbf{x}) = 1$, o que resulta em $PF_{ZDT4} = PF_{ZDT1}$. A melhor frente ótima de Pareto local é obtida fazendo $g(\mathbf{x}) = 1.25$. Note-se que nem todas as PF_{ZDT4} ótimas locais são distinguíveis no espaço de objetivos.

- O problema ZDT6 apresenta uma distribuição não-uniforme de pontos no PF_{ZDT6} . Ademais, a densidade de soluções factíveis é a menor possível nas regiões próximas à PF_{ZDT6} e maior em regiões mais afastadas:

$$\begin{aligned}
 f_1(x_1) &= 1 - \exp(-4x_1)\text{sen}^6(6\pi x_1) \\
 g(x_2, \dots, x_n) &= 1 + 9\left(\left(\sum_{i=2}^n x_i\right)/(n-1)\right)^{\frac{1}{4}} \\
 h(\mathbf{x}) &= 1 - (f_1(x_1)/g(\mathbf{x}))^2
 \end{aligned} \tag{B.10}$$

em que $\mathbf{x} \in [0, 1]^n$. A frente ótima de Pareto $PF_{ZDT6} = \mathbf{f}(\Omega^*) \subset PF_{ZDT2}$ é obtida fazendo $g(\mathbf{x}) = 1$, i.e., $\Omega^* = \{\mathbf{x} \in \Omega : (f_2(\mathbf{x}) = 1 - x_1^2) \wedge (x_1 \in [0.2807753191, 1])\}$.

O problema ZDT5 é o único de natureza discreta no conjunto e não é aqui descrito, pois este trabalho visa otimização em espaços contínuos.

B.5 Problemas FDA para MOPs Dinâmicos Contínuos

O conjunto de cinco problemas FDA (Farina *et al.*, 2004) é uma das primeiras escolhas para testes experimentais com MOEAs aplicados a MOPs dinâmicos. Os problemas são projetados para refletir os tipos de mudança temporais discutidos no Capítulo 3, página 43. No FDA, definem-se os três primeiros problemas como uma extensão dos problemas ZDT, em que funções auxiliares são adicionadas aos dois objetivos com o propósito de guiar a dinâmica temporal dos problemas. Os dois últimos problemas FDA são formulados inicialmente com três objetivos, baseados nos MOPs DTLZ (Deb *et al.*, 2002), não descritos nesse Apêndice. A Figura B.3 mostra as $PF(t)$ s variantes no

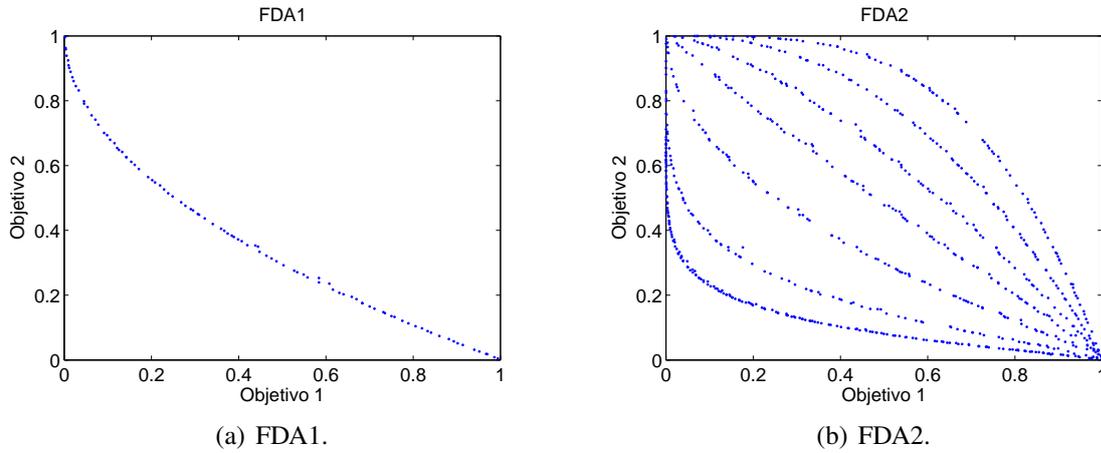


Figura B.3 Frentes de Pareto variantes no tempo ($PF(t)$) no conjunto FDA.

tempo para o FDA1 e FDA2, respectivamente. Note-se que as Figuras foram geradas a partir de execuções do algoritmo NSGA2 o qual, para diferentes instantes fixos de tempo, aproximou a $PF(t)$ com um conjunto de 100 pontos. Por ser um problema de Tipo 1, apenas o conjunto ótimo de Pareto ($\Omega^*(t)$) varia no tempo no FDA1, sendo a PF constante. No problema FDA2, por outro lado, a $PF(t)$ muda em passos discretos no tempo de um formato convexo para um formato côncavo. As formulações originais do FDA1, FDA2, FDA3, FDA4 e FDA5 (os MOPs dinâmicos FDA1 e FDA2 são utilizados nos experimentos relatados no Capítulo 6), são dadas a seguir.

- FDA1 – problema Tipo 1, definido como

$$\begin{aligned}
 f_1(X_I, t) &= x_1, \\
 g(X_{II}, t) &= 1 + \sum_{x_i \in X_{II}} (x_i - G(t))^2, \\
 h(f_1, g) &= 1 - \sqrt{\frac{f_1}{g}}, \\
 G(t) &= \text{sen}(0,5 \cdot \pi \cdot t), \\
 f_2(X_I, X_{II}, X_{III}, t) &= g \cdot h, \\
 X_I = (x_1) &\in [0, 1], X_{II}, X_{III} \in [-1, 1].
 \end{aligned}
 \tag{B.11}$$

- FDA2 – problema Tipo 3, definido como

$$\begin{aligned}
 f_1(X_I, t) &= x_1, \\
 g(X_{II}, t) &= 1 + \sum_{x_i \in X_{II}} x_i^2, \\
 h(X_{III}, f_1, g) &= 1 - \left(\frac{f_1}{g} \right)^{\left(H(t) + \sum_{x_i \in X_{III}} (x_i - H(t))^2 \right)^{-1}}, \\
 G(t) &= \text{sen}(0,5 \cdot \pi \cdot t), \\
 H(t) &= 0,75 + 0,7 \cdot \text{sen}(0,5 \cdot \pi \cdot t), \\
 f_2(X_I, X_{II}, X_{III}, t) &= g \cdot h, \\
 X_I &\in [0, 1], X_{II}, X_{III} \in [-1, 1].
 \end{aligned} \tag{B.12}$$

- FDA3 – problema Tipo 2, definido como

$$\begin{aligned}
 f_1(X_I, t) &= \sum_{x_i \in X_I} x_i^{F(t)}, \\
 g(X_{II}, t) &= 1 + G(t) + \sum_{x_i \in X_{II}} (x_i - G(t))^2, \\
 h(X_{III}, f_1, g) &= 1 - \sqrt{\frac{f_1}{g}}, \\
 G(t) &= |\text{sen}(0,5 \cdot \pi \cdot t)|, \\
 F(t) &= 10^{2 \cdot \text{sen}(0,5 \cdot \pi \cdot t)}, \\
 X_I &\in [0, 1], X_{II} \in [-1, 1].
 \end{aligned} \tag{B.13}$$

- FDA4 – problema Tipo 1, definido como

$$\begin{aligned}
 \min_{\mathbf{x}} f_1(\mathbf{x}, t) &= (1 + g(X_{II}, t)) \prod_{i=1}^{M-1} \cos\left(\frac{x_i \cdot \pi}{2}\right), \\
 &\vdots \\
 \min_{\mathbf{x}} f_k(\mathbf{x}, t) &= (1 + g(X_{II}, t)) \left(\prod_{i=1}^{M-k} \cos\left(\frac{x_i \cdot \pi}{2}\right) \right) \text{sen}\left(\frac{x_{M-k} + \pi}{2}\right), \\
 &\vdots \\
 \min_{\mathbf{x}} f_M(\mathbf{x}, t) &= (1 + g(X_{II}, t)) \text{sen}\left(\frac{x_1 \cdot \pi}{2}\right), \\
 g(X_{II}, t) &= \sum_{x_i \in X_{II}} (x_i - G(t))^2, \\
 G(t) &= |\text{sen}(0,5 \cdot \pi \cdot t)|, \\
 X_{II} &= (x_M, \dots, x_n), x_i \in [0, 1], i = 1 : n.
 \end{aligned} \tag{B.14}$$

