On the Evolutionary Optimisation of Many Objectives

Thesis submitted for candidature for the degree of PhD

Robin Charles Purshouse

September 2003

Department of Automatic Control and Systems Engineering
The University of Sheffield

Summary

Evolutionary algorithms based on the concept of Pareto dominance are popular and successful techniques for multi-objective optimisation. Such algorithms are required to obtain a good approximation (in terms of proximity and diversity) to trade-off surfaces of interest to the decision-maker.

To achieve good performance, the design of an evolutionary optimiser should be component-based and application-focused. Multi-objective algorithms tend to be analysed as composites using empirical approaches that lack statistical rigour. This severely limits the availability of process knowledge, itself essential for component-level design. The thesis addresses this issue with a new, structured, framework for algorithm development and assessment. A rigorous, nonparametric statistical methodology is applied to appropriate performance indicators for proximity and spread. Applications of the framework can be found throughout the thesis.

Evolutionary multi-objective theory mainly considers bi-objective problems. However, multi-objective applications often consider many more objectives. Thus, more theoretical work into the optimisation of many objectives is required. A platform for such research is established via consideration of the relationships that may exist between pairs of objectives. Three relationships — conflict, harmony, and independence — are identified, and issues and current research are discussed for each relationship.

The effect of many conflicting objectives on a class of popular evolutionary multi-objective optimisers is considered in a detailed, exploratory investigation. The study reveals that conclusions drawn from bi-objective analysis cannot be generalised to higher numbers of conflicting objectives. The probable mechanisms that underpin the observed variation in behaviour are also identified and discussed.

The thesis also demonstrates, using the rigorous empirical framework, that if independence exists in a multi-objective problem, then identification and exploitation of this relationship can produce improved optimiser results. An innovative method for identifying suitable objective-space decompositions on-line is subsequently developed, based on concepts from parallel evolutionary topologies and nonparametric statistics. Excellent results are obtained in a proof-of-principle assessment.

Acknowledgments

I would like to thank Peter Fleming for giving me the opportunity to do postgraduate research, for his expert support and advice, and for packing me off to nice places every so often.

My thanks also go to Nick Fieller for introducing me to randomisation methods and to the UK Engineering and Physical Sciences Research Council for research studentship support.

Further thank-yous are offered to my colleagues, my family, and my friends.

Statement of Originality

Unless otherwise stated in the text, the work described in this thesis was carried out solely by the candidate. None of this work has already been accepted for any other degree, nor is it being concurrently submitted in candidature for any degree.

Candidate: _____

Robin C. Purshouse

Supervisor: _____

Peter J. Fleming

For Tina out of S-Club.

Or maybe Rachel.

Nomenclature

Acronyms (AGA to OR)

AGA	adaptive grid archiving
BMDA	bivariate marginal distribution algorithm
BOA	Bayesian optimisation algorithm
D1,2	decompositions of NSGA-II
DC	divide-and-conquer
DM	decision-maker
DTLZ	Deb, Thiele, Laumanns, and Zitzler
EA	evolutionary algorithm
EC	evolutionary computing
EDA	estimation of distribution algorithm
e-e	exploration versus exploitation
E$\mathbb{M}$O	evolutionary many-objective optimisation
EMO	evolutionary multi-objective optimisation
ES	evolution strategies
GA	genetic algorithm
IC	independent collection
IFAC	International Federation of Automatic Control
MCDM	multi-criteria decision-making
$\mathbb{M}$	many-objective
$\mathbb{M}$IDEA	multi-objective mixture-based iterated density estimation EA
MO	multi-objective optimisation
MOEA	multi-objective evolutionary algorithm
MOGA	multi-objective genetic algorithm
NFL	no-free-lunch theorem
NN	nearest neighbour
NPGA	niched Pareto genetic algorithm
NSGA	non-dominated sorting genetic algorithm
NSGA-II	elitist non-dominated sorting genetic algorithm
OR	operations research

Acronyms (PAES to ZDT)

PAES	Pareto archived evolution strategy
PBIL	population-based incremental learning
p.d.f.	probability density function
PESA(-II)	Pareto envelope selection algorithm (revised)
PROMETHEE	preference ranking organisation method for enrichment evaluations
ROI	region of interest
SBX	simulated binary crossover
SOM	self-organising map
SPEA(2)	strength Pareto evolutionary algorithm (revised)
TSP	travelling salesman problem
VEGA	vector evaluated genetic algorithm
ZDT	Zitzler, Deb, and Thiele

Greek Symbols

α	fitness sharing shaping parameter
β	parameter internal to polynomial mutation
γ_i	parent value for the ith decision variable
Δ	spread metric
δ	parameter internal to SBX
ϵ_i	DM indifference threshold on the ith objective
ζ_M	volume of the unit M-dimensional hypersphere
η_c	SBX distribution parameter
η_m	polynomial mutation distribution parameter
κ	dimension of quadratic bowl in DTLZ2 g functional
λ	maximum spread indicator
ρ	index in sorted list
σ_{mate}	mating restriction neighbourhood size
σ_{share}	fitness sharing niche size
ψ	number of independent collections

Roman Symbols (a to Q)

a	index to an objective vector instance
b	index to a different objective vector instance
c_i	child value for the ith decision variable
D	Hoeffding or Blum-Kiefer-Rosenblatt statistic
d	distance metric
f_i	fitness of the ith solution
g	proximity-related functional
GD	generational distance metric
h	smoothing parameter
i	index variable
I_P	proximity indicator
I_S	spread indicator
j	index variable
K	Kendall sample correlation coefficient
k	index in sorting of nearest neighbours to a candidate solution
K_e	Epanechnikov kernel density estimator
$\mathcal{L}$	local region of decision-space
l_i	lower bound on the ith decision variable
M	number of objectives
n	number of decision variables
$\tilde{n}$	number of decision variables per sub-problem
p_c	probability of applying recombination to a solution
p_e	probability of exchanging recombined decision variables
$\mathcal{P}_i$	ith sub-problem
p_{ic}	probability of applying recombination to a decision variable
p_m	probability of mutating a decision variable
p(mutate)	probability of applying mutation to a solution
p(recombine)	probability of applying recombination to a solution
$P[t]$	population at time t
Q	number of objectives in independent collection

Roman Symbols (q to $\boldsymbol{Z_R}$)

q	index variable
$\boldsymbol{R}$	transformation matrix
r_i	random number generated uniformly from [0 1]
S	hypervolume metric
s	selection process
Sh	sharing function
s_s	selection-for-survival process
s_v	selection-for-variation process
$\boldsymbol{T}$	sample covariance matrix
t	time or iteration number
$\mathcal{U}$	feasible region of decision-space
$\boldsymbol{u}$	candidate solution
u_i	upper bound on the ith decision variable
$\boldsymbol{v}$	candidate solution
v	variation process
$\boldsymbol{X_*}$	Pareto optimal solution set
$\boldsymbol{X_*^\epsilon}$	ϵ-Pareto optimal solution set
x_i	ith decision variable
$\boldsymbol{Z}$	set of all realisable objective vectors
$\boldsymbol{Z_*}$	Pareto optimal region of objective-space
$\boldsymbol{Z_A}$	approximation set
z_i	ith objective
$\boldsymbol{Z_R}$	region of interest in objective-space

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Outline of the Thesis 2

1.3 Contributions 3

2 Review of EMO **6**

2.1 Introduction 6

2.1.1 Solving Problems 6

2.1.2 Contents of the Review 9

2.2 Multi-Objective Optimisation 10

2.2.1 Fundamental Concepts 10

2.2.2 Advanced Concepts 12

2.2.3 Requirements of a Multi-Objective Optimiser 13

2.3 Evolutionary Algorithms 14

2.3.1 Fundamental Concepts 14

2.3.2 A Generalised Evolutionary Algorithm 16

2.3.3 Advanced EA Concepts 20

2.3.4 Advantages and Disadvantages 25

2.4 Evolutionary Multi-Objective Optimisation 26

2.4.1 Why Use EAs for MO? 26

2.4.2 History of EMO 27

2.4.3 Methods for Obtaining Good Proximity 30

2.4.4 Methods for Obtaining Good Diversity 33

2.4.5 Preference-Based Methods 42

2.5 Summary 46

3 Enhancing MOGA **48**

3.1 Introduction 48

3.2 Experimental Framework 50

3.2.1 Overall Methodology 50

3.2.2 Test Suite 50

3.2.3 Measuring Performance 51

3.2.4 Analysing Performance 55

3.3 Baseline MOEA 58

3.4 Obtaining Good Proximity: The Ranking Strategy 65

3.4.1 Introduction 65

3.4.2 Preliminary Analysis 65

3.4.3 Evaluation 66

3.5 Promoting Diversity: The Sharing Strategy 66

3.5.1 Introduction 66

3.5.2 Epanechnikov Fitness Sharing 68

3.5.3 Rank-Based Methods 70

3.6 Obtaining Good Proximity *and* Good Diversity: an Elitist Strategy 77

3.7 State-of-the-Art MOEA: Enhanced s_s and s_v Mechanisms 81

3.8 Summary 87

4 Many Objectives: An Introduction **89**

4.1 Introduction 89

4.2 Relationships Between Objectives 90

4.2.1 Classification 90

4.2.2 Notation 92

4.3 Conflicting Objectives 93

4.3.1 Definitions of Conflict 93

4.3.2 Effect on EMO 93

4.3.3 Remedial Measures 94

4.4 Harmonious Objectives 96

4.4.1 Definitions of Harmony 96
4.4.2 Effect on EMO 97
4.4.3 Remedial Measures 98
4.5 Independent Objectives 99
4.5.1 Independence in the Context of EMO 99
4.5.2 Effects of Independence on EMO 99
4.6 Existing Methods for Identifying Pair-Wise Relationships 100
4.6.1 Qualitative Methods 100
4.6.2 Quantitative Methods 102
4.7 Summary 103

5 Many Conflicting Objectives 106
5.1 Introduction 106
5.2 Algorithms Considered 108
5.2.1 Introduction 108
5.2.2 Selection Mechanisms 110
5.2.3 Representation and Variation 112
5.3 Inquiry Design 115
5.3.1 Introduction 115
5.3.2 Scalable Optimisation Task 115
5.3.3 Performance Indicators 117
5.3.4 Investigative Framework 118
5.3.5 Introduction to Results 120
5.4 Results: Mutation 122
5.5 Results: Recombination 126
5.6 Analysis 130
5.6.1 Introduction 130
5.6.2 Active Diversity Promotion Mechanisms 131
5.6.3 Inactive Diversity Promotion Mechanisms 135
5.6.4 Comparison Between Mutation and Recombination 138
5.7 Effect of Population Size 139
5.7.1 Introduction 139

5.7.2 Mutation Results 140
5.7.3 Recombination Results 143
5.7.4 Analysis 144
5.8 Summary 147

6 Independence Within Many Objectives 149
6.1 Introduction 149
6.1.1 Problem Decomposition 149
6.1.2 Decomposition-Based Evolutionary Algorithms 150
6.1.3 Chapter Overview 151
6.2 Experimental Methodology 152
6.2.1 Baseline Algorithm 152
6.2.2 Test Functions 152
6.2.3 Performance Metrics 153
6.2.4 Analysis Methods 153
6.2.5 Presentation of Results 155
6.3 The Effect of Independence 156
6.4 Exploiting Independence via Objective-Space Decomposition 160
6.4.1 Overview of the Methodology 160
6.4.2 Population Management 162
6.4.3 Tests for Independence 163
6.4.4 Decision-Space Decomposition: An Aside 164
6.5 Preliminary Results 165
6.5.1 Blum-Kiefer-Rosenblatt D Results 165
6.5.2 Kendall K Results 167
6.5.3 Discussion 168
6.6 Summary 171

7 Conclusions 173
7.1 Evolutionary Multi-Objective Optimisation 173
7.2 Performance Assessment 174
7.3 Evolutionary Many-Objective Optimisation 176

7.4 Future Perspectives 179
7.4.1 Many-Objective Optimisation 179
7.4.2 General EMO Issues 180

A ZDT Benchmark Suite **185**

References **188**

Chapter 1

Introduction

1.1 Motivation

The human species sometimes finds inspiration in the natural world when searching for methods to solve problems. Such methods excite interest because of the aid to understanding embodied in the use of metaphor, and the potential for the methods to solve problems regarded as otherwise difficult because of the limitations of intuition and standard mathematical techniques.

One observation from the natural world is that organisms living in a particular environment have special attributes that enable them to be successful in that environment. Evolutionary biology explains that the organisms have arisen, over time, through a process of adaptation. Thus, can an abstraction of this concept be used as a methodology for finding solutions that perform well in a given problem environment?

This question has been explored by, amongst others, engineers, evolutionary biologists, and theoretical computer scientists for much of the second half of the twentieth century. These research efforts have resulted in the modern-day concept of the evolutionary algorithm, which is now a popular and successful multi-disciplinary subject area.

A solution to a problem will generally need to be assessed against multiple performance objectives. If conflict is experienced between the objectives then multiple solutions, each representing a particular performance trade-off between the objectives, can be considered optimal. Evolutionary algorithms have special utility for multi-objective problems since they can search for a family of individually distinct solutions and do not require objectives to be

aggregated.

Evolutionary computing methodologies that explicitly recognised the multi-objective nature of problems were first introduced in the mid-1980s. Many evolutionary algorithms have since been developed specifically as multi-objective optimisers, and have subsequently been applied to real-world problems. As a measure of the continuing popularity of this branch of evolutionary computing, the first international conference devoted to the subject took place in 2001 in Zürich, Switzerland (Zitzler, Deb, Thiele, Coello and Corne 2001). This has now been established as a successful biennial event, with the second conference held in 2003 in Faro, Portugal (Fonseca, Fleming, Zitzler, Deb and Thiele 2003).

This thesis explores the methodology of evolutionary multi-objective optimisation (EMO). In particular, it seeks to identify which processes within a multi-objective evolutionary algorithm have a significant effect on performance within a particular environment. A largely unanswered question in the research field is how the algorithms behave when **many** objectives must be optimised simultaneously. The thesis aims to promote this area of research, improve the knowledge of optimiser behaviour in these circumstances, and develop an underlying understanding of this behaviour.

1.2 Outline of the Thesis

Chapter 2 provides the necessary background material for the thesis. It discusses the processes that are involved in problem-solving, and introduces theoretical concepts from multi-objective optimisation. The requirements of a multi-objective optimiser are also discussed. Evolutionary algorithms are introduced as a general methodology at both a fundamental and a more advanced level. The various multi-objective evolutionary techniques are subsequently described from the perspective of how they address the required optimisation aims.

Chapter 3 develops an empirical framework for identifying the components of a global algorithm that are responsible for the greater part of the observed performance of that algorithm. The methodology argues the importance of baselining and the use of appropriate statistical confidence tests. The framework is then applied to an exploration of multi-objective evolutionary algorithm components on some popular bi-objective problems.

Chapter 4 argues the requirement for more research into the simultaneous optimisation of many objectives. This need is based on the reality that most theoretical development is

undertaken in the bi-objective domain, whilst applications can feature many more objectives. A platform for new research in this area is introduced in the chapter, based on the consideration of the different relationships that can exist between objectives. Known issues and related research from both the evolutionary and classical multi-objective communities are summarised for the relationships of conflict, harmony, and independence.

Chapter 5 considers the optimisation of many conflicting objectives. Using a slightly modified framework to that described in Chapter 3 to cope with the increase in dimensionality, the behaviour of a class of modern evolutionary optimisers (represented by a popular algorithm and its derivatives) is studied as the number of objectives is varied.

Chapter 6 explores the effect of independence between objectives in a many-objective problem. The effectiveness of *a priori* decomposition approaches are analysed. Subsequently, an adaptive divide-and-conquer strategy (that automatically determines a suitable decomposition) is proposed and tested on a benchmark problem.

Chapter 7 presents conclusions on the work undertaken and offers perspectives on the future of evolutionary multi-objective optimisation.

1.3 Contributions

The main contributions of this thesis are:

- **A contemporary review of EMO from a new perspective.** The review considers isolated components that have been devised to meet the three major aims of multi-objective optimisation (a diverse and accurate representation of the trade-off surface in a defined region of interest). This approach contrasts to the standard holistic review of algorithm brands.

- **A new framework for empirical algorithm analysis.** The framework enables the effect of components and component structures to be analysed, through the use of baselining and the randomisation testing of appropriate performance indicators. The framework is applied to a standard benchmark suite to analyse the performance of various components, including a new method for diversity promotion. This contribution has been published as Purshouse and Fleming (2002).

- **Promotion of the requirement for more research into optimisation problems with many objectives.** A platform for such research is established via consideration of the relationships between objectives. This contribution has been published as Purshouse and Fleming (2003b).