- FDA5 – problema Tipo 2, definido como

$$\begin{aligned}
 \min_{\mathbf{x}} f_1(\mathbf{x}, t) &= (1 + g(X_{II}, t)) \prod_{i=1}^{M-1} \cos\left(\frac{y_i \cdot \pi}{2}\right), \\
 &\vdots \\
 \min_{\mathbf{x}} f_k(\mathbf{x}, t) &= (1 + g(X_{II}, t)) \left(\prod_{i=1}^{M-k} \cos\left(\frac{y_i \cdot \pi}{2}\right) \right) \text{sen}\left(\frac{y_{M-k} + \pi}{2}\right), \\
 &\vdots \\
 \min_{\mathbf{x}} f_M(\mathbf{x}, t) &= (1 + g(X_{II}, t)) \text{sen}\left(\frac{y_1 \cdot \pi}{2}\right), \\
 g(X_{II}, t) &= G(t) + \sum_{x_i \in X_{II}} (x_i - G(t))^2, \\
 y_i &= x_i^{F(t)}, i = 1, \dots, (M-1), \\
 G(t) &= |\text{sen}(0,5 \cdot \pi \cdot t)|, \\
 F(t) &= 1 + 100\text{sen}^4(0,5 \cdot \pi \cdot t), \\
 X_{II} &= (x_M, \dots, x_n), x_i \in [0, 1], i = 1 : n.
 \end{aligned} \tag{B.15}$$



Outros Algoritmos Evolucionários Multiobjetivo

Apresentam-se nesse Apêndice alguns dos principais MOEAs propostos na literatura e ressaltam-se as diferentes estratégias empregadas para a obtenção do conjunto de soluções não-dominadas \mathcal{F}_1 , bem como se enfatizam as principais diferenças entre três aspectos essenciais para a eficácia do algoritmo em aproximar PF , mesmo em MOPs com alto grau de dificuldade quanto à existência de regiões não-convexas, descontinuidades e multimodalidade no espaço de objetivos: (i) estratégia de atribuição de valores de aptidão; (ii) elitismo; e (iii) mecanismos de geração e manutenção de diversidade. Em particular, o item (iii) é o de maior interesse nesse trabalho.

C.1 Algoritmo Genético com Avaliação de Vetores

Schaffer (1985) foi o primeiro a propor um MOEA o qual aproximasse o conjunto ótimo de Pareto, culminando na criação do *Vector Evaluated Genetic Algorithm* (VEGA), o qual utiliza diferentes atribuições de aptidão baseadas em cada objetivo do MOP. O VEGA divide a população P em m subpopulações de tamanhos iguais, P_1, \dots, P_m , nas quais a cada indivíduo $\mathbf{x}_j \in P_i$ é atribuído uma aptidão baseada na função objetivo f_i . A seleção de indivíduos para a etapa de variação em cada subpopulação é realizada proporcionalmente aos valores de aptidão, de modo que, em termos práticos, cada ciclo de operação realizada em subpopulações particulares equivale à execução de uma geração completa de um AG de um único objetivo.

No tocante às estratégias de elitismo e geração/manutenção de diversidade, o VEGA não apresenta quaisquer novidades e nem sequer faz uso de tais mecanismos. O procedimento para a aproximação do conjunto de Pareto em MOPs através do VEGA (supondo

Pseudocódigo 8 Vector Evaluated Genetic Algorithm (VEGA)

Seja $N_S = |P|/m$ o tamanho das subpopulações.
 InicializarPopulação($P^{(0)}$), $t \leftarrow 0$
enquanto condição de término não satisfeita **faça**
 para todo objetivo $f_i \in F_m$ **faça**
 $P_i^{(t)} = \emptyset$
 para $k = 1$ to N_S **faça**
 $\mathbf{x}_k \leftarrow \text{SelecaoProporcionalAptidão}(P, f_i)$
 Aptidão(\mathbf{x}_k) $\leftarrow f_i(\mathbf{x}_k)$
 $P_i^{(t)} \leftarrow P_i^{(t)} \cup \{\mathbf{x}_k\}$
 fim para
 fim para
 $P^{(t)} = \bigcup_i^m P_i^{(t)}$
 $P^{(t+1)} \leftarrow \text{OperadoresGeneticos}(P^t)$
 $t \leftarrow t + 1$
fim enquanto
retorne $P^{(t)}$

maximização dos objetivos) é dado no Pseudocódigo 8. Como resultado de sua execução, reportou-se (Schaffer, 1985) que VEGA possui a capacidade de gerar um conjunto de soluções localmente não-dominadas, em que “localmente” refere-se ao fato de que, no trabalho cujo VEGA foi proposto, computaram-se as relações de dominância considerando apenas a população atual. Apesar de computar tais informações, o VEGA não utiliza as mesmas para guiar a busca em direção à PF .

Em Schaffer (1985), citado por Coello Coello (2000), observa-se a principal limitação na abordagem utilizada pelo VEGA: um efeito indesejado da especiação, o qual tende a fazer com que a população convirja em direção às regiões extremas do espaço objetivo F_m , devido à consideração dos objetivos separadamente no operador de seleção. Schaffer (1985) denotou esse efeito como “*middling*”. Ainda, Richardson *et al.* (1989) observaram que o mecanismo de ordenação aleatória da população para a aplicação posterior dos operadores genéticos — cruzamento e mutação — equivale a uma seleção proporcional em relação a uma combinação linear dos objetivos com pesos dependentes da distribuição dos indivíduos no espaço de soluções. Nos experimentos realizados naquele trabalho, o VEGA apresentou as mesmas limitações dos métodos clássicos baseados em soma ponderada dos objetivos, dentre as quais, uma pobre cobertura da PF quando F_m apresenta regiões côncavas na superfície de *tradeoff*.

Apesar de apresentar importantes limitações, a simplicidade e facilidade de implementação do VEGA ainda tem motivado a aplicação desse algoritmo a uma variedade de

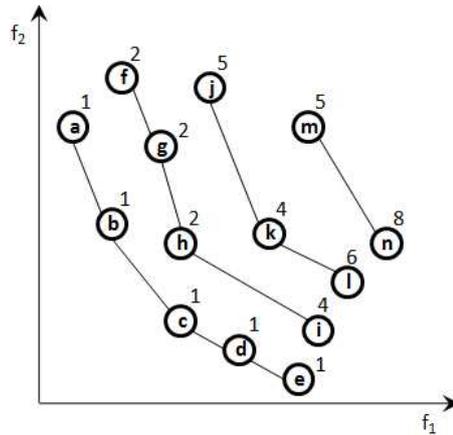


Figura C.1 Exemplo da atribuição de *ranks* empregada no MOGA (adaptado de Konak *et al.* (2006)).

problemas, e.g. ao projeto de geradores de números pseudoaleatórios utilizando autômatos celulares (Guan e Zhang, 2003), com a obtenção de resultados promissores. O VEGA também é comumente implementado como referência em *benchmarks* de avaliação de novos MOEAs. Embora existam modificações que objetivam melhorar o desempenho do VEGA em MOPs (c.f. Cvetkovic *et al.* (1998)), concentramo-nos nos modelos originais ao abordar esse e os demais MOEAs discutidos nesse trabalho.

C.2 Algoritmo Genético multiobjetivo

Denominou-se a primeira abordagem apresentada na literatura baseada em *rank* de Pareto — na qual as informações de dominância são consideradas para guiar a busca evolucionária — como *Multi-objective Genetic Algorithm* (MOGA) (Fonseca e Fleming, 1995). Seja Q um dado conjunto de pontos no espaço de busca. O *rank* de um ponto \mathbf{x} no conjunto Q é dado pelo número de soluções que dominam \mathbf{x} , i.e., $\text{rank}(\mathbf{x}, Q) := |\{y \in Q | x \prec y\}|$. O *rank* no MOGA consiste em um mapeamento dos indivíduos (\mathbf{x}) em um valor relativo ao número de soluções que dominam \mathbf{x} na população $P^{(t)}$ — relativa à t -ésima geração —, $\text{rank}(\mathbf{x}, P^{(t)})$, descrito como

$$r(\mathbf{x}, t) = 1 + \text{rank}(\mathbf{x}, P^{(t)}). \quad (\text{C.1})$$

O esquema de *rank* penaliza as soluções localizadas nas regiões do espaço de objetivos F_m cuja densidade da região ocupada por soluções na \mathcal{F}_1 aproximada seja alta. Um exemplo de como o *rank* é atribuído a partir das relações de dominância pode ser

visualizado na Figura C.2. Por exemplo, a solução **i** é dominada por **c**, **d** e **e**, e, portanto, de acordo com a Equação (C.1), lhe será atribuído o valor 4 como *rank*.

A preservação da diversidade baseada em técnicas explícitas de especiação (*niching*) também fora utilizada. A técnica utilizada no MOGA é conhecida como *fitness sharing* (compartilhamento de aptidão) (Goldberg e Richardson, 1987) e tem como premissa a distribuição da população em diferentes picos do espaço de busca, com a população sendo distribuída conforme a altura daquele pico. Para tanto, calcula-se o alcance do nicho (*niche counter*), expresso por

$$nc(\mathbf{x}, t) = \sum_{\mathbf{y} \in P, r(\mathbf{y}, t) = r(\mathbf{x}, t)} \max \left\{ \frac{\sigma_{share} - df(\mathbf{x}, \mathbf{y})}{\sigma_{share}}, 0 \right\}, \quad (C.2)$$

em que σ_{share} controla o alcance do nicho no espaço objetivo e df é dado como

$$df(\mathbf{x}, t) = \sqrt{\sum_{i=1}^m \left(\frac{f_i(\mathbf{x}) - f_i(\mathbf{y})}{\lceil f_i \rceil - \lfloor f_i \rfloor} \right)^2}, \quad (C.3)$$

representando a distância entre duas soluções \mathbf{x}, \mathbf{y} normalizadas em F_m no intervalo $[0, 1]$. Os valores de aptidão de cada indivíduo são posteriormente divididos pelo alcance do nicho, de modo que o valor $nc(\mathbf{x}, t)$ indica quantos indivíduos compartilham a região ocupada pela solução \mathbf{x} . Esse processo é representado como

$$\mathbf{f}'(\mathbf{x}) = \frac{1}{nc(\mathbf{x}, t)} \mathbf{f}(\mathbf{x}). \quad (C.4)$$

Dessa forma, o MOGA guia a busca evolucionária em direção ao *PF*, enquanto tenta manter a diversidade (i.e., espalhamento) das soluções não-dominadas na população atual. Para um melhor entendimento, apresenta-se o funcionamento geral do MOGA no Pseudocódigo 9.

As principais desvantagens do MOGA e de outros métodos baseados em *rank* de Pareto são o alto esforço computacional requerido para identificar soluções não-dominadas na população, o baixo grau de escalabilidade (em termos de eficiência computacional e desempenho) quando novos objetivos são adicionados ao problema e quando o tamanho da população cresce. Além dessas limitações, não há forma sistemática de determinar o melhor valor para o parâmetro σ_{share} , o qual influencia diretamente o desempenho do algoritmo e é dependente do problema (Coello Coello, 2000).

Pseudocódigo 9 Multi-objective Genetic Algorithm (MOGA)

InicializarPopulação($P^{(0)}$), $t \leftarrow 0$
enquanto condição de término não satisfeita **faça**
 para todo $\mathbf{x}_j \in P^{(t)}$ **faça**
 $r_j \leftarrow r(\mathbf{x}_j, t)$
 $z_j = \text{Aptidão}(\mathbf{x}_j, t) \leftarrow N - \sum_{i=1}^{r_j-1} n_i - 1/2 \times (n_{r_j} - 1)$
 $z'_j \leftarrow f(\mathbf{x}_j, t) / nc(\mathbf{x}_j, t)$
 $z''_j \leftarrow \frac{z'_j n_{r_j}}{\sum_{\mathbf{y} \in P^{(t)}, r(\mathbf{y}, t) = r(\mathbf{x}, t)} z'_j(\mathbf{y}, t)} f(\mathbf{x}, t)$ // Seja n_i o nº de indivíduos tal que $r(\mathbf{x}, t) = i$
 fim para
 $Q^{(t)} = \text{SelecaoParesEstocastica}(P^{(t)}, z''_j)$
 $P^{(t+1)} = \text{OperadoresGeneticos}(P^{(t)})$
 $t \leftarrow t + 1$
fim enquanto
retorne $P^{(t)}$

C.3 Algoritmos Genéticos com Nicho de Pareto

O principal diferencial do NPGA — proposto em Horn *et al.* (1994) — em relação a outros MOEAs está na utilização de um operador de seleção para reprodução por torneios, tendo como base a dominância de Pareto (Coello Coello *et al.*, 2005). Nessa proposta, dois indivíduos aleatoriamente selecionados são comparados com uma amostra de suporte da população, também escolhida ao acaso. Os dois indivíduos então competem entre si da seguinte forma: se um deles for dominado por alguma solução daquele subconjunto e o outro não, então o indivíduo não-dominado vence a disputa e é selecionado para a etapa de reprodução. Caso haja empate (ambos são dominados ou não-dominados), o vencedor é determinado mediante uma comparação por meio de compartilhamento de aptidão, e.g. o valor do parâmetro σ_{share} da Equação (C.2).

Horn *et al.* (1994) justificam a comparação baseada em dominância com uma amostra de suporte da população (geralmente 10% da população) pela baixa pressão de dominância ao se comparar os dois indivíduos do torneio entre si. No NPGA, a pressão de dominância é controlada pelo parâmetro t_{dom} , que fornece o tamanho da amostra de suporte utilizada nos torneios. Quanto maior o valor de t_{share} , melhor será a estimativa da dominância de Pareto das soluções envolvidas no torneio.

O esquema de especiação adotado no NPGA, entretanto, possui algumas modificações essenciais para o funcionamento desejado na presença do operador de seleção por torneio e de múltiplos objetivos. Primeiro, para evitar o típico comportamento caótico observado

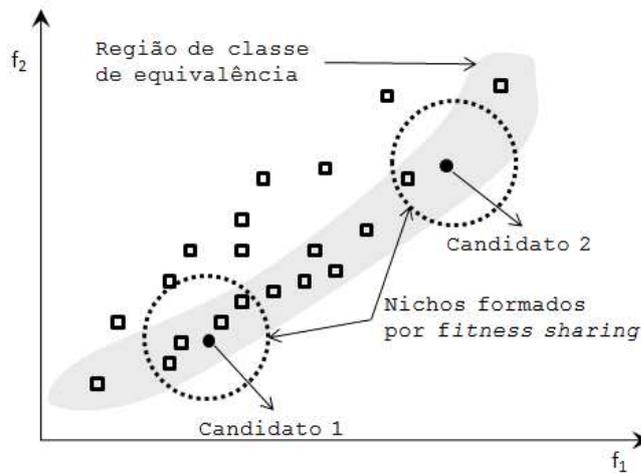


Figura C.2 Representação do processo de compartilhamento de classes de equivalência no NPGA (adaptado de Horn *et al.* (1994)).

em AGs ao se combinar compartilhamento de aptidão com seleção por torneio (Horn *et al.*, 1994), os autores utilizam-se do “*compartilhamento continuamente atualizado*”, proposto por Oei *et al.* (1991), em que o alcance do nicho, $nc(\mathbf{x}, t)$, é calculado utilizando parte dos indivíduos produzidos para a próxima geração. Finalmente, uma vez que o interesse quanto à preservação de diversidade em MOEAs é focado nas Frentes de Pareto e que indivíduos contidos na mesma classe de equivalência (Frente) são considerados equivalentes sob o ponto de vista da otimização multiobjetivo, Horn *et al.* (1994) não adotam a degradação das funções de aptidão baseado em especiação, como ocorre, por exemplo, no MOGA, mas determinam o melhor candidato como aquele de menor número de indivíduos em seu nicho, ou seja, aquele cujo alcance do nicho $nc(\mathbf{x}, t)$ é mínimo. Denominou-se essa abordagem por compartilhamento de classes de equivalência. A visualização do processo de especiação no NPGA, considerando-se dois objetivos, com a maximização do primeiro e a minimização do segundo, pode ser visualizado na Figura C.3.