- **New knowledge and understanding of behaviour under many conflicting objectives.** The inquiry into many-objective conflict has identified for the first time that difficulties, in terms of approximation set quality, can arise for larger numbers of conflicting objectives that are not evident for smaller numbers of such objectives. The factors responsible for this behaviour — primarily active diversity promotion in combination with dominance resistance — are also believed to have been identified. This knowledge should assist in the development of new algorithms, especially for real-world applications. This contribution has been published as Purshouse and Fleming (2003c).

- **First demonstration of the benefits of many-objective problem decomposition.** An innovative technique for multi-objective divide-and-conquer is also proposed, using concepts from nonparametric statistics and parallel evolutionary algorithms. The methodology is successfully evaluated using the general framework proposed in the thesis. This contribution has been published as Purshouse and Fleming (2003a).

A number of additional contributions have resulted from the work undertaken during the development of this thesis. For the sake of brevity and clarity, these contributions are not explicitly documented within the monograph itself but are briefly described below:

- **A contemporary review of evolutionary algorithms in control systems engineering.** The survey, commissioned as an International Federation of Automatic Control (IFAC) Professional Brief, considers applications in which the special properties of evolutionary algorithms have been most notably exploited (Fleming and Purshouse 2001b). This contribution has been published as Fleming and Purshouse (2002) and, in an abridged form, as Fleming and Purshouse (2001a).

- **Enhancement of evolutionary algorithm toolbox for MATLAB[1].** Many new components have been developed for the University of Sheffield's *GA Toolbox* (Chipperfield,

[1] MATLAB is a software package for technical computing, developed by The MathWorks, Inc.

Fleming, Pohlheim and Fonseca 1994) during the course of the work documented in this thesis. The availability of these components will facilitate the practice of further component-based design in the spirit of Michalewicz and Fogel (2000).

Chapter 2

Review of Evolutionary Multi-Objective Optimisation

2.1 Introduction

2.1.1 Solving Problems

From planning a walk of the English coast-to-coast path from St Bees to Robin Hood's Bay, to solving Fermat's Last Theorem, to the design of a nuclear fusion power plant, the *problem* is ubiquitous. When a problem exists, a *solution* is naturally sought. Michalewicz and Fogel (2000) offer the following definitions:

> A problem exists when there is a recognized disparity between the present and desired state. Solutions, in turn, are ways of allocating the available resources so as to reduce the disparity between the present and desired state.

Michalewicz and Fogel (2000) also point out that for a problem to exist there must be a corresponding "purpose-driven decision-maker". The *decision-maker* (DM) is some entity that has defined present and desired states and aims to transform the former into the latter. Problem-solving also involves an entity called the *analyst*, who helps the decision-maker to perform actions with the available resources to transform the state and thus solve the problem. The analyst and the decision-maker roles are not restricted to humans, although this is commonly the case, and may each be composed of multiple, conflicting elements. It is also possible that a single entity may encapsulate both roles.

Problem-solving requires consideration of the following questions:

1. Who are the **decision-makers** and stakeholders in the problem?
2. What **objectives** and **preferences** are possessed by each of the DMs?
3. What are the **decision variables** that comprise a solution to the problem?
4. What **constraints** exist for the problem?
5. By what methods can a candidate solution be **assessed** in the context of the problem?
6. What methods will be used to **search** for good candidate solutions?
7. How can a final solution be **validated**?

Each decision-maker has a set of performance criteria or objectives that a candidate solution to the problem will be evaluated against. For example, when choosing a new home, these may include the purchase price, the amount of renovation required, the proximity to good schools, the proximity to good transport links, local crime levels, and the size and orientation of the garden. In almost all cases, it is not possible to obtain the best performance across all the objectives simultaneously. Thus, the DM makes use of preference information for each objective in order to ultimately focus on a single solution. The DM may express a preference hierarchy or *priority* between objectives, for example, "proximity to a good school is more important than proximity to a motorway". The DM may also specify an attainment level or *goal* that is desired to be reached in an objective, such as "the purchase price should be less than £150,000". At a more abstract level of reasoning, the DM may wish to make comparisons at a holistic level between competing candidate solutions, for example "52 Festive Road is far better than 29 Acacia Road". The actual expression of DM preferences must be flexible. A technique should be chosen with which the DM is comfortable. From the above discussion, the utility of fuzzy methods in preference capture is clear.

A candidate solution to a problem involves a number of actions that must be performed. These relate to settings for decision variables in a problem. In the case of buying a house, one major decision variable is the address of the property. Other decision variables may relate to the choice of mortgage, and choice of solicitors. A further variable is the offer price to be made. Decision variables may be of different types: for example, the choice of solicitor

is from a discrete set of alternatives, whereas the offer price is of a more continuous nature. Different settings for decision variables correspond to different candidate solutions to the problem. Candidate solutions can be fairly complex in nature — such as settings for 10,000 Bezier curves for an aircraft airfoil — and may be of a hierarchical nature. The level of influence of each decision variable on the objectives can differ, and the influence of a single decision variable may vary within the context of choices for other variables.

Typically the solution to a problem will have a number of constraints placed upon it. These may be *hard* constraints that must be satisfied, or may be *soft* constraints where the degree of satisfaction may be subject to compromise. Some constraints arise naturally in the problem domain (such as rate limits on an actuator). An example of a hard constraint is the stability of a system. Goals for objectives, as discussed earlier, are often representative of softer constraints. Note that in some cases, such as scheduling applications, it can often prove very difficult to find any solution that will simultaneously meet all the problem constraints.

In order to find a good solution to the problem, some means of assessing the worth of a candidate solution (against the defined set of objectives) is required. This is achieved by using models of the real-world problem developed by the analyst. Models may take many forms, such as analytical, empirical, and simulation, and are a simplification of the actual processes involved. For example, in order to assess the levels of drag on an airfoil design, a finite element computer simulation may be employed. In order to assess job waiting times in a scheduling system, a discrete event simulation may be used. The model should be as accurate as possible. However, there is a trade-off between accuracy and complexity (the latter being measured in terms of development resources required, computational resources required, and amenability to an analytical solution). At the very least, a model should be able to provide accurate discrimination between the performance of two competing solutions.

Once a model has been developed to permit evaluation of candidate solutions, a search technique can be used to find a good solution for the model. If the model is mathematically well-behaved (that is, it is unimodal, continuously differentiable, deterministic, and convex) then standard analytical techniques can be used to find a globally optimal solution. The key difficulty here is that the assumptions required to generate such a model are usually so great that the optimal solution to the model is not a good solution from the perspective of the true problem. Thus, in order to find good solutions to high fidelity models, less

orthodox techniques are required. The class of techniques known as *metaheuristics* (those that work with knowledge of previous solutions to generate new solutions, using processes that are metaphors of observed 'natural' behaviour) are popular and successful methods for solving general models. A key advantage is that domain-specific information can be readily incorporated into the search. The main disadvantage of such methods is that they can (in most cases) only *approximate* the globally optimal solution.

Once the DM has selected a solution from the set recommended by the search method, this must be validated prior to implementation in the real-world. A key stage of validation is *sensitivity analysis*, in which the appropriateness of the solution is tested for small (likely) perturbations of both the solution and model. Thus, the behaviour of the solution in the presence of uncertainty is tested. Uncertainties can arise, for example, because of manufacturing tolerances on solution components or modelling errors. The solution may also be appraised on higher fidelity models or limited real-world test environments (which would have been prohibitive for use at the search stage because of the high complexity costs involved).

Problem-solving is generally a dynamic and iterative process. Experience of early abortive attempts at the solving process aid the development of new approaches. As the search progresses, the extent of the early uncertainty is decreased, and the DM has the opportunity to learn more about the problem environment. This may lead the DM to revise the pre-specified objectives and goals. The model may also be subject to change as more information is gathered, or the actual problem itself changes (for example, if product demand fluctuates). This stresses the need for flexibility in the solving techniques being used.

2.1.2 Contents of the Review

The notion of optimising multiple objectives simultaneously is formally introduced in Section 2.2. The fundamental concepts of solution comparison and optimality are described in Section 2.2.1. The aims of a generic multi-objective optimiser are outlined in Section 2.2.3. The *soft computing* methodology known as *evolutionary computing* (EC) is introduced in Section 2.3, and its special utility for multi-objective optimisation (MO) is explained in Section 2.4.1. A thorough review of the evolutionary methods developed to meet the aims of multi-objective optimisation is provided in Section 2.4.

2.2 Multi-Objective Optimisation

2.2.1 Fundamental Concepts

As discussed in Section 2.1.1, a general problem to be solved will have a defined set of objectives to be optimised. This is formally defined in Definition 2.1. If a consistent DM *line of preference* (Edgeworth 1932) exists for each objective then candidate solution comparisons, and consequently optimality, can be defined using *Pareto* concepts (Coello, Veldhuizen and Lamont 2002). Minimisation is assumed throughout without loss of generality.

Definition 2.1 (Multi-Objective Optimisation Problem) *Minimise the M components of a vector function $\boldsymbol{z}$ with respect to a vector variable $\boldsymbol{x} = (x_1, \ldots, x_n)$ in a universe $\mathcal{U}$, i. e.,*

$$min\ \boldsymbol{z}(\boldsymbol{x}) = [z_1(\boldsymbol{x}), \ldots, z_M(\boldsymbol{x})]$$

An example multi-objective problem domain is shown in Figure 2.1. The left-hand illustration shows a decision-space with two decision vectors. Constraints on these vectors lead to a feasible region, $\mathcal{U}$, as identified in grey. Each decision vector in the decision-space maps uniquely to an objective-vector in objective-space as shown in the right-hand illustration. Note that the inverse mapping may be non-unique. In the example, two objectives are shown. However, in general, the number of objectives may be any positive integer. This is also true of the number of decision variables.

Consider a pair of candidate solutions, $\boldsymbol{u}$ and $\boldsymbol{v}$. Solution $\boldsymbol{u}$ can be considered superior to $\boldsymbol{v}$ if $\boldsymbol{u}$ performs as least as well as $\boldsymbol{v}$ across all the objectives and performs better than $\boldsymbol{v}$ in at least one objective. In the language of multi-objective optimisation, solution $\boldsymbol{u}$ *dominates* solution $\boldsymbol{v}$. This comparison is formalised in Definition 2.2 and shown graphically, for the two objective case, in Figure 2.2.

Definition 2.2 (Pareto Dominance) *Given two candidate solutions $\boldsymbol{u}$ and $\boldsymbol{v}$ from $\mathcal{U}$, vector $\boldsymbol{z}(\boldsymbol{u})$ is said to* dominate *vector $\boldsymbol{z}(\boldsymbol{v})$ (denoted by $\boldsymbol{z}(\boldsymbol{u}) \prec \boldsymbol{z}(\boldsymbol{v})$) if and only if,*

$$\forall i \in \{1, \ldots, M\}, z_i(\boldsymbol{u}) \leq z_i(\boldsymbol{v}) \wedge \exists i \in \{1, \ldots, M\} : z_i(\boldsymbol{u}) < z_i(\boldsymbol{v})$$

If no feasible solution, $\boldsymbol{v}$, exists that dominates solution $\boldsymbol{u}$ then $\boldsymbol{u}$ is classified as a *non-dominated* or *Pareto optimal* solution. This is formally defined in Definition 2.3. There

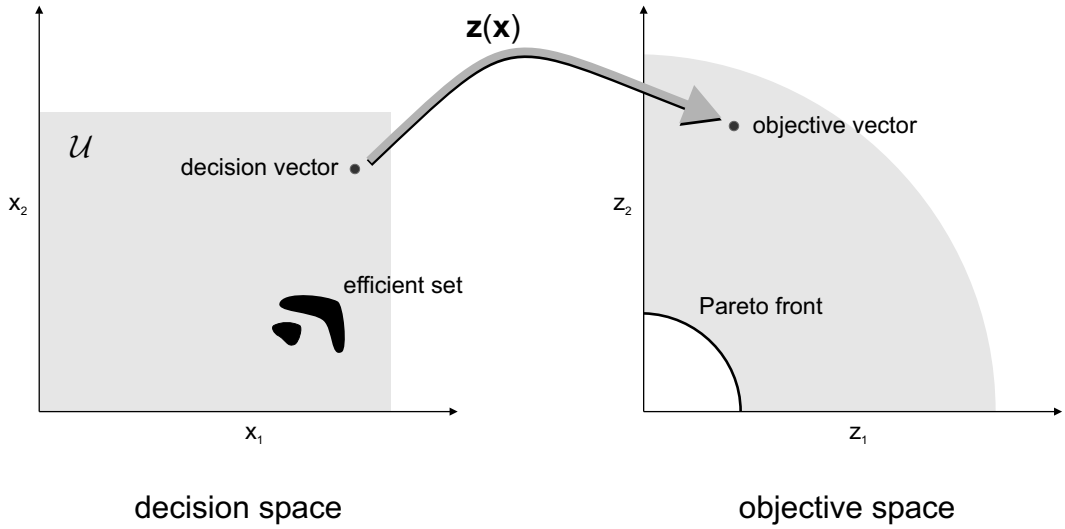


Figure 2.1: The multi-objective problem domain

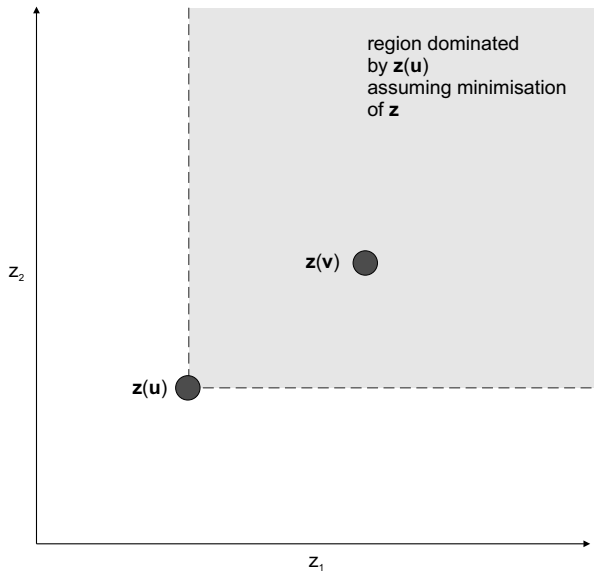


Figure 2.2: Example of Pareto dominance in two objectives

are no superior solutions to the problem than $\boldsymbol{u}$, although there may be other equally good solutions in the absence of further DM preference information.

Definition 2.3 (Pareto Optimality) *The candidate solution* $\boldsymbol{u} \in \mathcal{U}$ *is* Pareto optimal *if and only if,*

$$\neg\exists \boldsymbol{v} \in \mathcal{U} : \boldsymbol{z}(\boldsymbol{v}) \prec \boldsymbol{z}(\boldsymbol{u})$$

The collection of all candidate solutions that meet Definition 2.3 are known as the *Pareto optimal*, or *efficient*, *set*, $\boldsymbol{X}_*$. The corresponding objective vectors are described as the *Pareto front* or *trade-off surface*. An example of the relationship is shown in Figure 2.1. The terminology *non-dominated* is sometimes used when referring to either the Pareto optimal set or the Pareto front (although strictly this only applies to the latter vectors). The above definitions are all provided in their global sense. Local definitions (such as *locally non-dominated*) can be obtained by substituting universe $\mathcal{U}$ for a local space $\mathcal{L} \subseteq \mathcal{U}$.

2.2.2 Advanced Concepts

An approximate version of the Pareto dominance concept, known as ϵ-*Pareto dominance* was established in the operations research (OR) community during the mid-1980s (refer to Helbig and Pateva (1994) for a review). Under ϵ-dominance, the conditions required for one solution to dominate another are relaxed. Various forms of ϵ-dominance have been proposed, of which the *multiplicative* form is provided in Definition 2.4. From a practical perspective, the value of ϵ can be regarded as the maximum change in the value of an objective that is regarded as unimportant by the decision-maker (the so-called *indifference threshold*).