Uma versão revisada do NPGA, NPGA2, foi proposta por Erickson *et al.* (2001), em que o *rank* de Pareto é utilizado (Seção C.2, Equação (C.1)), mantendo a seleção para reprodução por torneio, como no NPGA original. No entanto, o torneio utilizado no NPGA2 considera o parâmetro t_{size} , que controla o número de soluções aleatoriamente escolhidas para o torneio, e opera da seguinte forma: se dentre as soluções competidoras há uma única solução de menor *rank* de Pareto em relação às demais (ou seja, tal solução é a “menos dominada”), tal indivíduo é declarado vencedor do torneio e é separado

para a reprodução. Caso contrário, se todas as soluções de menor *rank* no torneio são não-dominadas ou são iguais em *rank*, o torneio termina empatado. Nesse último caso, Erickson *et al.* (2001) utilizam um critério de desempate baseado na estimação de densidade de indivíduos no entorno de uma solução (*crowding*), processo que se assemelha àquele realizado para o desempate no algoritmo NSGA2 (Capítulo 3, página 42).

A versão aqui descrita como NPGA2 foi utilizada com sucesso no projeto de tubulações de água subterrânea Erickson *et al.* (2001). Os autores observam que o parâmetro t_{size} afeta a pressão de seleção no NPGA2 de forma a produzir soluções mais espalhadas sobre a \mathcal{F}_1 encontrada para valores altos de t_{size} e, em contrapartida, soluções inferiores para valores muito altos daquele parâmetro. Além disso, o parâmetro σ_{share} utilizado para controlar o processo de especiação e, por conseguinte, a diversidade na população, parece não afetar a qualidade das soluções encontradas. Por outro lado, moderados graus de especiação (valores positivos menores que 1 para σ_{share}) ajudam na obtenção de uma melhor cobertura na \mathcal{F}_1 obtida pelo NPGA2.

C.4 Algoritmos Evolucionários por Força de Pareto

O algoritmo SPEA (Zitzler e Thiele, 2002) combina os pontos fortes de diversos MOEAs previamente existentes de forma a obter um desempenho superior em relação àqueles algoritmos utilizados como base. Segundo os autores, o SPEA possui os seguintes pontos em comum em relação a alguns métodos anteriores:

- armazena as soluções não-dominadas encontradas externamente;
- utiliza-se da dominância de Pareto para atribuir um valor de aptidão aos indivíduos;
- realiza um processo de agrupamento para reduzir o número de soluções não-dominadas sem destruir as características da \mathcal{F}_1 encontrada até o momento.

Por outro lado, os autores indicam, ainda, as seguintes característica — únicas em relação aos demais MOEAs conhecidos na ocasião da publicação do SPEA:

- combina, pela primeira vez, as três técnicas supracitadas em um único algoritmo, em que a aptidão de um indivíduo é computada a partir dos indivíduos armazenados no conjunto externo de soluções não-dominadas — as relações de dominância entre os membros da população são irrelevantes;

- todo o conjunto externo de soluções não-dominadas é utilizado para a determinação da aptidão dos indivíduos na população;
- utiliza-se de um novo método de especiação para a preservação de diversidade na população baseado em Pareto e que não requer parâmetros de distância (e.g. σ_{share}).

Seja N_{PF} o tamanho máximo, definido pelo usuário, do conjunto externo de soluções não-dominadas. A operação do SPEA é então descrita no Pseudocódigo 10.

Pseudocódigo 10 Strength Pareto Evolutionary Algorithm (SPEA)

```

InicializarPopulação( $P$ )
 $PF \leftarrow \emptyset$  // em que  $\mathcal{F}_1$  é o conjunto externo de soluções não-dominadas
enquanto condição de término não satisfeita faça
     $PF \leftarrow P \cup \{\mathbf{x} \in P : \nexists \mathbf{y} \in P, \mathbf{y} \prec \mathbf{x}\}$ 
     $PF \leftarrow PF \setminus \{\mathbf{y} \in P' : \exists \mathbf{z} \in PF, \mathbf{z} \prec \mathbf{y}\}$ 
    se  $|PF| > N_{PF}$  então
         $PF \leftarrow \text{Agrupar}(\mathcal{F}_1, N_{PF})$  // Reduz  $\mathcal{F}_1$  a  $N_{PF}$  soluções por meio de agrupamento
    fim se
    Aptidão( $P$ ), Aptidão( $\mathcal{F}_1$ )
     $P_{\text{pais}} \leftarrow \text{SeleçãoPorTorneio}(P + PF, 2)$  // Seleção por Torneios binários
     $P \leftarrow \text{OperadoresDeVariação}(P_{\text{pais}})$ 
fim enquanto
retorne  $P$ 

```

No Pseudocódigo 10, cumpre mencionar que os valores de aptidão são calculados em duas etapas distintas, uma para as soluções contidas em \mathcal{F}_1 e outra para aquelas contidas em P : na primeira, calcula-se um valor de utilidade para uma solução não-dominada $\mathbf{y} \in PF$, $s_{\mathbf{y}}$, denominado *strength*,

$$s_{\mathbf{y}} = \frac{n}{1 + |P|}, \quad \text{(C.5)}$$

o qual indica uma relação entre os n indivíduos dominados por \mathbf{y} em P e o tamanho da população, $|P|$, e atribui-se o mesmo como valor de aptidão de \mathbf{y} , i.e., $f(\mathbf{y}) = s_{\mathbf{y}}$. Por sua vez, para cada indivíduo \mathbf{x} pertencente à população, são atribuídos valores de aptidão baseados na soma de *strengths* das soluções que o dominam, i.e.,

$$f(\mathbf{x} \in P) = 1 + \sum_{\mathbf{y} \in PF, \mathbf{y} \prec \mathbf{x}} s_{\mathbf{y}}, \text{ em que } f(\mathbf{x}) \in [1, N]. \quad \text{(C.6)}$$

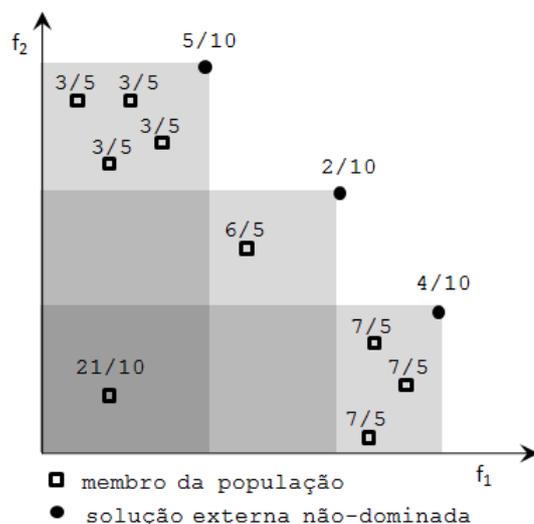


Figura C.3 Exemplo da atribuição de aptidão realizada na população e no conjunto externo de soluções não-dominadas do SPEA (adaptado de Zitzler e Thiele (2002)).