Definition 2.4 (ϵ-Pareto Dominance) *If* $\boldsymbol{z}(\mathcal{L}) \subseteq \mathbb{R}^{+M}$ *then, given two candidate solutions* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *in* $\mathcal{L}$*, vector* $\boldsymbol{z}(\boldsymbol{u})$ *is said to* ϵ-dominate *vector* $\boldsymbol{z}(\boldsymbol{v})$ *(denoted by* $\boldsymbol{z}(\boldsymbol{u}) \prec_\epsilon \boldsymbol{z}(\boldsymbol{v})$*) if and only if,*

$$\forall i \in \{1, \ldots, M\}, (1 - \epsilon_i) z_i(\boldsymbol{u}) \leq z_i(\boldsymbol{v}),$$

where ϵ_i *is the DM indifference threshold on the* i*th objective.*

An illustration of multiplicative ϵ-dominance in the bi-objective case is shown in Figure 2.3. The shaded area indicates the region of objective-space that is ϵ-dominated by solution $\boldsymbol{u}$. This region is composed of the area that would normally be dominated by $\boldsymbol{u}$ plus

areas that would otherwise be non-dominated with respect to $\boldsymbol{u}$ or would in fact dominate $\boldsymbol{u}$. Thus, solution $\boldsymbol{v}$ that is non-dominated with respect to $\boldsymbol{u}$ (according to the pure definition of dominance) is also ϵ-dominated by $\boldsymbol{u}$.

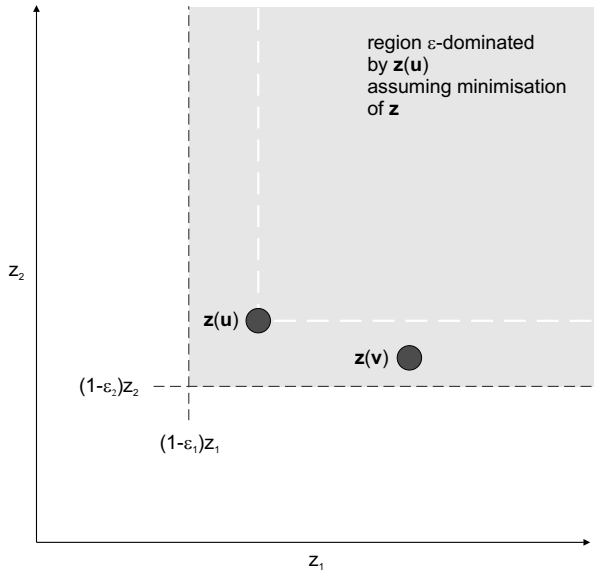


Figure 2.3: Example of multiplicative ϵ-Pareto dominance in two objectives

A set of solutions that together ϵ-dominate the universe $\mathcal{U}$ is known as an *ϵ-approximate Pareto optimal set*. Such a set is generally non-unique. This concept was refined further by Laumanns, Thiele, Deb and Zitzler (2002) to the case where all the solutions within the set are also globally Pareto optimal. This set is then termed *ϵ-Pareto Optimal* and is formalised in Definition 2.5 below.

Definition 2.5 (ϵ-Pareto Optimal Set) *The set of candidate solutions $\boldsymbol{X}_*^{\epsilon} \in \mathcal{L}$, where $\boldsymbol{z}(\mathcal{L}) \subseteq \mathbb{R}^{+M}$, is an* ϵ-Pareto optimal set *if and only if,*

$$\forall \boldsymbol{v} \in \mathcal{L}, \exists \boldsymbol{u} \in U_{\epsilon} : \boldsymbol{z}(\boldsymbol{u}) \prec_{\epsilon} \boldsymbol{z}(\boldsymbol{v}) \wedge \boldsymbol{u} \in \boldsymbol{X}_*$$

2.2.3 Requirements of a Multi-Objective Optimiser

The globally optimal trade-off surface of a multi-objective problem can contain a potentially infinite number of Pareto optimal solutions. The task of a multi-objective optimiser is to provide an accurate and useful representation of the trade-off surface to the decision-maker.

The set of solutions generated by the optimiser is known as an *approximation set* (Zitzler, Thiele, Laumanns, Fonseca and Grunert da Fonseca 2003). Three aspects of solution set quality can be considered. These are listed below, and shown graphically in Figure 2.4.

- **Proximity.** The approximation set should contain solutions whose corresponding objective vectors are close to the true Pareto front.
- **Diversity.** The approximation set should contain a good distribution of solutions, in terms of both extent and uniformity. Good diversity is commonly of interest in objective-space, but may also be required in decision-space. In objective-space, the approximation set should extend across the entire range of the true Pareto front with a parametrically uniform distribution across the surface.
- **Pertinency.** The approximation set should only contain solutions in the decision-maker region of interest (ROI). In practice, and especially as the number of objectives increases, the DM is interested only in a sub-region of objective-space. Thus, there is little benefit in representing trade-off regions that lie outside the ROI. Focusing on pertinent areas of the search space helps to improve optimiser efficiency and reduces unnecessary information that the DM would otherwise have to consider.

2.3 Evolutionary Algorithms

2.3.1 Fundamental Concepts

Evolutionary computation is a search discipline based on the evolutionary biology concepts of *natural selection* (Darwin 1859) and *population genetics* (Fisher 1930). The term *evolutionary algorithm* (EA) has no rigorous definition, but can generally be used to describe any population-based, stochastic, direct search method. Many such algorithms exist.

Contemporary evolutionary algorithms have their origins in two originally independent and highly influential evolutionary models: the German *evolution strategy* (ES) (Rechenberg 1973) and the American *genetic algorithm* (GA) (Holland 1975). The distinctions between the fields have weakened considerably over the last thirty years, with modern researchers encouraged to think in broader EA terms (Michalewicz and Fogel 2000). Note that other,

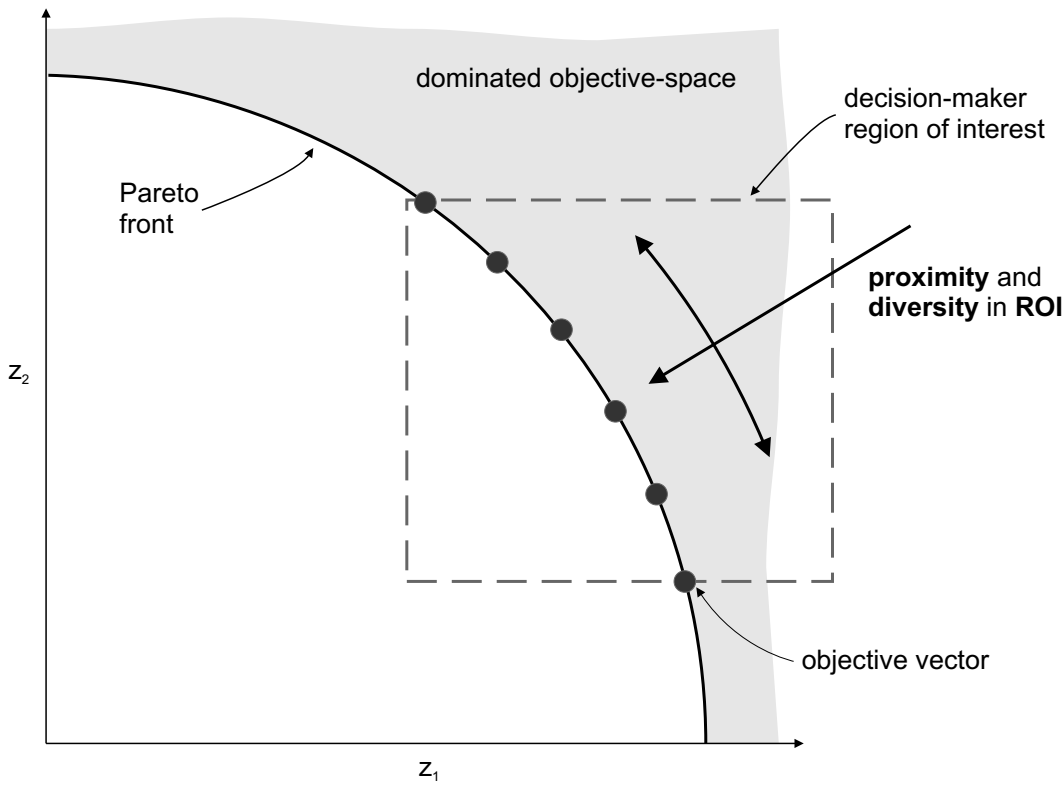


Figure 2.4: The ideal solution to a multi-objective optimisation problem

similar, early evolutionary models were also proposed, but these have not enjoyed the same level of influence as GAs and ES.

The fundamental idea in an evolutionary search is to iteratively apply *variation* (v) and *selection* (s) processes to a *population* of candidate solutions until some termination criterion is satisfied. This core concept is described in Equation 2.1, in which $P[t]$ is the population at time t (Michalewicz and Fogel 2000).

$$P[t+1] = s(v(P[t])) \tag{2.1}$$

Probabilistic changes are made to solutions in the population to yield a set of new candidate solutions. Solutions are then selected for inclusion in the population at the next iteration (or *generation*), $P[t+1]$, based on their ability to solve the problem at hand. Variation is then applied to the new population to generate yet another set of new solutions, and so on. The aim of the EA is thus to progressively develop better solutions to the problem by making modifications to previous solutions that exhibited good performance within their peer group.

2.3.2 A Generalised Evolutionary Algorithm

Expanding slightly on the fundamental evolutionary search mechanism given in Equation 2.1, the general form of a modern EA can be captured as shown in Equation 2.2. The central difference is that the overall selection operation has been split into two processes, *selection-for-variation* (s_v) and *selection-for-survival* (s_s), which are described in more detail below.

$$P[t+1] = s_s(v(s_v(P[t])), P[t]) \tag{2.2}$$

Selection-for-Variation

At the selection-for-variation stage, solutions from $P[t]$ are chosen for inclusion in the so-called *mating pool*, to which variation operators are applied to create new solutions. The performance of every solution is evaluated for the problem at hand, by whatever means. This raw evaluation is then transformed into a scalar *fitness* value that is an overall measure of solution performance. Larger fitness values correspond to better solutions. All EA selection operators use fitness value as the selection discriminator. Fitness ultimately represents the expected number of times that a solution will be selected. The mapping between the two

quantities is often implicit to the selection procedure used.

Several different selection operators have been proposed in the literature, including *tournament selection* (Brindle 1981) and *stochastic universal sampling* (Baker 1987). The reader is referred to Goldberg and Deb (1991) for a review and analysis. The ideal selection mechanism has a low computational complexity, is parallelisable, has minimal spread (defined as the maximum sample deviation from the expected number of selections of an individual), and minimal bias (defined as the difference between the sampling probability of selection for an individual and the true expected value). Such an ideal mechanism is difficult to obtain in practice.

All selection operators have a property known as *selective pressure*. This is the ratio between the number of expected selections of the best performing solution and the mean performing solution. In proportionate selection schemes this value is explicitly defined by the user, whilst it is implicit to tournament selection mechanisms.

Variation

The variation operators use the genetic material of the solutions in the mating pool to create new candidate solutions. The operators work on a representation of the solutions (known as the *genotype*) rather than the actual solutions themselves (known as the *phenotype*). Thus, coding and decoding processes are required between genotype and phenotype. In genetic algorithms, the genotype is traditionally a binary string, whereas in evolution strategies it is a concatenation of real numbers (direct operation on the phenotype). However, there is no restriction as to representation, providing that variation operators can be devised to handle it. Contemporary wisdom is to use whatever representation seems most appropriate for the task at hand (Michalewicz and Fogel 2000). For instance, genetic algorithms solving real-parameter function optimisation problems generally use real-parameter variation operators (Herrera, Lozano and Verdegay 1998). Block diagram representations, and other structures such as trees, are also popular (Koza 1992). In one example, Gray, Murray-Smith, Li, Sharman and Weinbrunner (1998) describe a potential methodology for the structural identification of a system using a block diagram approach.

Variation operators can be classified according to whether one solution or multiple solutions (known as *parents*) are used to create the new solutions (known as *children* or *offspring*).

Single parent operators are commonly described as *mutations*. The classic example for the original genetic algorithm is known as *bit-flipping mutation*, in which each element, or *locus*, of the genotype is probabilistically tested for possible mutation. Consider the parent genotype `1111`: if only the third locus of the genotype is to be mutated then the resulting child genotype is `1101`. In the classic evolution strategies approach, mutation is achieved by means of a Gaussian probability density function (p.d.f.) centred on the parent decision variable. The p.d.f. is sampled to obtain a child value on the real number range.

Multi-parent variation operators are commonly known as *recombination* or *crossover* operators. This can be illustrated by the original genetic algorithm technique known as *single-point binary crossover*, in which two parents are partitioned at randomly determined equivalent positions and then the genetic material is swapped between the two genotypes in one of the segments to form two children. Consider an example with parents `1111` and `0000`: if the crossover site is chosen as the third locus, then the children are `1100` and `0011`. Recombination was absent from early ES approaches but special versions have now been developed for this domain (Whitley 2001).

Every variation operator has an associated probability of application. Good values for this probability depend on the problem to be solved and the interaction with selection mechanisms. In genetic algorithms, general heuristics exist for both crossover and mutation (Goldberg 1989). The probability of crossover between a parent pair is usually quite high (values such as 0.7 or 0.8 are common). Conversely, the probability of applying mutation to a particular locus on the genotype is usually quite low (often set to be equivalent to an expectation of one mutation per solution). These values are based on the assumption that both operators are to be applied at each generation. In evolution strategies, a *self-adaptive* approach is usually employed, in which the variation probability is itself subject to evolution (Whitley 2001).

Selection-for-Survival

Following the application of the variation operators, two sets of solutions exist: the current population, $P[t]$, and the child population, $v(s_v(P[t]))$. Since the size of an EA population over the generations is typically static, more solutions now exist than can be retained. Thus, a selection-for-survival stage is required in order to determine the new population, $P[t+1]$. In ES, this concept is embodied in the so-called (μ, λ) and $(\mu + \lambda)$ schemes, where μ is the

current population and λ is the child population. In the (μ, λ) method, only child solutions can survive to $P[t+1]$, whereas in the $(\mu + \lambda)$ approach the child and $P[t]$ solutions compete for inclusion in the subsequent population. In the genetic algorithm field the concept of a *generational gap* was conceived, in which sufficient children were generated to replace a certain percentage of $P[t]$. Selection-for-survival is often a deterministic process in which the best solutions are selected from a fitness-based hierarchy. However, stochastic schemes are also possible. s_s schemes that force the retention of high-performance members of $P[t]$ are described as *elitist*.

Exploration versus Exploitation

One of the fundamental issues in any search algorithm is the trade-off between *exploration* of the undiscovered regions of search space and detailed *exploitation* of promising areas already identified. Good performance can be heavily dependent on an appropriate choice of *e-e trade-off*. In an EA, this choice is controlled by means of EA operators and associated free parameters. Classically, in the EA community, variation is seen as an explorative operator, whilst selection is viewed as exploitative. This links cleanly with the evolutionary biologist's view of natural selection as being a process that reduces the variation within a population, and the effect of genetic mutations to increase the aforementioned variation. However, in the context of a general EA search, this is an oversimplification. In particular, the e-e trade-off can often be heavily varied entirely within the EA variation operator, by means of a single control parameter: *blend crossover* is one such example (Eshelman and Schaffer 1993). The selective pressure, described earlier, is another means of control.

Obtaining the correct e-e balance for the task at hand is the black art of EA design (although this area is becoming increasingly understood). Essentially, it requires the exploitation of *a priori* knowledge about the problem landscape under consideration. Note that the required e-e setting may change as the optimisation progresses (for example, expected variation perturbations should probably be reduced as the algorithm converges on a final solution). Some automatic schemes for controlling the e-e trade-off within the variation operator have been suggested: examples are described by Bäck (1996) and Deb and Agrawal (1995).

2.3.3 Advanced EA Concepts

Many variants on the original genetic algorithms and evolution strategies have been proposed in the literature. Some of the key areas of extension are briefly reviewed in this section.

Probabilistic Models

There exists a class of search techniques, known as *estimation of distribution algorithms* (EDAs), that attempt to build a probabilistic model of current high performance areas of the search space and generate new candidate solutions from this model. These solutions are then used to update the model, depending on their performance (Pelikan, Goldberg and Lobo 1999).

EDAs are very closely related to evolutionary algorithms. The main difference is that the variation operators of the EA are replaced by probabilistic sampling from a model. Probabilistic modelling and sampling can create a substantial overhead in the algorithm (significantly larger than that of standard variation operators), but offers the advantage that relationships between decision variables are explicitly considered and exploited to provide a more efficient search in terms of required solution evaluations.

EDAs can be classified according to the complexity of the models they are capable of building. Higher fidelity models can effectively solve a wide range of problems, but the costs of model building and sampling can be considerable. A comprehensive review of EDAs can be found in Pelikan *et al.* (1999). Simple EDAs, such as *population-based incremental learning* (PBIL) (Baluja 1994), assume that there are no interactions between variables. Such approaches will work well so long as the epistasis in the problem is limited. The next step in complexity is to consider pair-wise relationships between decision variables, as in Pelikan and Mühlenbein's (1999) *bivariate marginal distribution algorithm* (BMDA). Various pair-wise topologies are possible, of which BMDA's *forest* is the most flexible. In order to overcome the limitations of the pair-wise interaction algorithms, research has continued into more powerful and complicated algorithms that are capable of handling general multivariate interactions. These require computationally expensive methods to build the associated models. The methods also tend to be greedy, thus reducing the probability of generating a globally optimal model. The *Bayesian optimisation algorithm* (BOA) models promising solutions by learning Bayesian networks (Pelikan, Goldberg and Cantú-Paz 1998).

Most EDAs require that the decision variables be represented by fixed length strings defined over a finite alphabet, which can seriously hamper the flexibility of the algorithms as a general problem-solving tool. Some work has, however, been undertaken into the incorporation of real-parameter representations (Bosman and Thierens 1999). Research has also considered representing the high-performance solution region by more than one probability model, through the use of solution clustering and *mixture distributions* (Bosman and Thierens 2000, Pelikan and Goldberg 2000).

Population Topology

The topology of an EA population can be any of three basic forms: *global*, *island*, or *diffusion* (Chipperfield and Fleming 1995). In the global model, all the candidate solutions reside within a single population, within which every solution competes with every other solution for selection. Also, for multi-parent variation schemes, any solution can usually participate with any other. This scheme is the most popular in practice, largely because it is simple and can produce acceptable results.

In the island model, separate sub-populations of solutions (known as *demes*) exist which evolve largely independently. *Migration* is possible, in which some solutions on one island are transferred to another. An example island model is depicted in Figure 2.5, in which four sub-populations are shown in a ring topology. Migration is only possible between neighbouring islands on the ring. Note that other topologies are also (equally) acceptable.

Island models require extra design choices to be made, including the number and arrangement of the sub-populations, the standard EA settings for the evolution of each sub-population, and also the choice and frequency of migrants. However, the island concept offers coarse-grained parallelisation through the assignment of different demes to different processors, and there is some empirical evidence to suggest that island schemes can produce good results in fewer overall solution evaluations than an equivalent global model (see, for example, Cohoon, Martin and Richards (1991)).

A finer-grained parallelisation can be obtained through use of the diffusion model. In this approach, single solutions or small groups are assigned to nodes in a grid (with each node potentially hosted on a separate processor). Various grid geometries are possible, such as ring, torus, and hypercube. EA operations are then performed in local neighbourhoods

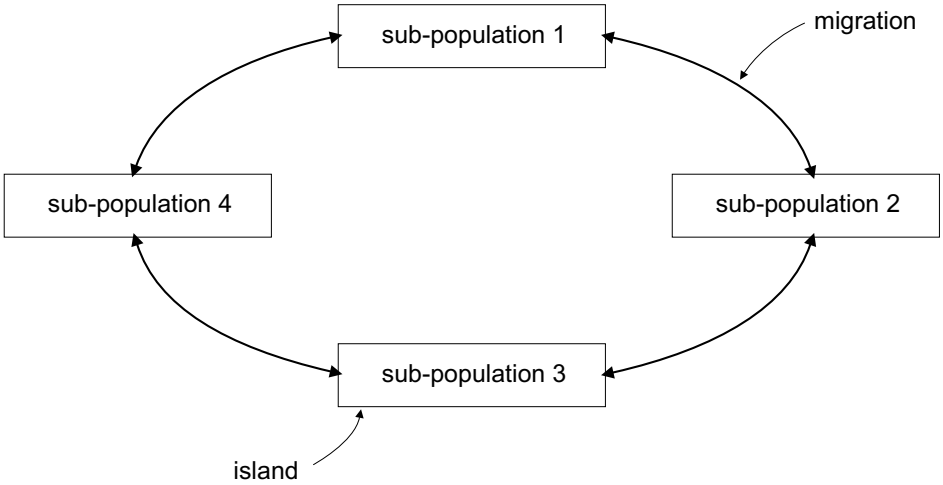


Figure 2.5: The island model: four demes arranged in a ring topology

within the overall grid. It has been claimed that the diffusion model is more efficient (in terms of required solution evaluations) than the island model (Gordon, Whitley and Böhm 1992). However, the diffusion topology is the least frequently used, perhaps because of the limited availability of fine-grained parallel computing systems.

An example diffusion model is shown in Figure 2.6. Twenty-five population nodes are shown, each of which will be assumed to contain a single candidate solution. A simple rectangular grid configuration is shown. This model could be easily extendible to a torus by allowing nodes on one edge to communicate directly with nodes on the opposite edge. In the example a solution has been chosen for recombination, as indicated by the black circle. A new solution is then created for consideration at this node by performing recombination with one of the neighbouring nodes on the grid (these are indicated in Figure 2.6 by the grey circles). The neighbourhood connections are indicated by the heavier lines on the grid.

Genetic Drift Countermeasures

If multiple solutions in the current EA population have identical fitness values then, over generations, the population will tend to converge to one of these solutions. This behaviour is also known to occur in biological population genetics and is termed *genetic drift* (De Jong 1975). The drift occurs because of cumulative sampling errors on expected selection rates due to the use of a finite population.

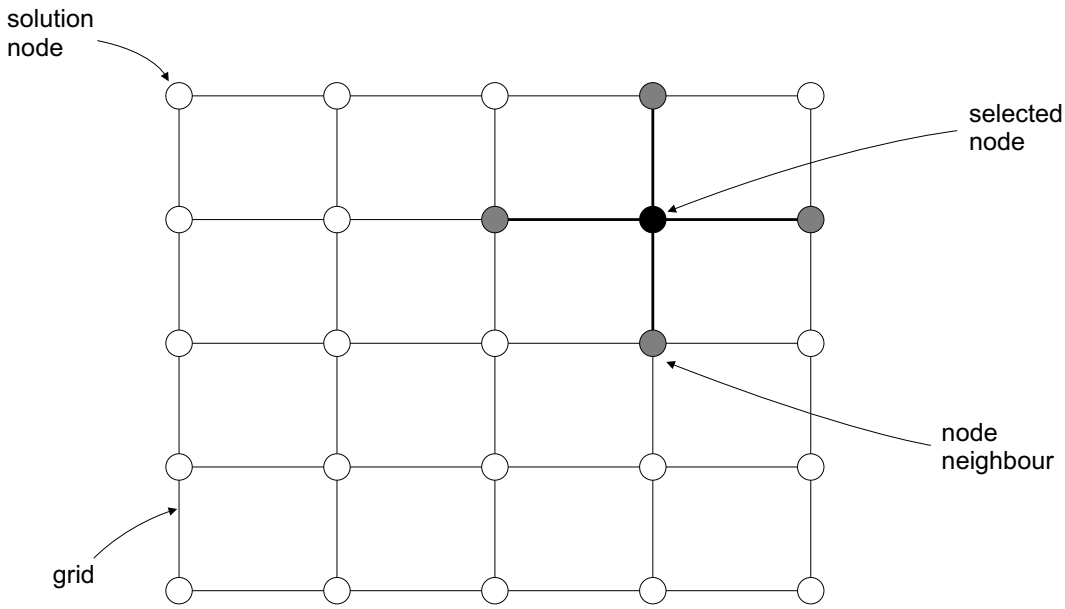


Figure 2.6: The diffusion model: 25 solutions arranged in a simple rectangular grid

Explicit processes are required in an EA to prevent the problem of genetic drift and to enhance diversity in the population. The methods are aimed at so-called *niche formation*: the maintenance of multiple, distinct sub-populations within the context of the global population. The two classic niching methods are *crowding* (De Jong 1975) and *sharing* (Goldberg and Richardson 1987). In the former scheme, the solution from a sample of $P[t]$ that is most similar to a child solution is replaced by the child in $P[t+1]$ at the selection-for-survival stage of the EA. In the sharing method, solution fitnesses are degraded if they are within a certain distance of other solutions. The notions of similarity and distance are usually problem-dependent.

Goldberg and Richardson's (1987) sharing concept makes use of the power law sharing function shown in Equation 2.3, where σ_{share} is a parameter to control the extent of sharing, d is a distance metric, and α shapes the function.

$$Sh(d) = \begin{cases} 1 - \left(\frac{d}{\sigma_{\text{share}}}\right)^{\alpha} & \text{if } d < \sigma_{\text{share}}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

The fitness of the ith solution is modified by all the other solutions (including itself) that reside within the neighbourhood defined by σ_{share} as shown in Equation 2.4, where $\bar{f}_i$ is the

modified fitness, f_i is the original fitness, N is the number of solutions in the population, and $d_{i,j}$ is the distance between solution i and solution j.

$$\bar{f}_i = \frac{f_i}{\sum_{j=1}^{N} Sh(d_{i,j})} \tag{2.4}$$

A suitable setting for σ_{share} is critical to the success of the method. A method to assist in the choice of the parameter in the context of single-objective optimisation (sharing in decision-space) has been proposed by Deb and Goldberg (1989). Approaches have also been proposed specifically for multi-objective optimisation (where sharing is normally computed in objective-space). These methods are discussed further in Section 2.4.4 but it should be noted that most of the schemes can equally be applied to decision-space.

Constraint Handling Methods

Most real-world problems involve constraints on the solution. If a solution can meet all constraints it is described as *feasible*. If any of the constraints are breached it is described as *infeasible*. In applications such as scheduling, the identification of even a single feasible solution can be a challenging task. Several methods have been developed to handle constraints in EAs, as comprehensively reviewed by Coello (2002). A summary of the main approaches is detailed below:

- **Coding.** The most efficient approach is to design the representation and variation operators in such a way that infeasible solutions cannot be generated. In practice, this can be quite a difficult task to achieve. *Decoders* are sometimes used to translate an indirect solution representation into the solution itself. The genotype is often 'read' sequentially, with the interpretation of later parts of the genotype depending on the decoding of previous elements. This approach can be of great help in ensuring feasibility but it can prove very difficult to analyse the effect of variation operators in these conditions. An example of decoder use is described by Shaw and Fleming (1997).

- **Penalties.** The use of penalty functions has proved popular in the EA community (Coello 2002). In this approach, the fitness values of infeasible solutions are reduced, usually as some function of their 'distance' to being feasible. The use of penalties is rather inelegant, but has proved effective in practice.

- **Repairs.** Specialist repair algorithms have been suggested to convert infeasible solutions into related feasible solutions. Typically, the fitness of such a solution is reduced in relation to the 'cost' of the associated repair.

- **Multi-objective formulation.** Rather than modifying an infeasible solution or reducing its overall fitness, it is also possible to draw a distinction between the performance of a solution on the objectives and its performance on the constraints. In a multi-objective formulation, a constraint can thus become a further axis of performance. In these circumstances Pareto-type solutions to the problem can be sought, using comparison operators such as *preferability* (Fonseca and Fleming 1998a) and *constrained-domination* (Deb, Pratap, Agarwal and Meyarivan 2002).

Competent Evolutionary Algorithms

In order to use an EA, in addition to domain-specific considerations such as representation and variation, a set of design choices must be made with respect to population size, selective pressure, and variation operator application probabilities. It is a non-trivial task to determine good settings for these parameters, which requires substantial practical experience of EA implementations. However, based on theoretical results from the binary coded genetic algorithm field, Lobo and Goldberg (2001) and Reed, Minsker and Goldberg (2000) have devised design methodologies for so-called *competent* evolutionary algorithms which help the user to determine appropriate parameter settings. This is an area that demands urgent further work, but will always carry a high degree of complexity because of the intrinsic flexibility of the EA methodology for general problem-solving. In particular, guidelines need to be as representation-free as possible.

2.3.4 Advantages and Disadvantages

The fundamental advantage of the EA methodology over alternative schemes is the astonishing flexibility that the method permits. Since the EA is a direct search method, there are no theoretical restrictions on the evaluation function: simulation and human response to solutions are both possible. In addition, there are no restrictions on the representation of a solution. Any data type is permitted for which suitable variation operators can be devised. Variation operators can, and should, be domain-specific. Since the EA is a population-based

and stochastic methodology, it offers increased robustness to multimodality and randomness over conventional methods.

The flexibility of the EA is a two-sided coin, however. The design of an EA requires a number of choices to be made. Poor choices, for example the wrong probability of mutation for the landscape at hand in the context of the selective pressure, will lead to poor performance. Furthermore, the 'Swiss army knife' nature of the EA means that it will tend not to be competitive with domain-specific tools (Michalewicz and Fogel 2000). If a good method exists to solve the task at hand without the involvement of any harmful assumptions or simplifications then that method should certainly be used. EAs rely on an iterative search over a population of solutions. Thus, if solution evaluation is time-consuming then the resulting optimisation process can be a very computationally intensive business. EA design in such circumstances requires special care. The requirement for a completed search within a reasonable time-frame has motivated research into parallel evolutionary algorithms (such as the island and diffusion models described earlier) and *metamodelling* (the approximation of solution performance using faster, lower accuracy models) (Bull 1999).

2.4 Evolutionary Multi-Objective Optimisation

Evolutionary algorithms are a popular tool for multi-objective optimisation. The mechanisms that underpin their special utility are described in Section 2.4.1. A brief history of the *evolutionary multi-objective optimisation* (EMO) field is provided in Section 2.4.2. As discussed previously in Section 2.2, a multi-objective optimiser is required to produce an approximation set that is close to globally Pareto optimal and that contains a rich distribution of solutions in regions of interest to the decision-maker. Distinct EA components have been developed to address each aspect of approximation set quality. In Section 2.4.3, methods for obtaining good proximity to the global Pareto front are reviewed. In Section 2.4.4, methods for obtaining a suitable distribution are discussed. Techniques for developing and focusing on the DM ROI are outlined in Section 2.4.5.

2.4.1 Why Use EAs for MO?

In addition to the general benefits of using an EA as a search tool, outlined in Section 2.3.4, the EA concept is particularly suitable for multi-objective tasks. The population-based nature

of the algorithm permits objectives to be treated distinctly through the notion of Pareto dominance and permits a *family* of trade-off solutions to be produced in a single execution of the algorithm. The fundamental benefit of this latter factor over multiple-start strategies is the potential for a cooperative search for ultimately different solutions, thus saving on the total number of solution evaluations required. As discussed in Section 2.4.4, given the implementation of a suitable process, a multi-objective evolutionary algorithm (MOEA) can work toward a good solution distribution. In multi-start strategies that rely on particular parameter settings to provide direction toward a particular area of the Pareto front, there is generally no guarantee that a good distribution of parameter settings will ultimately lead to a good distribution of solutions on the trade-off surface.

The key EMO benefit of not requiring objectives to be aggregated in some way to form an overall cost function cannot be overemphasised. It is generally very difficult to aggregate objectives in a manner that precisely captures the DM preferences. Also, the required normalisation of non-commensurable objectives can be far from straightforward. The EMO Pareto-based approach offers flexibility and information-richness with regard to solution performance discrimination, and assists the DM in *learning* about the problem as the search progresses.

2.4.2 History of EMO

The first evolutionary algorithms that were purposefully designed to obtain an approximation set were proposed in the mid-1980s (Schaffer 1985, Fourman 1985, Kursawe 1991). In these schemes, a proportion of the population was selected according to each individual objective. The main difficulty with this approach is that it often creates a phenomenon known as *speciation*, in which solutions arise in the population that are particularly strong in a single objective and particularly poor in others. Thus, important compromise solutions remain undiscovered, since the recombination of solutions from different extreme regions of the trade-off surface cannot usually be assumed to generate an 'intermediate' compromise.

In the *weighted-sum* approach to MO, performance is captured in a single objective, calculated as a weighted-sum of individual performance in each of the original individual objectives. The well-known drawbacks of this approach are the difficulty in setting values for the weights, and the necessary condition for convexity of the trade-off surface that is

required to obtain all Pareto optimal solutions (Censor 1977). Thus, no combination of weights exists that can generate solutions in non-convex regions of the trade-off surface, as shown geometrically by Fleming and Pashkevich (1985). However, EMO schemes based on weighted-sums have also been proposed. Haleja and Lin (1992) included the weight vector in the solution genotype and allowed multiple weight combinations to be propagated through the population during evolution. Jin, Okabe and Sendhoff (2001a) varied the weight vector over the evolution, and have also provided theoretical justification for the method (Jin, Okabe and Sendhoff 2001b, Okabe, Jin and Sendhoff 2002).

Unlike these early attempts, the majority of modern EMO approaches are based on the concept of Pareto dominance given in Definition 2.2 (Coello *et al.* 2002). The use of Pareto dominance as a basis for solution comparison in EAs was first suggested by Goldberg (1989), together with the use of a niching technique to encourage solution distribution across the trade-off surface. In the early-1990s, three much-cited techniques emerged based on Goldberg's ideas: Fonseca and Fleming's (1993) *multi-objective genetic algorithm* (MOGA), Horn and Nafpliotis's (1993) *niched Pareto genetic algorithm* (NPGA) and Srinivas and Deb's (1994) *non-dominated sorting genetic algorithm* (NSGA), although early less well-known implementations by Ritzel (1992) and Cieniawski (1993) have also been reported (Horn and Nafpliotis 1993, Fonseca and Fleming 1995b). The techniques differ slightly in the way in which fitness is derived from Pareto comparisons of solutions. MOGA, NPGA, and NSGA all use fitness sharing for diversity promotion (Goldberg and Richardson 1987).