O processo de atribuição de aptidão no SPEA, considerando-se a maximização de dois objetivos, é exemplificado na Figura C.4, em que se percebe uma formação natural de nichos — representados pelas regiões sombreadas retangulares — com base na dominância de Pareto, o que, portanto, elimina a necessidade da existência de um parâmetro de compartilhamento de aptidão baseado em distância (e.g. σ share no MOGA e NPGA) no SPEA. Às soluções contidas em nichos populosos, dominados por uma solução externa, são atribuídos valores mais altos de aptidão (considera-se minimização da aptidão). Quanto mais próxima a solução da \mathcal{F}_1 , menor será a sua aptidão. Ainda, quanto maior a utilidade (*strength*) das soluções não-dominadas, pior avaliadas serão as soluções posicionadas nos nichos dominados por aquelas soluções. Desse modo, o SPEA ataca simultaneamente o problema da preservação da diversidade, por meio de especiação, e o da convergência à *PF*.

Finalmente, o algoritmo de agrupamento adotado para reduzir o conjunto externo de soluções não-dominadas (\mathcal{F}_1) no SPEA, denominado *average linkage* (Morse, 1980), considera cada solução externa contida em \mathcal{F}_1 como um grupo (*cluster*), realizando operações de fusão (*merge*) sucessivas entre pares de grupos que estiverem à menor distância dentre quaisquer outros pares contidos no conjunto externo. A cada operação de fusão, o centróide (vetor médio tomado no espaço genótipo) calculado sobre os elementos do conjunto fundido de pares de menor distância (digamos, $C = \{\mathbf{x}_i \cup \mathbf{x}_j\}$) substitui C em \mathcal{F}_1 até que $|PF| < N_{PF}$. Zitzler e Thiele (2002) ressaltam a capacidade do algoritmo

average linkage de preservar a estrutura original de \mathcal{F}_1 e de preservar a diversidade de \mathcal{F}_1 no processo de redução.

O SPEA foi avaliado na versão multiobjetivo do problema da mochila (*knapsack*), no qual várias mochilas — no *knapsack* de um único objetivo há apenas uma mochila (cada mochila acrescenta um objetivo ao problema) — são preenchidas com diversos itens, cada qual fornecendo um valor de lucro, e estando limitada a uma capacidade máxima de itens que podem ser carregados (Zitzler e Thiele, 2002). Os autores demonstraram que o SPEA foi capaz de produzir melhores soluções, além de uma melhor cobertura da \mathcal{F}_1 obtida, em relação a outros MOEAs, como NPGA, VEGA e NSGA (Seção 3.5.1).

Embora o SPEA tenha obtido resultados promissores quando comparado a outros métodos até então existentes, algumas limitações podem ser identificadas (Zitzler *et al.*, 2004):

- **Atribuição de aptidão:** indivíduos nos mesmos nichos recebem valores de aptidão idênticos, fato que, em casos cujo tamanho do conjunto de não-dominadas, \mathcal{F}_1 , é pequeno — por exemplo, quando há apenas um indivíduo em \mathcal{F}_1 —, resulta na atribuição dos mesmos valores de aptidão a todas as demais soluções na população, o que elimina a pressão de seleção e faz com que o SPEA comporte-se como um algoritmo de busca aleatória;
- **Estimação da densidade:** em situações nas quais um grande número de indivíduos são não-comparáveis entre si, as relações de dominância fornecem pouca informação para a construção de nichos úteis para a etapa de seleção de pares para reprodução. Daí a necessidade de técnicas que estimem a densidade de indivíduos no entorno de cada solução (i.e., *crowding*);
- **Poda do arquivo externo:** apesar do algoritmo *average linkage* preservar a estrutura da \mathcal{F}_1 no processo de redução, soluções extremas podem ser descartadas no processo, diminuindo os níveis de cobertura da \mathcal{F}_1 .

A versão revisada do SPEA, SPEA2 (Zitzler *et al.*, 2001), trata das limitações supracitadas da seguinte forma (Coello Coello *et al.*, 2005):

- utiliza uma atribuição de aptidão a qual considera tanto o número de indivíduos que dominam a solução quanto o número de indivíduos dominados pela mesma, o que permite a diferenciação de soluções contidas no mesmo nicho;
- utiliza a técnica *K-Vizinhos Mais Próximos* (VMP), a qual fornece a distância σ_x^K de um indivíduo x para o seu K -ésimo VMP no conjunto $P + PF$ (os autores

utilizam $K = 1$ por questões de eficiência). Um valor inversamente proporcional à $\sigma_{\mathbf{x}}^K$ é então interpretado como uma estimativa válida da densidade de soluções no entorno de \mathbf{x} e tem então o seu valor adicionado ao valor de aptidão original de \mathbf{x} — baseado nas relações de dominância — de forma a penalizar as soluções que estão contidas em nichos mais densos.

- utiliza um método mais eficiente de truncamento do arquivo externo de soluções não-dominadas, o qual garante a preservação das soluções extremas da \mathcal{F}_1 obtida.

Os resultados obtidos nos testes comparativos com o SPEA2 demonstraram que o mesmo superou a versão anterior (SPEA) em todos os casos. No quadro comparativo traçado pelos autores, o SPEA2 não obteve a melhor velocidade de convergência dentre todos os MOEAs testados devido ao fato — sugerido pelos autores — de que os níveis de elitismo mais intensos de outros MOEAs contribuem para a maior velocidade desses algoritmos, mas que, no entanto, falham ao não preservarem soluções extremas na \mathcal{F}_1 e não produzirem uma melhor cobertura do conjunto de Pareto. Ainda, avaliou-se o desempenho do *SPEA2* como estatisticamente comparável à do *NSGA2*, mas avaliou-se a escalabilidade desse em relação àquele como menor em problemas de mais alta dimensionalidade, nos quais o mecanismo de estimação de densidade proposto no *SPEA2* pode ter sido o fator a ter contribuído para um melhor desempenho.



Biografia e Publicações

D.1 Biografia Resumida

Carlos Renato Belo Azevedo é Bacharel em Ciência da Computação (2008) pela Universidade Católica de Pernambuco (UNICAP), tendo recebido a Láurea Acadêmica pelo seu desempenho no curso, e Mestre em Ciência da Computação pela Universidade Federal de Pernambuco (UFPE). Foi aluno de Iniciação Científica (IC) e de mestrado com bolsa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Os projetos desenvolvidos quando aluno de IC obtiveram a segunda e primeira colocações entre todos os trabalhos de IC da UNICAP na grande área Ciências Exatas e da Terra em 2007 e em 2008, respectivamente. Em 2010, foi um dos 80 selecionados dentre mais de 1.600 candidatos de todo o mundo para participar do *Graduate Studies Program* da Singularity University na NASA Ames Research Center, Califórnia, onde estudou tecnologias de crescimento exponencial, participando de pesquisa coordenada por Chris Lewick, ex-engenheiro sênior do Laboratório de Propulsão de Jato (JPL) da NASA, que atuou como diretor de vôo dos robôs de exploração Spirit e Opportunity e como gerente de navegação da sonda robótica *Phoenix* enviada a Marte em 2008. Carlos é membro estudante da ACM e do grupo de interesse em computação genética e evolucionária (ACM/SIGEVO) desde 2009, possuindo interesses de pesquisa experimental e teórica nos seguintes temas: computação evolucionária e bio-inspirada, redes neurais artificiais, otimização dinâmica multiobjetivo, reconhecimento de padrões, previsão de séries temporais, simulação estocástica e robótica. Atualmente encontra-se vinculado ao Programa de Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação na Universidade Estadual de Campinas (UNICAMP), onde desenvolve atividades de pesquisa como aluno de doutorado no Laboratório de Bioinformática e Computação Bio-Inspirada (LBiC).

Currículo Lattes: <http://lattes.cnpq.br/3516550996242948>

Página do LBiC: <http://www.lbic.fee.unicamp.br/>

Carlos pode ser contactado pelo endereço eletrônico azevedo@dca.fee.unicamp.br.