In the late-1990s, new methods were proposed to improve on the performance of the earlier Pareto-based algorithms. The innovations were usually evaluated on bi-objective test problems. Research efforts have focused particularly on the selection-for-survival aspect of an MOEA, with new methods for preserving and using identified (relatively) good solutions. Techniques for population density estimation and its use in diversity-promotion schemes have also been the focus of contemporary research. These ideas have been implemented in algorithms such as Zitzler and Thiele's (1998) *strength Pareto evolutionary algorithm* (SPEA), Corne, Knowles and Oates's (2000) *Pareto envelope-based selection algorithm* (PESA), and Deb, Pratap, Agarwal and Meyarivan's (2002) *elitist non-dominated sorting genetic algorithm* (NSGA-II). A schematic that summarises the key developments in EMO research is provided in Figure 2.7.

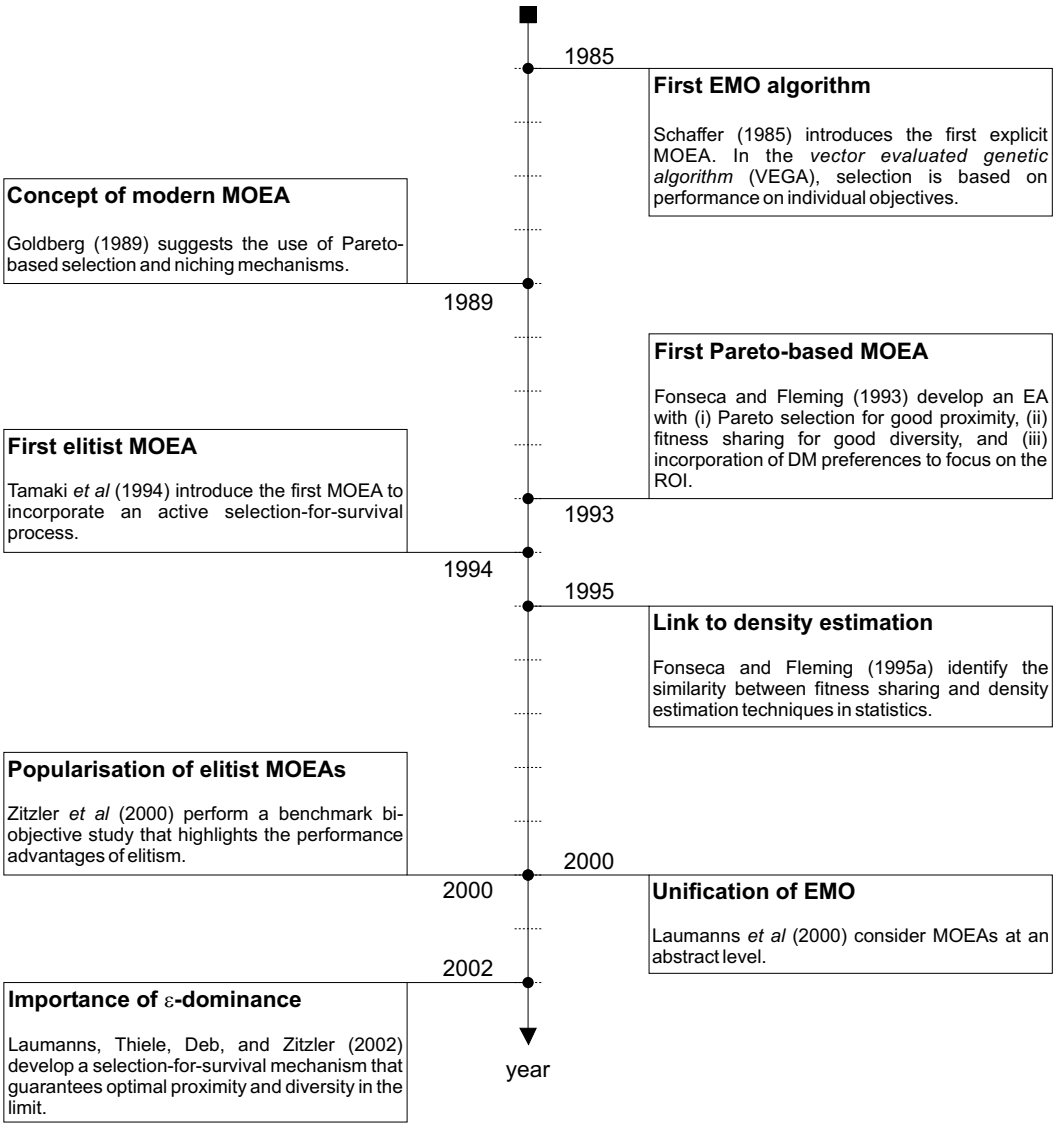


Figure 2.7: Key developments in EMO history

2.4.3 Methods for Obtaining Good Proximity

The selection methods that discriminate in favour of locally non-dominated solutions over dominated counterparts are the corner-stone for evolving the population toward the globally optimal surface. The methodologies are based on a partial ordering, or ranking, of the population. Following the terminology of Zitzler (2002), these schemes make use of the following information for each solution that can be drawn from the current population:

- **Dominance rank.** The number of solutions in the population that dominate the solution under consideration.
- **Dominance count.** The number of solutions in the population that are dominated by the solution under consideration.
- **Dominance depth.** The rank of the solution in the *non-dominated sorting* of the population.

Non-dominated sorting was the original Pareto-based EA approach proposed by Goldberg (1989). The locally non-dominated solution in the population are identified, assigned rank 0, and are removed temporarily. In the remaining population, the new locally non-dominated solutions are identified, assigned rank 1, and are removed. This process is continued until all solutions have been assigned a rank.

Fonseca and Fleming (1993) achieved a ranking with greater resolution simply through use of the dominance rank, in a scheme called *Pareto-based ranking* which was originally implemented in the MOGA. A further example of its use can be found in Thierens and Bosman's (2001b) *multi-objective mixture-based iterated density estimation evolutionary algorithm* ($\mathbb{M}$ID$\mathbb{E}$A). Note that Khan, Goldberg and Pelikan (2002) also implemented a probabilistic model-building scheme based on non-dominated sorting. A still greater degree of granularity is achieved in Zitzler and Thiele's (1998) strength-based approach, originally implemented in SPEA, which makes use of both dominance rank and dominance count. A modified version has also been proposed in the SPEA2 (Zitzler, Laumanns and Thiele 2001).

In another variation on the dominance rank theme, Lu and Yen (2002) proposed an *automatic accumulated ranking strategy* in which the rank of a solution is the sum of the ranks of the solutions that dominate it. A sample-based approximation to a Pareto-based ranking of the entire population was proposed by Horn and Nafpliotis (1993).

Proximity Promotion in Selection-for-Variation

Fonseca and Fleming (1993) and Srinivas and Deb (1994) map rank values to fitness values, via some (typically linear) transformation, and then apply proportional selection methods in order to form the mating pool: *stochastic universal sampling* and *stochastic remainder sampling* respectively (Baker 1987). Binary tournament selection-for-variation operates directly on the strength-based fitness measure in the SPEA family approaches (Zitzler and Thiele 1998, Zitzler, Laumanns and Thiele 2001) and on the dominance depth in Deb, Pratap, Agarwal and Meyarivan's (2002) NSGA-II.

In some methodologies, such as the PESA family of algorithms (Corne *et al.* 2000, Corne, Jerram, Knowles and Oates 2001), selection-for-variation operates purely on locally non-dominated solutions and thus no Pareto-based selection is required at this stage. Comparisons are based purely on density issues, as discussed in the forthcoming Section 2.4.4.

Proximity Promotion in Selection-for-Survival

In the early Pareto-based MOEAs, such as MOGA and NSGA, all of the child solutions that result from the variation stage replace all the parents during selection-for-survival. This is a non-elitist concept that is identical to the (μ, λ)-ES approach. Note that, in these schemes, there is an often implicit assumption that an infinite memory of all non-dominated solutions found during the course of the optimisation is maintained (the so-called *off-line archive*). Selection-for-survival can be considered inactive in these approaches.

Active selection-for-survival systems, often labelled as *elitism*, *on-line archiving*, or *secondary population*, have proved very popular in recent years (Deb 2001a). Much of this popularity springs from a study that compared elitist algorithms favourably with non-elitist algorithms on a variety of real-parameter bi-objective test functions (Zitzler, Deb and Thiele 2000). Elitism (in certain forms) is also a required element for an EA to guarantee convergence in the limit to the global optimum, but the general aim of active selection-for-survival is to improve the quality of approximation sets (in terms of both proximity and diversity) that can be obtained within a given finite number of candidate solution evaluations. The proximity elements of such schemes are considered here.

Following the terminology of Zitzler (2002), selection-for-survival can be incremental or *en bloc*. In the former scheme, solutions are sequentially considered for inclusion in the new

population (thereby the order of presentation can affect the outcome). In the latter approach, the solutions are considered simultaneously (order is unimportant).

Selection-for-survival often contains a large deterministic element. The $(\mu + \lambda)$-ES technique is used by Deb, Pratap, Agarwal and Meyarivan (2002) in the NSGA-II. The best ranked individuals from the combined child and current population pools are retained to form the new population. Tied ranks are decided by diversity considerations, followed by random selection if required. For the most part, this is a deterministic method. The scheme is *en bloc*.

In SPEA, Zitzler and Thiele (1998) used the classical GA generational gap approach, in which a proportion (symbolised by Zitzler and Thiele (1998) as an external population) of the old population is preserved. The actual proportion chosen represents a form of *elitism intensity* (Laumanns, Zitzler and Thiele 2001). As usual, locally non-dominated solutions are preserved. According to Zitzler (1999), this type of approach was first used by Tamaki, Mori, Araki, Mishima and Ogai (1994). Since the number of non-dominated individuals may be greater than the designated fraction of the population to be retained, diversity-based clustering is used to reduce the set as detailed in Section 2.4.4. Zitzler (1999) generalises this scheme as *universal elitism* and relates the mechanism back to its origins in single-objective EAs (De Jong 1975) and earlier multi-objective extensions. In these latter approaches, the reduction criterion could be random (Cieniawski, Eheart and Ranjithan 1995, Ishibuchi and Murata 1996), time-based (Parks and Miller 1998), or single-objective-based (Anderson and Lawrence 1996, Murata, Ishibuchi and Tanaka 1996, Todd and Sen 1997). In the SPEA2 mechanism, dominated points are also preserved if space permits (Zitzler, Laumanns and Thiele 2001). The generational gap strategy, when combined with truncation, is an *en bloc* approach.

In the PESA family of methods, objective-space is divided into a set of non-overlapping hyperboxes (Knowles and Corne 1999, Corne *et al.* 2000, Corne *et al.* 2001). Non-dominated points are retained in the population until a pre-specified maximum size is reached. Solutions are then rejected based on a histogram-type density count of the number of solutions in each hyberbox. Only locally non-dominated solutions are preserved from one generation to the next, regardless of the number of such solutions discovered in relation to the population ceiling. The inclusion of non-dominated children (and removal of dominated or crowded

current population members) is done on an incremental basis, in contrast to universal elitism.

The methods describe above have all been implemented with a global population topology, although they are easily extendible to the island model (Veldhuizen, Zydallis and Lamont 2003). The promotion of proximity was considered for the diffusion topology for EMO by Rowe, Vinsen and Marvin (1996). In this approach, dominance-based comparisons are made between the child and parent at a grid node to determine the survivor.

Proximity Promotion Under Uncertainty

Objective-vector values may have an element of uncertainty due to general modelling errors and the stochastic nature of some models. The critical issue here is how dominance should be interpreted in order to provide correct proximity-based discrimination. Hughes (2001) developed methods to determine the probability of correctly determining dominance between two vectors with Gaussian uncertainty. The probability can be used directly in the ranking process. Hughes (2001) implemented the method within the context of Fonseca and Fleming's (1993) Pareto-based ranking. Teich (2001) developed similar probabilities for uniform uncertainty distributions.

Schemes for good proximity do not directly address the MO requirement for a good distribution of solutions in the approximation set. Thus, diversity-promotion elements have also been considered in EMO. These processes may be entirely distinct from the proximity enhancement processes, but are often embedded and intertwined with these latter processes in the overall system. A review of diversity enhancement mechanisms is provided below.

2.4.4 Methods for Obtaining Good Diversity

Density Estimation in EMO

Many of the EMO methodologies devised to obtain a good distribution of solutions in the approximation set require some measure of the *density* of the population in the neighbourhood of each solution. Various density estimators have been proposed and used in an EMO context. Some of the estimators have been specifically devised for EMO, whilst others have been imported from other scientific disciplines.

The importance of niching in EMO was motivated by Goldberg (1989). The original density estimator proposed for EMO was fitness sharing which is a classical EA method

of effecting niching behaviour. Fitness sharing, as discussed earlier in Section 2.3.3, was originally proposed to combat genetic drift on single-objective, multimodal, cost landscapes. Multi-objective problems can also in a sense be considered multimodal, since the optimal solutions that describe the trade-off surface form one or more plateaus of equal fitness from a Pareto perspective. Since its implementation in MOEAs by Fonseca and Fleming (1993) and Srinivas and Deb (1994), in objective-space and decision-space respectively, fitness sharing has been a very popular approach to diversity enhancement. There has been some discussion in the literature concerning which space (objective or decision) fitness sharing should be performed in. Essentially, as argued by Horn (1997), sharing should be performed in the space within which the DM is most concerned to obtain a good distribution. This is likely to be objective-space, may possibly be decision-space, and indeed a good distribution may be required in both spaces simultaneously. A combined objective-decision scheme is described by Rowe *et al.* (1996). Horn and Nafpliotis (1993) used a *continuously-updated* sharing scheme (to avoid possible instability when used with tournament selection (Oei, Goldberg and Chang 1991)). Horn and Nafpliotis (1993) also suggested joint objective-space-decision-space sharing methods.

As described in the introduction to fitness sharing provided earlier, the success of the technique relies to a significant extent on the chosen value for the niche size parameter, σ_{share}. Methods for selecting the niche size have been developed by Deb and Goldberg (1989), Fonseca and Fleming (1993), Tan, Lee and Khor (1999), and Ray, Kang and Chye (2001). The essential similarity between fitness sharing and kernel density estimation was first noted by Fonseca and Fleming (1995a). The importance of this finding is that it exposes EMO to a new set of techniques for finding a parameter analogous to σ_{share} that already exist in the statistics domain. Fonseca and Fleming (1995a) used Silverman's (1986) approach for the *Epanechnikov* estimator to provide an automatic method of niche size selection. This technique has subsequently been pursued successfully in several EMO applications, such as that described by Griffin, Schroder, Chipperfield and Fleming (2000).

Estimators other than fitness sharing have also been proposed and used in EMO. Histogram-based techniques are a popular alternative, in which the density estimate is a count of the number of solutions that reside in a particular hyperbox of objective-space. The hyperboxes are obtained through a simple gridding of objective-space. In Knowles and Corne's (1999)

Pareto archived evolution strategy (PAES), an adaptive grid spacing is determined by a user-specified number of bisections of the objective range defined by the locally non-dominated solutions. This scheme has been modified to provide restricted limit convergence properties (Knowles and Corne 2003b). The main advantage of this *adaptive grid archiving* (AGA) method is that it permits a pre-specified approximation set size (unlike the method of Laumanns, Thiele, Deb and Zitzler (2002) that also provides convergence in the limit under less strict assumptions).

Histogram density estimation techniques have also been suggested by Lu and Yen (2002), wherein an adaptive cell count similar to PAES is used in the *rank density-based genetic algorithm*. Coello and Toscano Pulido's (2001) *micro-genetic algorithm* is a further method that uses histogram binning. The main advantage of these methods is that they do not require a distance metric that combines non-commensurable objectives. This feature (when expressed in terms of the dominance operator) is argued as one of the key reasons for preferring an MOEA over classical, aggregation-based methods but, as critics have pointed out, the use of aggregation-based density estimators (such as fitness sharing) then undermines these claims of superiority. The main disadvantages of grid-based methods are that the imposed grid structure may be unsuitable for the actual structure of the trade-off surface and the computational complexity of the estimator is exponential in the number of objectives.