D.2 Publicações Seleccionadas

1. Aranha, C., Azevedo, C. R. B. e Iba, H. (aceito para publicação). Money in Trees: How Memes, Trees and Isolation Can Optimize Financial Portfolios. *Information Science*. Elsevier.
2. Bispo Jr., E. L., Azevedo, C. R. B., Lopes, W. T., Madeiro, F., Alencar, M. S. (2010). Methods to Accelerate a Competitive Learning Algorithm Applied to VQ Codebook Design. *Tendências em Matemática Aplicada e Computacional*, v. 11(3), 193–203. Sociedade Brasileira de Matemática Aplicada e Computacional.
3. Azevedo, C. R. B. *et al.* (2010). Exponential Technologies Impact Report: Space (relatório de pesquisa). Singularity University/NASA Ames.
4. Azevedo, C. R. B. e Gordon, V. S. (2009). Adaptive Terrain-Based Memetic Algorithms. GECCO'09. 747–754. Association for Computing Machinery Inc. (ACM).
5. Azevedo, C. R. B., Azevedo, F. E. A. G., Lopes, W. T. e Madeiro, F. (2009). Terrain-Based Memetic Algorithms for Vector Quantizer Design. *Studies in Computational Intelligence*, v. 236, 197–211. Springer.
6. Azevedo, C. R. B., Ferreira, T. A. F., Lopes, W. T. e Madeiro, F. (2008). Improving Image Vector Quantization with a Genetic Accelerated K-Means Algorithm. ACIVS'08. *Lecture Notes in Computer Science*, v. 5259, 67–76. Springer.
7. Azevedo, C. R. B., Azevedo, R. A., Bispo Jr. E. L., Ferreira, T. A. E., Lopes, W. T. e Madeiro, F. (2008) Um Algoritmo Memético para a Otimização de Quantizadores Vetoriais. *Learning and Nonlinear Models*, v. 6, 1–15. Sociedade Brasileira de Redes Neurais.
8. Azevedo, C. R. B., Ferreira, T. A. E. (2007). Time Series Forecasting with Qubit Neural Networks. IASTED ASC'07. 13–18. ACTA Press.

Referências Bibliográficas

- Altrock, P. M., Traulsen, A., Reeves, R. G., e Reed, F. a. (2010). Using underdominance to bi-stably transform local populations. *Journal of Theoretical Biology*, **267**(1), 62–75.
- Araújo, A. e Garrozi, C. (2010). Mulroga: A multicast routing genetic algorithm approach considering multiple objectives. *Applied Intelligence*, **32**, 330–345. 10.1007/s10489-008-0148-5.
- Assenza, A., Valle, M., e Verleysen, M. (2008). A comparative study of various probability density estimation methods for data analysis. *International Journal of Computational Intelligence Systems*, **1**(2), 188.
- Auger, A., Bader, J., Brockhoff, D., e Zitzler, E. (2009). Theory of the hypervolume indicator: Optimal μ -distributions and the choice of the reference point. Nos *Anais do Tenth ACM SIGEVO Workshop on Foundations of Genetic Algorithms*, FOGA '09, páginas 87–102.
- Azevedo, C. R. B. e Gordon, V. S. (2009). Adaptive terrain-based memetic algorithms. Nos *Anais da 11th Annual Conference on Genetic and Evolutionary Computation*, GECCO '09, páginas 747–754.
- Bader, J. M. (2010). *Hypervolume-Based Search for Multiobjective Optimization: Theory and Methods*. Swiss Federal Institute of Technology Zurich, Zurich, 1ª edição.
- Barakat, H. (2009). Multivariate order statistics based on dependent and nonidentically distributed random variables. *Journal of Multivariate Analysis*, **100**(1), 81–90.
- Benayoun, R., De Montgolfier, J., Tergny, J., e Laritchev, O. (1971). Linear programming with multiple objective functions: Step method (STEM). *Mathematical Programming*, **1**(1), 366–375.
- Beume, N., Naujoks, B., e Emmerich, M. (2007). SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, **181**(3), 1653–1669.
- Beume, N., Laumanns, M., e Rudolph, G. (2011). Convergence rates of (1+1) evolutionary multiobjective optimization algorithms. Nos *Anais da Parallel Problem Solving from Nature - PPSN XI*, volume 6238 da *Lecture Notes in Computer Science*, páginas 597–606. Springer Berlin / Heidelberg.

- Bilgin, Z. e Khan, B. (2010). A dynamic route optimization mechanism for AODV in MANETs. Nos *Anais da IEEE International Conference on Communications (ICC)*, páginas 1–5.
- Bosman, P. e Thierens, D. (2003). The balance between proximity and diversity in multi-objective evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, **7**(2), 174–188.
- Branke, J. (1999). Memory enhanced evolutionary algorithms for changing optimization problems. Nos *Anais do Congress on Evolutionary Computation*, volume 3, páginas 1875–1882.
- Burke, E., Gustafson, S., e Kendall, G. (2004). Diversity in genetic programming: An analysis of measures and correlation with fitness. *IEEE Transactions on Evolutionary Computation*, **8**(1), 47–62.
- Carmichael, D. (1980). Computation of pareto optima in structural design. *International Journal for Numerical Methods in Engineering*, **15**(6), 925–952.
- Chen, J., Xin, B., Peng, Z., Dou, L., e Zhang, J. (2009). Optimal contraction theorem for exploration/exploitation tradeoff in search and optimization. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, **39**(3), 680–691.
- Cheng, H. e Yang, S. (2010). Genetic algorithms with immigrants schemes for dynamic multicast problems in mobile ad hoc networks. *Engineering Applications of Artificial Intelligence*, **23**(5), 806–819.
- Cobb, H. e Grefenstette, J. (1993). Genetic algorithms for tracking changing environments. Nos *Anais da International Genetic Algorithms Conference*, páginas 1–8.
- Coello Coello, C. A. (2000). An updated survey of GA-based multiobjective optimization techniques. *ACM Computing Surveys (CSUR)*, **32**(2), 143.
- Coello Coello, C. A., Pulido, G. T., e Montes, E. M. (2005). *Current and Future Research Trends in Evolutionary Multiobjective Optimization*, páginas 213–231. Springer-Verlag, Berlin, Heidelberg.
- Corley, H. (1984). Multivariate order statistics. *Communications in Statistics - Theory and Methods*, **13**(10), 1299–1304.

- Cvetkovic, D., Parmee, I., e Webb, E. (1998). *Multi-Objective Optimisation and Preliminary Airframe Design*, páginas 255–267. Springer-Verlag.
- Das, I. e Dennis, J. (1996). Normal-boundary intersection: An alternate method for generating pareto optimal points in multicriteria optimization. Relatório técnico, NASA Langley Research Center.
- Deb, K. (2001). *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley.
- Deb, K. e Agrawal, R. (1995). Simulated binary crossover for continuous search space. *Complex Systems*, **9**(2), 115–148.
- Deb, K. e Goyal, M. (1996). A combined genetic adaptive search (GeneAS) for engineering design. *Computer Science and Informatics*, **26**(4), 30–45.
- Deb, K., Agrawal, S., Pratap, A., e Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. Nos *Parallel Problem Solving from Nature PPSN VI, Lecture Notes in Computer Science*, páginas 849–858. Springer.
- Deb, K., Thiele, L., Laumanns, M., e Zitzler, E. (2002). Scalable multi-objective optimization test problems. Nos *Anais do Congress on Evolutionary Computation. CEC'02*, volume 1, páginas 825–830. IEEE.
- Deb, K., Bhaskara Rao, N. U., e Karthik, S. (2007). Dynamic multi-objective optimization and decision-making using modified NSGA-II: A case study on hydro-thermal power scheduling. Nos *Anais da 4th international conference on Evolutionary multi-criterion optimization, EMO'07*, páginas 803–817.
- Ding, N. e Zhou, S. (2008). *Linkages Detection in Histogram-Based Estimation of Distribution Algorithm*, páginas 25–40. Springer.
- Duda, R., Hart, P., e Stork, D. (2000). *Pattern Classification*. Wiley, 2ª edição.
- Eiben, A. E. e Schippers, C. A. (1998). On evolutionary exploration and exploitation. *Fundam. Inf.*, **35**, 35–50.
- Eichfelder, G. (2008). *Adaptive Scalarization Methods in Multiobjective Optimization*. Springer Verlag.
-

- Erickson, M., Mayer, A., e Horn, J. (2001). The niched pareto genetic algorithm 2 applied to the design of groundwater remediation systems. Nos *Anais da First International Conference on Evolutionary Multi-Criterion Optimization*, EMO '01, páginas 681–695.
- Eshelman, L. J., Caruana, R. A., e Schaffer, J. D. (1989). Biases in the crossover landscape. Nos *Anais da Third International Conference on Genetic Algorithms*, páginas 10–19.
- Farina, M., Deb, K., e Amato, P. (2004). Dynamic multiobjective optimization problems: Test cases, approximations, and applications. *IEEE Transactions on Evolutionary Computation*, **8**(5), 425–442.
- Fleischer, M. (2003). The measure of pareto optima applications to multi-objective metaheuristics. Nos *Anais da 2nd International Conference on Evolutionary Multi-Criterion Optimization*, EMO'03, páginas 519–533.
- Fonseca, C. e Fleming, P. (1995). Multiobjective genetic algorithms made easy: Selection, sharing and mating restriction. Nos *Anais da First IEE/IEEE International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications*, páginas 42–52.
- Freund, D. E., Burlina, P., Banerjee, A., e Justen, E. (2009). Comparison of kernel based pdf estimation methods. Nos *Anais do SPIE*, páginas 733508–733508–14.
- Fu, L., Liu, Q., e Calamai, P. (2003). Real-time optimization model for dynamic scheduling of transit operations. *Transportation Research Record*, **1857**(1), 48–55.
- Garrozi, C. e Araujo, A. (2006). Multiobjective genetic algorithm for multicast routing. Nos *Anais da IEEE International Conference on Evolutionary Computation*, páginas 2513–2520.
- Geman, S. e Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**(6), 721–741.
- Georgakis, A., Kotropoulos, C., Xafopoulos, A., e Pitas, I. (2004). Marginal median SOM for document organization and retrieval. *Neural Networks*, **17**(3), 365–77.
- Goh, C.-K. e Tan, K. C. (2009). A competitive-cooperative coevolutionary paradigm for dynamic multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, **13**(1), 103–127.

- Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.
- Goldberg, D. E. e Richardson, J. (1987). Genetic algorithms with sharing for multimodal function optimization. Nos *Anais da Second International Conference on Genetic Algorithms and Their Application*, páginas 41–49.
- Gouvêa Jr., M. M. (2009). *Algoritmo Evolucionário Adaptável em Problemas Dinâmicos Multimodais*. Tese de doutorado, Universidade Federal de Pernambuco.
- Gouvêa Jr., M. M. e Araújo, A. F. R. (2010). *Diversity-Based Adaptive Evolutionary Algorithms*, páginas 1–16. InTech, Rijeka, Croatia.
- Guan, S.-u. e Zhang, S. (2003). An evolutionary approach to the design of controllable cellular automata structure for random number generation. *IEEE Transactions on Evolutionary Computation*, **7**(1), 23–36.
- Haimes, Y., Ladson, L., e Wismer, D. (1971). Bicriterion formulation of problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC1**(3), 296–297.
- Handl, J. e Knowles, J. (2007). An evolutionary approach to multiobjective clustering. *IEEE transactions on Evolutionary Computation*, **11**(1), 56–76.
- Handl, J., Lovell, S. C., e Knowles, J. (2008). Multiobjectivization by decomposition of scalar cost functions. Nos *Anais da 10th International Conference on Parallel Problem Solving from Nature: PPSN X*, páginas 31–40.
- Hassan, G. N. A. (2010). *Multiobjective Genetic Programming for Financial Portfolio Management in Dynamic Environments*. Tese de doutorado, University College London.
- Hohm, T. e Zitzler, E. (2009). Multiobjectivization for parameter estimation: A case-study on the segment polarity network of drosophila. Nos *Anais da 11th Annual Conference on Genetic and Evolutionary Computation, GECCO '09*, páginas 209–216.
- Horn, J. (1997). The nature of niching: Genetic algorithms and the evolution of optimal, cooperative populations. Relatório técnico, University of Illinois at Urbana-Champaign, Champaign, IL, USA.

REFERÊNCIAS BIBLIOGRÁFICAS

- Horn, J., Nafpliotis, N., e Goldberg, D. E. (1994). A niched pareto genetic algorithm for multiobjective optimization. Nos *Anais da International Conference on Evolutionary Computation*, páginas 82–87.
- Hornby, G. S. (2009). Steady-state ALPS for real-valued problems. Nos *Anais da 11th Annual Conference on Genetic and Evolutionary Computation, GECCO '09*, páginas 795–802.
- Hruschka, E. R., Campello, R. J. G. B., Freitas, A. A., e De Carvalho, A. C. P. L. F. (2009). A survey of evolutionary algorithms for clustering. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, **39**(2), 133–155.
- Ishibuchi, H. e Narukawa, K. (2005). Recombination of similar parents in EMO algorithms. Nos C. A. C. Coello, A. H. Aguirre, e E. Zitzler, editors, *Anais da Third International Conference on Evolutionary Multi-Criterion Optimization (EMO'05)*, volume 3410 da *Lecture Notes in Computer Science*, páginas 265–279. Springer.
- Jähne, M., Li, X., e Branke, J. (2009). Evolutionary algorithms and multi-objectivization for the travelling salesman problem. Nos *Anais da 11th Annual Conference on Genetic and Evolutionary*, páginas 595–602.
- Jain, A., Murty, M., e Flynn, P. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, **31**(3), 264–323.
- Jensen, M. T. (2004). Helper-objectives: Using multi-objective evolutionary algorithms for single-objective optimisation. *Journal of Mathematical Modelling and Algorithms*, **3**(4), 323–347.
- Jeong, I. e Kim, K. (2005). D-STEM: A modified step method with desirability function concept. *Computers and Operations Research*, **32**(12), 3175—3190.
- Jin, Y. e Branke, J. (2005). Evolutionary optimization in uncertain environments: A survey. *IEEE Transactions on Evolutionary Computation*, **9**(3), 303–317.
- Jones, D., Mirrazavi, S., e Tamiz, M. (2002). Multi-objective meta-heuristics: An overview of the current state-of-the-art. *European Journal of Operational Research*, **137**(1), 1–9.
- Kennedy, J. e Eberhart, R. (1995). Particle swarm optimization. Nos *Anais da IEEE International Conference on Neural Networks*, volume 4, páginas 1942–1948.

- Kirkpatrick, S., Gelatt, C. D., e Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, **220**(4598), 671–680.
- Knowles, J. D., Watson, R. A., e Corne, D. (2001). Reducing local optima in single-objective problems by multi-objectivization. Nos *Anais da First International Conference on Evolutionary Multi-Criterion Optimization*, EMO '01, páginas 269–283.
- Konak, A., Coit, D., e Smith, A. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, **91**(9), 992–1007.
- Korkmaz, E., Du, J., Alhaji, R., e Barker, K. (2006). Combining advantages of new chromosome representation scheme and multi-objective genetic algorithms for better clustering. *Intelligent Data Analysis*, **10**(2), 163–182.
- Kullback, S. e Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, **22**(1), 79–86.
- Loesch, C. e Hein, N. (2008). *Pesquisa Operacional: Fundamentos e Modelos*. Saraiva, São Paulo, 1ª edição.
- Marler, R. e Arora, J. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, **26**(6), 369–395.
- Masud, A. S. M. e Ravindran, A. R. (2008). *Multiple Criteria Decision Making*, páginas 95–135. CRC Press, Boca Raton, FL.
- Messac, A. (1996). Physical programming - effective optimization for computational design. *AIAA Journal*, **34**(1), 149–158.
- Miller, B. L. e Goldberg, D. E. (1996). Genetic algorithms, selection schemes, and the varying effects of noise. *Evolutionary Computation*, **4**(2), 113–131.
- Montgomery, D. C. (2004). *Design and Analysis of Experiments*. John Wiley & Sons.
- Morrison, R. W. e De Jong, K. A. (2002). Measurement of population diversity. *Population (English Edition)*, **1**(2), 31–41.
- Morse, J. (1980). Reducing the size of the nondominated set: Pruning by clustering. *Computers & Operations Research*, **7**(1-2), 55–66.
- Müller, M., Sperlich, S., Werwatz, A., HSrdle, W., e Mnller, M. (2004). *Nonparametric and Semiparametric Models*. Springer.