Nearest-neighbour (NN) estimators are also prevalent in the EMO community. Zitzler, Laumanns and Thiele (2001) use the kth NN in Euclidean objective-space as a density estimate, where k is determined according to a statistical heuristic based on the square root of the sample size. Abbass, Sarker and Newton (2001) proposed the mean Euclidean distance of the two NN of a solution as a density estimator, which was extended to the mean of the M nearest solutions by Sarker, Liang and Newton (2002). This is similar to the NSGA-II *crowding distance* estimator, which is defined as the mean side length of the hypercube formed using the first NN in each objective as vertices. The advantage of the NN estimators is that they are conceptually simple and have a computational complexity that is linear in the number of objectives (although quadratic in sample size). The fundamental disadvantage is that they require the forced cohesion of potentially non-commensurable objectives. In addition to scaling requirements, the methods may also be sensitive to the parameter k (techniques based on first NN are sometimes misleadingly regarded as 'parameter free'). Thus, a typical

implementation of the NN methodology would be problem-specific. Euclidean distances tend to be popular in the literature for the simple reason that real-valued objective functions tend to be considered in benchmark problems.

Diversity Promotion in Selection-for-Variation

The promotion of diversity is the sole aim of selection-for-variation operators from the PESA family, since all population members are locally non-dominated in these schemes. In PESA itself, binary tournament selection is used: from two individual solutions chosen at random from the population, the solution that is selected for inclusion in the mating pool is that with the lowest density, as defined by the hyperbox bin count (Corne *et al.* 2000). The major innovation in the related *PESA-II* algorithm is that selection-for-variation is *region-based* rather than (the usual) *individual-based*. In this new approach, the binary tournament takes place between populated hyperboxes of objective-space rather than between solutions. The hyperbox chosen during selection is then that with the lowest number of solutions within it. A solution from within this hyperbox is randomly chosen for inclusion in the mating pool.

In methods that include the fitness sharing technique, diversity concerns tend to form a secondary means of fitness assignment. After initial fitness values have been prescribed by a transformation of Pareto rank information, fitness values are modified by the application of Equation 2.4 on a rank-wise basis (Fonseca and Fleming 1993). Thus, fitness sharing tends to solely differentiate between solutions of the same Pareto rank (although this is not guaranteed by the original method, as discussed by Deb (2001a)). The modified fitness values can then be used in any selection mechanism (typical selection schemes are discussed earlier for proximity promotion in Section 2.4.3).

Methods that are not directly based on fitness sharing often use the similar idea of diversity promotion as a secondary consideration to proximity promotion. This is understandable since, as stated by Bosman and Thierens (2003):

> ... since the goal is to preserve diversity *along* an approximation set that is as close as possible to the Pareto optimal front, rather than to preserve diversity in general, the exploitation of diversity should not precede the exploitation of proximity.

In the SPEA2 fitness assignment scheme, the kth NN estimator is used to add a value

to solution cost, with a maximum value that is less than the smallest addition that can be made via dominance (Zitzler, Laumanns and Thiele 2001). Thus, a discriminator is provided that will differentiate only between solutions of the same Pareto rank. This is very similar in nature to the NSGA-II *crowded comparison operator* (Deb, Pratap, Agarwal and Meyarivan 2002). In this latter technique, binary tournaments are primarily decided based on respective membership of dominance equivalence classes. If the solutions have the same rank, then the individual with the smallest density estimate (as calculated via the crowding distance estimator) is selected. Note that, in the case of a tie on density estimate, selection is random.

Diversity Promotion in Selection-for-Survival

During selection-for-survival, diversity promotion tends to become the central consideration when there exist more locally non-dominated solutions than can be retained in the population. In these circumstances, diversity considerations are used to reduce the number of non-dominated solutions to a representative subset.

In the s_s methodology implemented in PAES, if the population is full then a child solution that is non-dominated with respect to both other child solutions and the current population can only enter the archive if the hyperbox into which it would be placed (its density estimate) is less populated than the most populated hyperbox in the current population. In this case, the new solution is included and a solution with the greatest density is removed from the current population. In the PESA schemes, the 'less-than' requirement on the density estimate is relaxed to 'less-than-or-equal-to' and selection is thus random in the case of a tie. This method forms one step toward creating the new population using the PESA incremental update strategy.

Laumanns, Thiele, Deb and Zitzler (2002) also suggested a histogram-based scheme, in which the grids are defined using ϵ-dominance according to the method proposed by Papadimitriou and Yannakakis (2000). A maximum of only one solution is permitted to reside in each hyperbox. A child solution can only be accepted for inclusion in the new population if it resides in a non-dominated hyperbox that is otherwise empty or contains a solution from the current population that the child dominates. Such current solutions are removed from the population, as are those that reside in dominated hyperboxes. Laumanns, Thiele, Deb and

Zitzler's (2002) method ensures convergence in the limit to an ϵ-Pareto optimal set (refer to Section 2.2.2 for a brief discussion on ϵ-dominance) with a finite, if possibly unknown, bound on the population size. The distribution of the solutions is guaranteed to be optimal in the limit, according to the ϵ-Pareto definition of optimality. The disadvantage of the method is that the resolution of ϵ cannot be increased from its initial specification during the optimisation run (although it can be reduced by merging neighbouring hyperboxes). Laumanns, Thiele, Deb and Zitzler (2002) suggests the use of a multiple restart strategy to overcome this problem.

In the universal elitism approach promoted by Zitzler (1999), if the combined non-dominated solutions from the child and current population pools are more numerous than can be preserved, density-based clustering methods are used to reduce the number of solutions to the required quantity. Methods of this nature were previously suggested in the classical OR community (Morse 1980, Rosenman and Gero 1985). The *average linkage method*, based on the Euclidean distance metric, is used in SPEA (Zitzler and Thiele 1999). A problem with this approach is that boundary solutions can be lost. To remedy this, an alternative clustering method was proposed in SPEA2. In this approach, at each step of the truncation algorithm the solution identified for removal is that with the minimum Euclidean distance to the kth NN, where k is incremented sequentially from unity until a single solution can be identified (note that removal may have to be random in some pathological cases).

In the SPEA-type clustering methods, density information must be updated after the removal of each rejected solution. Another population truncation scheme of this nature was implemented by Abbass *et al.* (2001), in which non-dominated solutions are systematically rejected based on the highest densities (the estimator was mean distance to the first and second NN). However, in some approaches, the density information is only calculated at the start of the truncation process. In NSGA-II selection-for-survival, elements from the child and current solution populations are included in the new population based primarily on dominance. If the cut-off point for inclusion resides within a particular dominance-based equivalence class then s_s is made from this class based on a hierarchy of density estimates. The lowest density estimates, as measured by the crowding distance, are deterministically included (Deb, Pratap, Agarwal and Meyarivan 2002).

An innovative approach to selection-for-survival was developed by Farhang-Mehr and

Azarm (2002), who noted that the information metric *entropy* could also be used as a density estimator. In this approach, the selected solutions from the combined child and current population pools are those that maximise an entropy measure. This process is an optimisation task in itself, and is handled by Farhang-Mehr and Azarm (2002) using Monte Carlo methods.

The advantages and disadvantages of the different schemes can broadly be captured by the trade-off between the capability to capture a representative subset and the computational complexity of the technique. For example, the standard NSGA-II approach uses a simple estimator and requires only one density calculation for each solution during the entire s_s stage. By contrast, SPEA2 uses a more complex estimator and requires a complete re-estimation of the remaining population for each solution identified for rejection. This form of clustering can be very computationally expensive. However, the quality of the approximation set distribution that can be developed by the latter method has been shown to be significantly higher than that of the former method (Deb, Thiele, Laumanns and Zitzler 2002). This motivated Deb, Mohan and Mishra (2003) to consider SPEA2 and ϵ-dominance diversity techniques within the context of NSGA-II mechanisms.

The rejection of locally non-dominated solutions at the selection-for-survival stage can cause an MOEA to exhibit partial deterioration and consequent oscillatory behaviour, wherein solutions that are 'lost' are rediscovered and preserved later by the algorithm. This can hamper progress toward the global surface. The ϵ-dominance gridding technique proposed by Laumanns, Thiele, Deb and Zitzler (2002) can be used to prevent this behaviour whilst also permitting rejection of non-dominated solutions. Alternatively, as argued by Everson, Fieldsend and Singh (2002), all non-dominated solutions can be preserved during selection-for-survival. This can lead to a large (potentially infinite) number of solutions in the EA population. Thus, Everson *et al.* (2002) implemented diversity-based selection-for-variation based on a uniform sampling from the current set of locally non-dominated solutions. Computationally efficient methods, based on tree-like data structures, are used to perform the dominance checks.

Diversity Through Parallel Evolutionary Topologies

Distributed population topologies have occasionally been used to provide good approximation set diversity. Rowe *et al.* (1996) used the diffusion model to obtain a natural diversity

across the parallel topology, but pointed out the potential for redundancy in the structure. The island model has also been used in EMO. Okuda, Hiroyasu, Miki and Watanabe (2002) used such a model with $M + 1$ islands (recall that M is the number of objectives to be optimised). In this approach, one of the islands contains a multi-objective EA, whilst the others are single-objective EAs for each of the objectives in the problem. The best solutions for the ith objective are migrated using a star topology (with the MOEA at the centre). Deb, Zope and Jain (2003) also used an island model, with each island focusing on a particular region of objective-space. Each MOEA uses the *guided domination* approach of Branke, Kaußler and Schmeck (2001) to bias the search in a particular direction. A method based on directed cosines is used to obtain the necessary weights for each modified dominance definition in order to cover the entire trade-off surface with minimal overlap. In a further island-type approach, Hiroyasu, Miki and Watanabe (2000) determined the population composition of each island by dividing the global population based on the scale of a particular objective every few generations, with the aim of achieving diversity across the range of the objective. A comprehensive review, discussion, and extensions of parallel MOEAs is provided by Veldhuizen *et al.* (2003).

Diversity Through Probabilistic Model-Building

After initial empirical analysis suggested that an approach based on a single model could not adequately search for and represent a complete approximation set, Thierens and Bosman (2001a) considered the use of mixture models to preserve diversity across the trade-off surface. In this approach, clustering methods are used to group elements of the population that reside in similar areas of objective-space. Model estimators are then fitted to each cluster. Laumanns and Ocenasek (2002) stressed the importance of diversity maintenance in decision-space when using probabilistic models in order to avoid redundancy and provide enough information to build an acceptable model.

Miscellaneous Diversity Enhancement Concepts

Mating restriction. In the context of a study on multimodality, Deb and Goldberg (1989) suggested that the recombination of parent solutions from separate regions (of either objective or decision-space) often lead to poor child solutions, known as *lethals*. Thus, these researchers suggested that recombination between remote solutions should be prohibited where possible,

in an approach known as *mating restriction*. This method was originally conceived of by Booker (1982) in order to promote diversity in the population (Deb and Goldberg 1989). Since the Pareto front spans a (potentially expansive) region of objective-space, it has therefore been suggested that mating restriction might be necessary to improve MOEA search efficiency (recombination from spatially remote areas of the trade-off surface may produce lethals) (Fonseca and Fleming 1993). The key parameter in mating restriction is the distance over which recombination will be prohibited, σ_{mate}. This is often obtained in the same way as for the fitness sharing parameter σ_{share}. In practice, the definition of distance may also be challenging to determine, since (as for fitness sharing) it may involve the forced cohesion of non-commensurable objectives.

Lateral diversity. Sometimes it may prove necessary to preserve diversity in dominated areas of the search space in order to ultimately achieve an approximation set with good proximity and diversity. Thus a trade-off exists between diversity exploitation and proximity exploitation (Bosman and Thierens 2003). Methods for controlling this trade-off, through the preservation of *lateral diversity*, have been proposed by Deb and Goel (2001), Laumanns *et al.* (2001), Laumanns and Ocenasek (2002), and Bosman and Thierens (2003). In Deb and Goel's (2001) scheme, a proportion of the population is allocated to each non-dominated front according to a pre-specified plan. Interestingly, in the classic fitness sharing scheme for EMO implemented in MOGA, dominated points can receive a greater selection probability than non-dominated counterparts, thus promoting lateral diversity (Fonseca and Fleming 1993). Until recently, this attribute was regarded as a weakness of the MOGA framework (Deb 2001a).

Target vectors approach. The use of a *target vector* in objective-space has been proposed in the classical OR community, and used in EAs, as a method for scalarising performance. The task is to minimise some scalar measure of solution distance to this vector (Wienke, Lucasius and Kateman 1992). This approach has been extended in the multi-objective EA community to permit multiple target objective vectors (Lohn, Kraus and Haith 2002). Candidate solutions gain fitness by meeting various target vectors. If a vector is met by several solutions then the fitness associated with that vector is shared between them, thus encouraging solution diversity.

2.4.5 Preference-Based Methods

Introduction

Decision-maker preferences are required in order to focus the approximation set on to a desired sub-region (ROI) of the overall trade-off surface. Several preference-based schemes exist in the EMO literature, although this particular facet of research tends to have been somewhat overlooked.

Preference-based schemes can be classified according to when the preference information is used to influence the search. Thus, *a priori* schemes exist, in which DM preferences are incorporated before the search begins. In *progressive* methods, DM preferences are incorporated during the search. The key advantage of these techniques over *a priori* methods is that the DM may be unsure of his or her preferences at the beginning of the procedure and may be informed and influenced by information that becomes available during the search. The final class of methods is *a posteriori*, in which a solution is chosen from the approximation set returned by the optimiser. Many EMO researchers apparently view this approach as standard, with the actual preference articulation process lying outside the pure EMO domain. This mindset is possibly due to the EMO focus on bi-objective problems.

A review of the literature on preference methods for EMO is offered below. Refer to Coello (2000) for an alternative survey.

A Priori Preference Methods

Branke *et al.* (2001) suggested the creation of a weighted-sum of objectives *for each objective* in turn. This is achieved by adding weighted terms of other objectives together with a unit weight on the initial objective. Each weight represents how much gain would be required in another objective for a unit loss in the objective under consideration. Solutions are then compared in terms of performance on these new weighted-sums. This approach, in effect, modifies the dominance operator such that each solution dominates a potentially larger region (it can also be viewed as applying the standard definition of dominance to a transformation of objective-space). This method suffers from the standard difficulties of a weighted-sum approach: it cannot handle non-convex regions of the search space and requires *a priori* knowledge of the trade-offs between objectives.

An alternative method of the use of weights was proposed by Parmee, Cvetković, Watson

and Bonham (2000). The weights are obtained by using fuzzy binary preference relations on objectives (which can also be used to obtain an ordering on the importance of the objectives). Note that this method assumes transitivity and infers relations in order to prevent possible contradictions (Cvetković and Parmee 2002). Once the weights have been obtained, Parmee *et al.* (2000) propose that these are used in a standard weighted-sum (prior to solution by a single-objective EA) or are used in the *weighted Pareto method*. In this latter approach, the global dominance relation is decomposed into a count of pair-wise 'less-than-or-equal-to' comparisons between the objectives. The results of each comparison can be weighted prior to being included in the count. The necessary total required for dominance can also be adjusted.

Existing preference articulation approaches developed in the operational research community can also be integrated within an evolutionary computing scheme. Rekiek, De Lit, Pellichero, L'Eglise, Falkenauer and Delchambre (2000) used the *PROMETHEE II* outranking system to order the population of an EA in terms of preference at each generation. The ordering was then used to select individuals for reproduction. An *a priori* approach was taken, although a progressive scheme could be formulated by updating PROMETHEE II during the search.

The aim of preference articulation is to concentrate on particular, distinct sub-regions of the trade-off surface. Hence, instead of developing a uniform distribution of solutions (as is the case for a standard MOEA) a preference-based algorithm aims to develop a biased distribution. One of the key EA techniques to obtain a suitable distribution is fitness sharing, as described in Section 2.3.3. Deb (1999b) exploited this operator by introducing bias into its calculations. Normalised weights for each objective are included in the distance metric computation. Thus, the density of solutions should become more numerous in more important objectives. Note that this approach requires weights to be specified.

Deb (2001b) also investigated the use of MOEAs for *goal programming* (Aouni and Kettani 2001). The traditional approach to goal programming is to minimise a weighted-sum of the deviations from unmet goals. However, this approach encounters the usual difficulties with weighting non-commensurable objectives, normalising the objectives, and with handling non-convexity. Thus, Deb (1999a) formulated the problem as one of minimising deviations from unmet goals, where each deviation is treated as a separate objective. An NSGA-based algorithm was used to find a family of Pareto optimal solutions (in terms of deviations from

goals). Note that the solutions in the 'Pareto optimal' set may only be *satisficing* solutions (acceptable to the DM but not globally optimal), since the method does not attempt to optimise past the goal levels of the DM.

In a somewhat different approach to those mentioned above, Shaw and Fleming (2000) embedded preference information within the workings of the EA (rather than expressing direct preferences on the objectives). In a factory scheduling problem, the DM can set pseudo-fuzzy preferences on product-to-line mappings. This data is then used by a schedule builder, which interprets EA chromosomes in order to produce a schedule (refer to Section 2.3.3). In the application, these mappings were elicited prior to the run of the optimiser, but they could be provided in a progressive manner if required. Shaw and Fleming (2000) also suggested the use of Fonseca and Fleming's (1998a) progressive objective-level preference system, which is described next.

Progressive Preference Methods

The first truly progressive MOEA scheme was introduced by Fonseca and Fleming (1998a) as an extension to the Pareto-based ranking described in Section 2.4.3 (Fonseca and Fleming 1993). The *preferability* operator represents the state-of-the-art in MOEA preference articulation. In this scheme, the DM can set goal values and priority levels for any objective. This can be done at any time during the run of the MOEA and can be updated when required. The data feeds into a modified definition of dominance, which provides a unification of Pareto optimality, the lexicographic method, goal programming, constraint satisfaction, and constrained optimisation. All these methods, plus hybrids, can be derived from the preference operator. Fonseca and Fleming (1998a) also developed an on-line user interface that featured the *parallel coordinates* method of visualising trade-offs between objectives (Inselberg 1985).

Deb, Pratap, Agarwal and Meyarivan (2002) developed a *constrained-domination* approach that is very similar to the preferability operator. The main distinction is that, in this new scheme, an overall quantity of goal violation is calculated. This enhances the amount of information available to the search, but requires the forced cohesion of objectives. Another similar scheme known as *favour* has been proposed by Drechsler, Drechsler and Becker (2001). Tan, Khor, Lee and Sathikannan (2003) developed logical connectives to allow a DM to make alternative preference scenarios for a problem in the context of preferability-type

schemes.

Todd and Sen (1999) proposed an alternative progressive scheme, which incorporates learning and automation of DM preferences. Rather than setting goals and priorities, the DM is asked to make judgements on a set of potential solutions at various intervals during the optimisation process. This approach is described further in the section on automating the decision process. Note that this scheme may be more DM-intensive than Fonseca and Fleming's (1998a) method.

A Posteriori Preference Methods

Deb (2001a) suggests three possible techniques for obtaining a subset of solutions from a final population of Pareto optimal candidate solutions. *Compromise programming* can be used to select a solution that is closest to some specified reference point (Zeleny 1973). Alternatively, the solution with the 'best' marginal rate of substitution for each pair of objectives (the amount of improvement in one objective that can be obtained by making a unit sacrifice in another) can be chosen. Finally, pseudo-weight vectors can be computed for each solution in the family. A solution is chosen that corresponds to a vector that is closest to a set of weights specified by the DM. Massebeuf, Fonteix, Kiss, Marc, Pla and Zaras (1999) used the *PROMETHEE II* outranking method to choose a solution from the final set of solutions found by an MOEA.

Semi-Automation of the Decision Process

Preference articulation schemes can require the DM to study and comment on a large amount of information. It is important not to overwhelm the DM with information, since this will reduce the quality of the preference information gathered. This issue is especially pertinent to progressive methods. In order to reduce the demands placed on the DM, and to improve search efficiency, researchers are beginning to look at means of automating parts of the decision-making process. Two such schemes, which use neural networks to model the preference structure of the DM, are discussed below.

Todd and Sen (1999) used an elitist MOEA with a (conceptually) external population of Pareto optimal solutions. In addition to a Pareto rank, each solution also carried a preference score between 0 (least preferable) and 1 (most preferable). The preference score was used

as a fitness value to insert members of the Pareto population into the mating pool. The DM provided preference scores for a selection of ten individuals every ten generations. The individuals were chosen to provide a good spread of preference. This data was used to train a neural network model of the DM preferences. The neural network then determined preference values for all other individuals automatically.

Sun, Stam and Steuer (2000) also used a neural network to learn the DM preference structure, this time in the context of the multi-criterion decision-making (MCDM) *Tchebycheff* method (Steuer 2001). The DM is invited to assign values to, or to make pair-wise comparisons of, a subset of solutions. The assignment process results in a reciprocal comparison matrix, the normalised principal eigenvector of which can be viewed as the 'priority' given to each solution. Thus, a neural network is trained using normalised objective values as input data and preference values as output data. The network is used to filter non-dominated solutions found by the Tchebycheff method, with the predicted most preferred solutions being shown to the DM. The process then continues as before. Sun *et al.* (2000) reported that the technique provided superior results to the standard Tchebycheff procedure across benchmark problems of various magnitudes.

In both of the above methods, further research is required to explore how imprecise and inconsistent preference data can be successfully handled.

2.5 Summary

This chapter has sought to introduce multi-objective optimisation within the context of an overall problem-solving framework. The key concept of Pareto dominance has been explained, and the three key requirements of a multi-objective optimiser — to produce an approximation set with good proximity and diversity in regions of interest to the decision-maker — have been described.

The class of metaheuristic techniques known as evolutionary algorithms have been introduced and the special utility of these methods for multi-objective optimisation has been described. The various EMO components that have been developed to meet each of the above requirements — from the simple to the more complex — have been introduced in some detail.

Elitism and advanced diversity promotion methods have been particularly concentrated on in the contemporary EMO community. But can these new innovations really be said to

produce improved approximation sets over the basic mechanisms sketched by Goldberg (1989) and implemented in the MOGA, NPGA, and NSGA? In the following chapter this question is explored by considering the effect of advanced elitism and diversity-promotion schemes in the context of the original EMO algorithm, MOGA. The familiar benchmark problems are reconsidered using a rigorous experimental framework.

Chapter 3

Enhancements to the Multi-Objective Genetic Algorithm

3.1 Introduction

As evident from the review of evolutionary multi-objective optimisation in Chapter 2, there have been many algorithmic developments in this very active research field since the first published implementation of a Pareto-based MOEA by Fonseca and Fleming (1993). As a result, EMO practitioners are faced with a number of design choices beyond those encountered in a standard evolutionary algorithm. In order to exploit the true potential of the evolutionary meta-heuristic, the optimiser should be tailored to the application rather than used simply as a black-box (Michalewicz and Fogel 2000). The analyst should perhaps be encouraged to develop a bespoke MOEA rather than resort to an algorithm *brand* such as MOGA, NSGA-II, or SPEA2. Thus, the nature of a design choice would be, for example, 'What mechanisms should be used to promote diversity in this application?' rather than 'Should NSGA-II be used instead of SPEA2?'.

There is a tendency in the EMO research community to compare different brands of algorithms, treating each one as an inseparable entity. However, as is clear from the description of EMO methods in Section 2.4, the actual mechanisms responsible for performance reside *within* each algorithm. By considering algorithms at a decomposed and abstracted level, it should become easier to identify the underlying components, and interactions between components, to which the observed performance can be attributed. One of the rare examples

where this approach has been taken is the excellent study by Laumanns *et al.* (2001) based on abstract EMO mechanisms identified by Laumanns, Zitzler and Thiele (2000).

In order for the analyst to make informed EMO design choices, knowledge is required of how the performance of each fundamental component changes with context, such as the class of application and the partner components that together comprise the complete MOEA. It is also useful to understand the performance sensitivity of a design choice within some defined contextual boundary. With these considerations in mind, this chapter aims to more accurately expose the performance of the following popular EMO strategies using a rigorous and tractable experimental procedure:

Ranking: two methods for determining Pareto-based performance in order to achieve good proximity in the approximation set (see Section 2.4.3).

Sharing: two methods to modify the expected probability of selection in order to promote a good distribution in the approximation set (see Section 2.4.4).

Elitism: a selection-for-survival method aimed at obtaining both good proximity and good diversity (see Section 2.4.3).

The experimental framework is introduced in Section 3.2. The benchmark suite of test problems used in the study is described, together with suitable performance indicators. A method for statistical significance testing is introduced, as is an appropriate visualisation technique. A baseline MOEA is developed in Section 3.3, and its performance is established. The effects of the various advanced EMO design choices are then considered with reference to this baseline. The two most popular multi-objective ranking strategies in the literature are contrasted in Section 3.4. Selection-for-variation methodologies for the promotion of diversity are discussed in Section 3.5, in which the performances of the established *Epanechnikov* method and a generalised ranking-based method are compared. An elitist strategy is developed and tested in Section 3.6. In Section 3.7, a state-of-the-art MOEA incorporating both elitism and rank-based sharing is considered.

Table 3.1: Test function characteristics

Name	Attributes
ZDT-1	Convex front
ZDT-2	Non-convex front
ZDT-3	Piece-wise continuous convex front
ZDT-4	Many local fronts, single global front
ZDT-5	Deceptive problem, discrete front
ZDT-6	Non-uniform distribution across a non-convex front

3.2 Experimental Framework

3.2.1 Overall Methodology

Evolutionary algorithms are complicated non-linear systems that have proved very challenging to analyse. The large number of free parameters, and the interactions between them, can form obstacles to a confident interpretation of performance. The *EMO empirical inquiry framework* presented in this section seeks to increase the benefit of empirical testing of algorithms. The following attributes of the methodology are emphasised:

- Modular, traceable, MOEA configuration changes.
- Transparent, understandable, test problems with realistic properties.
- Appropriate, accurate, performance measures.
- Rigorous, informative, analysis, including tests for statistical significance and visualisation.

The test suite, performance indicators, and statistical and visual analysis techniques used in this study are discussed in detail in the following sub-sections.

3.2.2 Test Suite

The established set of test problems developed by Zitzler *et al.* (2000) is used in this study. The suite consists of six, tractable, bi-objective functions, with varying characteristics as summarised in Table 3.1. Equations for the various functions are provided in Appendix A.

These functions cover many of the features that may be found in real-world problems and are comparatively straightforward to analyse. The main concern is that, for each problem, one

of the objectives is a function of only a single decision variable (mapped without modification in the first four test problems). In particular, this direct mapping between decision-space and objective-space may cloud the issues surrounding diversity preservation. It should also be noted that these test functions consist of two objectives only. Much care should be taken before transferring conclusions drawn from these functions to problems with a higher, and more realistic, number of objectives.

3.2.3 Measuring Performance

As described in Section 2.2.3, the performance of an MOEA can be decomposed into three interacting criteria: (i) the proximity of the identified non-dominated solutions to the true Pareto front, (ii) the diversity of the approximation set across the trade-off surface, and (iii) the pertinence of the solutions to the decision-maker. In this study, the ROI will be assumed to be the entire Pareto front. Therefore, only proximity and diversity performance are considered further.

Various performance indicators have been proposed to measure these aspects of quality (Deb 2001a). They can typically be classified according to function (which part of overall performance they measure) and provide a scalar value that represents the quality of a locally non-dominated set. Many indicators are unary (they describe the absolute performance of one approximation set), although a few are binary (they describe the relative performance of two sets). Some unary indicators require a reference set with which a comparison can be made. This often necessitates that the true surface is known to the analyst and can be sampled. Unary methods are advantageous, however, in that conventional statistical tests can be straightforwardly applied. A review of performance indicators is provided by Deb (2001a).

Zitzler *et al.* (2003) have shown that no finite combination of unary measures can indicate whether one approximation set is superior to another (from the perspective of the dominance relation). Thus, care must be taken when making statements about global performance. This study adopts the *functional* approach described by Deb and Jain (2002). Specific unary indicators are used to evaluate specific aspects of performance. There is no attempt to describe global performance using a unary indicator or indeed a combination of such indicators. Thus the failure of an indicator to respect the dominance operator is not of immediate concern.

This study uses three known performance indicators: *generational distance* to measure proximity, *spread* to measure diversity, and *attainment surfaces* to provide visualisation of the results. These indicators are described in further detail below.

Generational Distance

The proximity of each non-dominated point produced by an MOEA can be measured in terms of its distance to the closest part of the global trade-off surface. These distances can be averaged to provide a measure of accuracy for a complete approximation set. The definition of distance is dependent on the problem domain. Euclidean distance is a natural choice for the ZDT test problems, and has been adopted in this work. Note that the objective values must be normalised if they are not of the same scale. The generational distance indicator is formalised in Equation 3.1, where GD is the generational distance, $\boldsymbol{Z_A}$ is the obtained approximation set, and d_i is the closest distance between the ith objective vector $\in \boldsymbol{Z_A}$ and any vector in $\boldsymbol{Z_*}$, where $\boldsymbol{Z_*}$ is the set of globally non-dominated objective vectors (Veldhuizen 1999).

$$GD = \frac{1}{|\boldsymbol{Z_A}|} \sum_{i=1}^{|\boldsymbol{Z_A}|} d_i \tag{3.1}$$

The main advantages of this indicator are its computational simplicity and its amenability to statistical analysis. The disadvantage is that the set $\boldsymbol{Z_*}$ is required. Given that the size of this set may be infinite, a finite representation is required for computational purposes. This latter set should be sufficiently numerous and should be (parametrically) uniformly distributed across the trade-off surface in order to avoid bias. Fortunately, for the ZDT problems, the global Pareto front is explicitly defined in each case. For the continuous and piece-wise continuous trade-off curves, uniform parametric sampling is quite straightforward (achieved, for example, using equations for curvature). The discrete trade-off surface of ZDT-5 can easily be enumerated.

Spread

Consider the distribution of distances between nearest-neighbour objective vectors. In the case of a uniform distribution, all such distances will be identical and will equal the mean of the distribution. In the general case, uniformity can thus be measured by considering the

difference between a nearest-neighbour distance and the mean of all such distances. Schott (1995) originally formulated the sum of all these differences as an indication of the uniformity of the identified trade-off surface. This was extended by Deb, Pratap, Agarwal and Meyarivan (2002) to include a measure of the extent of the obtained distribution. The resulting indicator is shown in Equation 3.2, where Δ is the spread, d_m^e is the Euclidean distance from the extreme point in $\boldsymbol{Z_A}$ to the extreme point in $\boldsymbol{Z_*}$ for the mth objective, M is the number of objectives, d_j is the Euclidean distance between consecutive objective vectors in $\boldsymbol{Z_A}$ and, $\bar{d}$ is the mean of all d_j.

$$\Delta = \left[\sum_{m=1}^{M} d_m^e + \sum_{j=1}^{|\boldsymbol{Z_A}-1|} |d_j - \bar{d}| \right] \Big/ \left[\sum_{m=1}^{M} d_m^e + |\boldsymbol{Z_A} - 1|\bar{d} \right] \tag{3.2}$$

The first term in the numerator of Equation 3.2 describes the extent of the trade-off surface that is not represented in $\boldsymbol{Z_A}$. The second term describes the non-uniformity of the $\boldsymbol{Z_A}$ distribution. The denominator seeks to normalise these measures with respect to the total magnitude of the trade-off surface. Smaller values of Δ indicate superior diversity to larger values.

The spread metric is a suitable diversity-measuring indicator for bi-objective problems. However, in the form presented in Equation 3.2, the spread metric cannot be used in problems with more than two objectives because the concept of consecutive vectors does not exist in higher dimensions.

Attainment Surfaces

Fonseca and Fleming (1993) introduced the concept of an *attainment surface*. Given a set of non-dominated vectors produced by a single run of an algorithm, the attainment surface is the boundary in objective-space that separates the region that is dominated by or equal to the set from the region that is non-dominated. Note that this is fundamentally different to interpolating between the vectors. This latter approach is not, in general, correct because there is no guarantee that any intermediate vectors actually exist and, even if this were the case, the corresponding solutions are unknown. The concept of the attainment surface is illustrated in Figure 3.1.

Attainment surfaces serve two very useful purposes. One the one hand, they provide a convenient means of visualising the results from multiple runs of an optimiser. On the

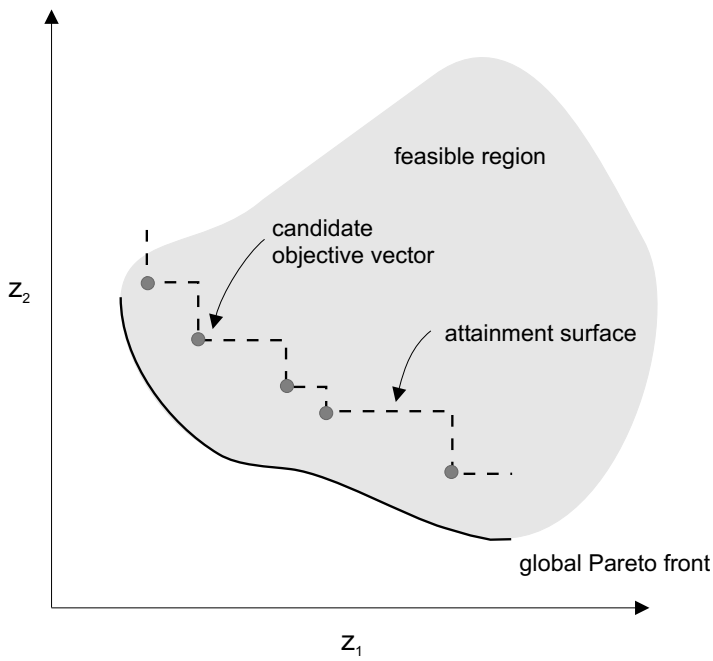


Figure 3.1: Example attainment surface

other, through the use of *auxiliary lines*, they allow for algorithm comparisons using well-known univariate statistical tests. In this study, the attainment surfaces are used purely for visualisation. Refer to Fonseca and Fleming (1996) and Knowles and Corne (2000) for examples of the comparative statistics work.