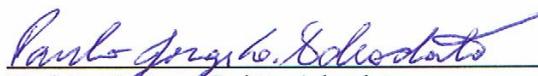
REFERÊNCIAS BIBLIOGRÁFICAS

- Nocedal, J. e Wright, S. (2006). *Numerical optimization*. Springer, Berlin Heidelberg, 2ª edição.
- Oei, C. K., Goldberg, D. E., e Chang, S. (1991). Tournament selection, niching, and the preservation of diversity. Relatório técnico, Illinois Genetic Algorithms Laboratory (IlliGAL), Urbana-Champaign.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, **33**(3), 1065–1076.
- Pereira, F. L. (2001). Control design for autonomous vehicles: A dynamic optimization perspective. *European Journal of Control*, **7**(2-3), 178–202.
- Pitas, I., Kotropoulos, C., Nikolaidis, N., Yang, R., e Gabbouj, M. (1996). Order statistics learning vector quantizer. *IEEE Transactions on Image Processing*, **5**(6), 1048–53.
- Reed, D. H. e Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology*, **17**(1), 230–237.
- Richardson, J., Palmer, M., Liepins, G., e Hilliard, M. (1989). Some guidelines for genetic algorithms with penalty functions. Nos *Anais da 3rd International Conference on Genetic Algorithms Table of Contents*, páginas 191–197.
- Ridley, M. (2003). *Evolution*. Wiley-Blackwell.
- Ross, S. (2001). *A First Course in Probability*. Prentice Hall, 6ª edição.
- Schaffer, J. (1985). Multiple objective optimization with vector evaluated genetic algorithms. Nos *Anais da 1st International Conference on Genetic Algorithms and Their applications*, páginas 93–100.
- Schwefel, H.-P. (1965). *Kybernetische Evolution als Strategie der experimentellen Forschung in der Strömungstechnik*. Dipl.-Ing. Thesis, Technical University of Berlin, Hermann Föttinger-Institute for Fluid Dynamics.
- Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization (Wiley Series in Probability and Statistics)*. Wiley.
- Shukla, P. K. e Deb, K. (2007). On finding multiple pareto-optimal solutions using classical and evolutionary generating methods. *European Journal of Operational Research*, **181**(3), 1630 – 1652.

- Smith, J. E. (1998). *Self Adaptation in Evolutionary Algorithms*. Tese de doutorado, University of the West England, Bristol.
- Sola, M. C. (2010). *Parallel Processing for Dynamic Multi-objective Optimization*. Tese de doutorado, Universidad de Granada.
- Tang, B. (1993). Orthogonal Array-Based Latin Hypercubes. *Journal of the American Statistical Association*, **88**(424).
- Ursem, R. K. (2002). Diversity-guided evolutionary algorithms. Nos *Anais da 7th International Conference on Parallel Problem Solving from Nature*, PPSN VII, páginas 462–474.
- Valdez-Peña, S. I., Hernández-Aguirre, A., e Botello-Rionda, S. (2009). Approximating the search distribution to the selection distribution in EDAs. Nos *Anais da 11th Annual Conference on Genetic and Evolutionary Computation*, GECCO '09, páginas 461–468.
- Wald, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. *The Annals of Mathematical Statistics*, **10**(4), 299–326.
- Watanabe, S. e Sakakibara, K. (2007). A multiobjectivization approach for vehicle routing problems. Nos *Anais da 4th International Conference on Evolutionary Multi-Criterion Optimization*, EMO'07, páginas 660–672.
- Webb, A. (2002). *Statistical Pattern Recognition*, volume 9. John Wiley & Sons, 2ª edição.
- Welch, B. L. (1947). The generalization of “student’s” problem when several different population variances are involved. *Biometrika*, **34**(1-2), 28–35.
- Wolpert, D. e Macready, W. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, **1**(1), 67–82.
- Woodward, J. (2010). The necessity of meta bias in search algorithms. Nos *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on*, páginas 1–4.
- Yang, S. (2008). Genetic algorithms with memory and elitism-based immigrants in dynamic environments. *Evolutionary Computation*, **16**(3), 385–416.
-

- Yang, S. e Yao, X. (2005). Experimental study on population-based incremental learning algorithms for dynamic optimization problems. *Soft Computing*, **9**(11), 815–834.
- Yang, S. e Yao, X. (2008). Population-based incremental learning with associative memory for dynamic environments. *IEEE Transactions on Evolutionary Computation*, **12**(5), 542–561.
- Yu, X., Tang, K., e Yao, X. (2008). An immigrants scheme based on environmental information for genetic algorithms in changing environments. Nos *Anais do IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, number 3 em CEC '08, páginas 1141–1147. IEEE Press.
- Yu, X., Tang, K., Chen, T., e Yao, X. (2009). Empirical analysis of evolutionary algorithms with immigrants schemes for dynamic optimization. *Memetic Computing*, **1**(1), 3–24.
- Yu, X., Tang, K., e Yao, X. (2010). Immigrant schemes for evolutionary algorithms in dynamic environments: Adapting the replacement rate. *Science in China Series F - Information Sciences*, **53**, 1–11.
- Zadeh, L. (1963). Optimality and non-scalar-valued performance criteria. *IEEE Transactions on Automatic Control*, **8**(1), 59–60.
- Zitzler, E. e Thiele, L. (2002). Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach. *IEEE Transactions on Evolutionary Computation*, **3**(4), 257–271.
- Zitzler, E., Laumanns, M., e Thiele, L. (2001). SPEA2: Improving the strength pareto evolutionary algorithm. Relatório Técnico 103, Swiss Federal Institute of Technology (ETH) Zurich, Zurich.
- Zitzler, E., Deb, K., e Thiele, L. (2002). Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, **8**(2), 173–195.
- Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C., e da Fonseca, V. (2003). Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, **7**(2), 117–132.
- Zitzler, E., Laumanns, M., e Bleuler, S. (2004). A tutorial on evolutionary multiobjective optimization. *Metaheuristics for Multiobjective Optimisation*, **535**, 3–37.

Dissertação de Mestrado apresentada por **Carlos Renato Belo Azevedo** à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título , “**Geração de Diversidade na Otimização Dinâmica Multiobjetivo Evolucionária por Paisagens de Não-Dominância**”, orientada pelo **Prof. Aluizio Fausto Ribeiro Araújo** e aprovada pela Banca Examinadora formada pelos professores:



Prof. Paulo Jorge Leitão Adeodato
Centro de Informática / UFPE

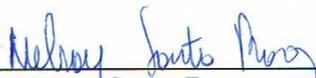


Prof. Carmelo José Albahez Bastos Filho
Departamento de Sistemas Computacionais / UPE



Prof. Aluizio Fausto Ribeiro Araújo
Centro de Informática / UFPE

Visto e permitida a impressão.
Recife, 4 de fevereiro de 2011.



Prof. Nelson Souto Rosa

Coordenador da Pós-Graduação em Ciência da Computação do
Centro de Informática da Universidade Federal de Pernambuco.