The superposition of multiple attainment surfaces, as shown for the bi-objective example in Figure 3.2, provides a qualitative indication of the performance of a particular MOEA configuration. The regions of objective-space partitioned by the surfaces can be interpreted probabilistically. Given that both objectives are to be minimised, the region below all the attainment surfaces contains performance vectors that were not matched by the MOEA in any run. The region above all the surfaces contains vectors that were exceeded by all runs. In the intermediate regions, the performance vectors were exceeded on an intermediate number of occasions. Thus, it is possible to obtain a family of vectors that, individually, would be obtained in a given percentage of runs. The heavy line in Figure 3.2 shows the 50%-attainment surface (akin to the median statistic). Similarly, the grey lines indicate the 25% and 75% surfaces (quartiles). The 0% and 100% surfaces are shown as the dotted lines. This method of visualisation is employed throughout this study. The attainment surfaces provide information

on location, dispersion, and skewness, in a similar manner to the box plot (Cleveland 1993). This methodology provides more reliable information than the unification-of-runs approach adopted by Zitzler *et al.* (2000). It also has a quantitative performance equivalent, developed by Grunert da Fonseca, Fonseca and Hall (2001).

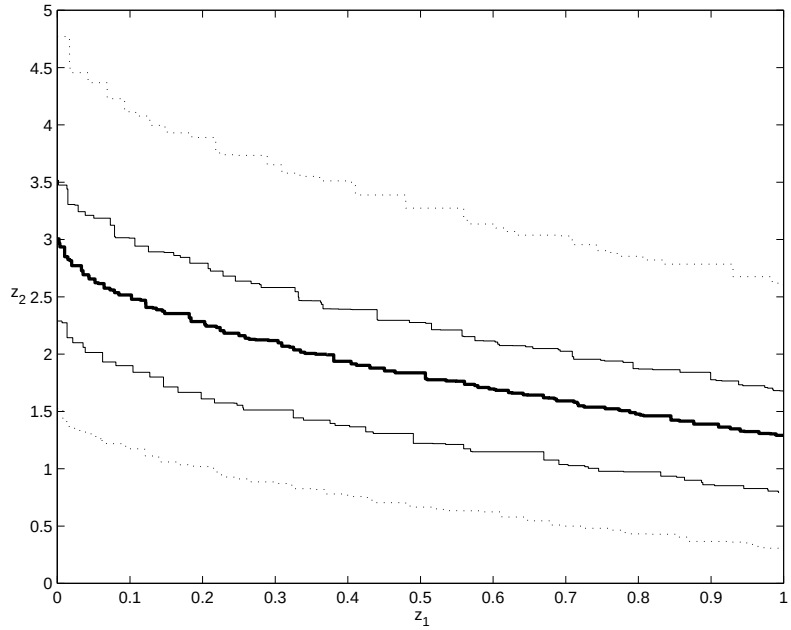


Figure 3.2: The superposition of multiple attainment surfaces

The attainment surface concept can be extended to any number of objectives, although visualisation becomes problematic at any dimension higher than three. Computational complexity also increases significantly.

3.2.4 Analysing Performance

Upon completion of a single run of a specific MOEA configuration on a particular problem, two sets of non-dominated objective vectors (and associated solutions) are obtained, namely:

final population the non-dominated vectors in the final population of the algorithm.

off-line archive the complete set of non-dominated vectors identified by the algorithm.

The first of these sets is used for analysis and comparison purposes in this study since it provides the most appropriate measure of the on-line trade-off surface *maintenance* capabilities of an algorithm. An evolutionary algorithm is a stochastic process and, thus, multiple runs (samples) are required in order to infer reliable conclusions as to its performance. Hence, 35 runs have been conducted for each MOEA configuration when applied to a particular test problem. The performance of the algorithm is expressed in the resulting distributions of generational distance and spread. A statistical comparison of two configurations is then possible through the use of a test statistic. In this study, the mean difference between two generational distance (or, alternatively, spread) distributions is taken as the test statistic. The significance of this observed result is then assessed using *randomisation testing*. This is a simple, yet effective, technique that does not rely on any assumptions concerning the attributes of the underlying processes, unlike conventional statistical methods (Manly 1991). The central premise of the method is that, if the observed result has arisen by chance, then this value will not appear unusual in a distribution of results obtained through many random relabellings of the samples. The randomisation method proceeds as follows:

1. Compute the difference between the means of the samples for each algorithm: this is the observed difference.

2. Randomly reallocate half of all samples to one algorithm and half to the other. Compute the difference between the means as before.

3. Repeat Step 2 until 5000 randomised differences have been generated, and construct a distribution of these values.

4. If the observed value is within the central 99% of the distribution, then accept the null hypothesis. Otherwise consider the alternative hypotheses. This is a two-tailed test at the 1%-level.

The null hypothesis is that the observed value has arisen through chance and so there is no performance difference between the two configurations. The alternative hypotheses are that the difference is unlikely to have arisen through chance and that one configuration has outperformed the other (depending on which side of the distribution the observed difference falls, and the direction in which the difference has been calculated). By demanding a 1%-level

of significance, the probability of making a *Type II* error (accepting the null hypothesis when it is false) is increased. This conservative criterion provides improved confidence that the detected difference is truly reflective of structural differences between algorithms.

Note that the observed value is included as one of the random relabellings since, if the null hypothesis is true, then this value is one of the possible randomisation results. 5000 randomisations is regarded as an acceptable quantity for a test at the 1%-level (Manly 1991). The results of randomisation testing are simple to visualise, as shown by the example in Figure 3.3. The randomised results are described by the grey histogram, whilst the observed result is depicted as a filled black circle. Each row shows the performance on a particular test function (from ZDT-1 at the top, to ZDT-6 at the bottom). The left-hand column indicates the relative performance regarding proximity, and the right-hand column shows the corresponding difference in diversity. It is usually clear from the figure whether or not the observed result is statistically significant, although it may occasionally prove necessary to resort to a closer analysis of the underlying randomisation data.

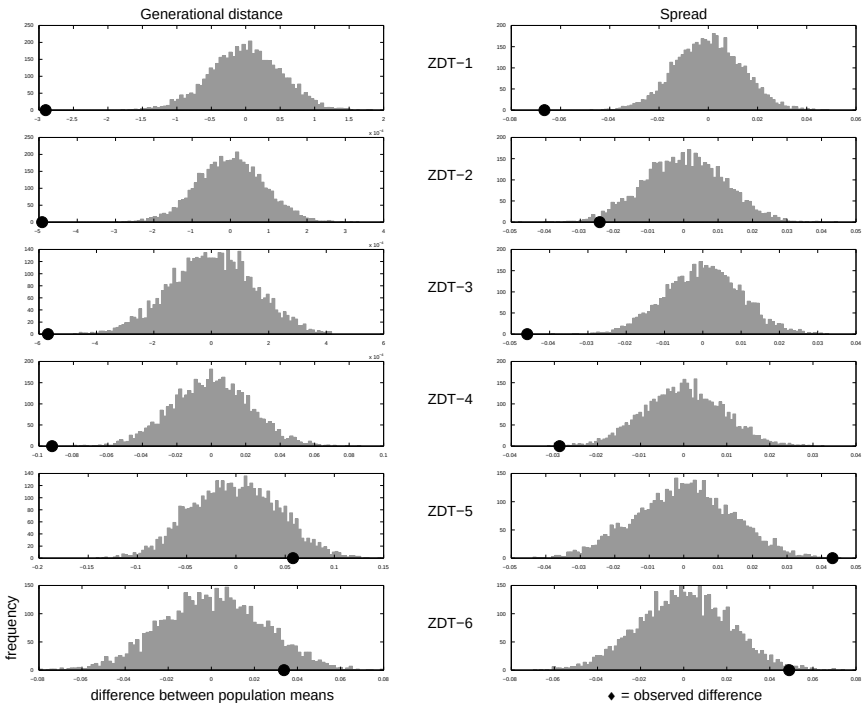


Figure 3.3: Randomisation testing — example results

In the example in Figure 3.3, given data for two algorithms A and B together with a test statistic of $mean(B) - mean(A)$, then the following results are observed:

- B obtains fronts closer to the true front than A for ZDT-1, 2, 3, and 4.
- B produces a superior distribution of objective vectors to A for ZDT-1, 3, and 4.
- A offers a superior distribution to B on ZDT-5.

No other results are significant at the 1%-level, although clearly some of these cases offer more evidence against the null hypothesis than others.

3.3 Baseline MOEA

The baseline optimiser used in this study has been developed according to the holistic design principles championed by Michalewicz and Fogel (2000) and has previously been shown to be effective at solving the ZDT test problems (Purshouse and Fleming 2001). A summary of the algorithm is provided in Table 3.2.

The multi-objective performance of a solution is scalarised using Fonseca and Fleming's (1998a) Pareto-based ranking procedure. A solution is ranked according to the number of solutions in the population that are *preferred* to it. If the entire Pareto front is to be identified, the preference relation collapses to a test for pure Pareto dominance.

When ranking is complete, initial fitness values can be prescribed. The population is sorted according to rank and fitnesses are assigned by interpolating between the highest fitness value for the best rank and the lowest fitness value for the worst rank. In the baseline algorithm, linear interpolation is used and fitness is varied between the population size (highest) and unity (lowest). The ratio of these two fitnesses is a definition of the selective pressure of the assignment mechanism, as explained in Section 2.3.2. Solutions of the same rank then have their fitnesses amended to the mean of the original assignments for that rank. Since part of this study is concerned with the effect of diversity-preserving mechanisms, no manipulation of the above fitnesses through sharing is undertaken.

Stochastic universal sampling has been chosen as the selection-for-variation mechanism (Baker 1987). This method achieves maximum spread with minimal bias, but is non-parallelisable. As part of this procedure, the above fitness values are normalised to provide an

Table 3.2: Baseline configuration

EMO Component	Strategy
General	
Population size	100 per generation.
Total generations	250
Representation	
Real parameter	Concatenation of real number decision variables. Accuracy bounded by machine precision.
Binary function	Binary string, 80 bits in length. Defined by the problem.
Selection-for-survival	Non-elitist: $(\mu, \lambda) = (100, 100)$ (no generational gap).
Selection-for-variation	[1] Fonseca and Fleming (1998a) Pareto-based ranking. [2] Linear fitness assignment with rank-wise averaging. [3] No modification of fitness to account for population density. [4] Mechanism: stochastic universal sampling. (Baker 1987)
Variation	
For real representations	[1] Naïve crossover. Probability = 0.8. [2] Gaussian mutation (initial search power of 40% of variable range; sigmoidal scaling set to 15; feasibility requirement of one standard deviation). Probability = Expected value of 1 variable per chromosome.
For binary representations	[1] Single-point two-parent crossover. Probability = 0.8. [2] Simple bit-flipping mutation. Probability = 1/80.

expected number of selections for each solution. In total, 100 selections are required since the chosen selection-for-survival strategy is that all offspring replace all parents (no generational gap) and since for the chosen recombination operators two parents are required to produce two offspring. In effect, this is the classic (μ, λ) approach.

Since five of the test problems feature real number decision variables, it is logical to use a real number representation for these problems. Hence, a candidate solution is described by a concatenation of phenotypic decision variables. This representation offers a number of advantages over a binary encoded approach: it is faster to manipulate, it is empirically associated with more consistent optimiser performance, it permits much greater precision, and it is intuitively closer to the problem-space (Michalewicz 1992). The other test problem explicitly uses binary variables, thus a binary representation is natural for this problem.

Different representations require different search operators. For the binary chromosome case, the familiar single-point two-parent crossover and bit-flipping mutation operators are employed. Good results are known to be achievable using this simple approach (Zitzler *et al.* 2000). Various operators for real representations have been suggested (Herrera *et al.* 1998). This study uses the so-called *naïve crossover* in conjunction with a *Gaussian mutation*

operator. The former of these search tools is a very simple single-point two-parent crossover operator, where the crossover sites are limited to points between decision variables. This offers quite a low level of exploration, since it cannot generate any values for decision variables that were not present in the original population. However, when coupled with a complementary highly explorative search tool, the resulting search capabilities are considerable. Gaussian mutation is one such operator.

Gaussian mutation was first used in the early ES algorithms and was briefly described in Section 2.3.2. Each decision variable within a candidate solution is probabilistically tested for individual mutation. A major benefit of the p.d.f. approach is that the exploration-exploitation trade-off is directly controllable through the standard deviation of the mutation distribution. In this study, the standard deviation is scaled using a function of the proportion of optimiser iterations completed (assuming that a maximum number of such iterations has been provided *a priori* as the termination criterion). A *sigmoidal* function is used herein because it can be shaped to allow both substantial periods of *macro-* and *micro-mutations* (respectively, large- and small-scale expected mutation perturbations with respect to the decision variable range).

A potential difficulty with Gaussian mutation is that it can produce infeasible solutions when the decision variable range is bounded, through the generation of child solutions with decision variable values outside of the defined range. A simple solution to this problem is to crop infeasible values to the nearest feasible equivalent, but this can cause a large bias in the search toward extreme decision variable values. Thus, in this study, the standard deviation of the operator is manipulated to control the amount of bias that is permitted. If the analyst predefines the number of standard deviations of the p.d.f. that must be associated with feasible child solutions (or, alternatively and equivalently, the probability that the mutation will produce a feasible solution) then the standard deviation can simply be rescaled on-line to ensure that this condition is met. In the study, the overall probability distribution is also manipulated to provide a more exploratory search away from the infeasible region by scaling the probability of mutation in either direction along the decision variable as a (linear) function of the standard deviation rescaling found to be required. Further details on the specialised Gaussian mutation operator are provided by Purshouse and Fleming (2001).

Attainment surfaces illustrating the performance of the baseline algorithm are shown in

Figures 3.4 through 3.9.

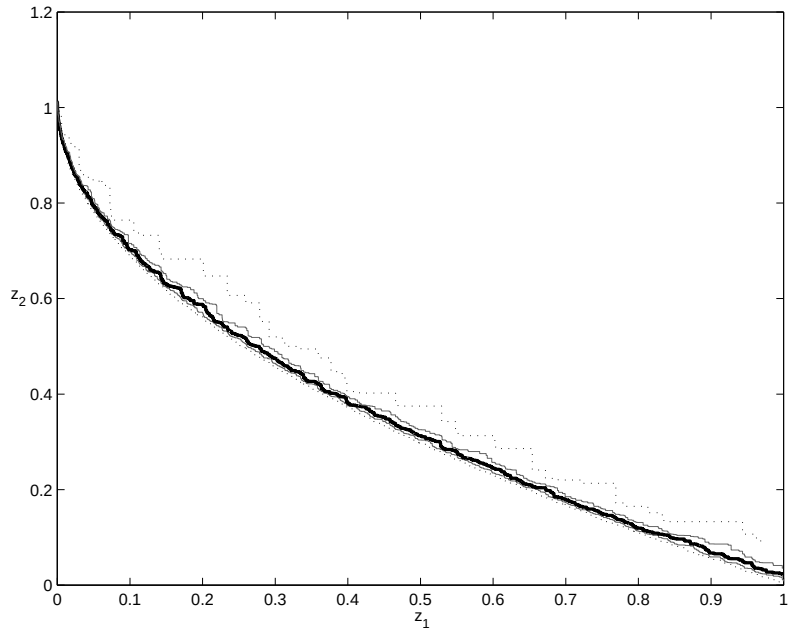


Figure 3.4: Attainment surface — baseline MOEA solving ZDT-1

Particularly good results were achieved for ZDT-1, ZDT-2, and ZDT-3 (Figures 3.4, 3.5, and 3.6 respectively) in terms of both proximity to the global Pareto front and diversity across the front. The tight envelopes of attainment indicate the high level of consistency achieved in these cases. Proximity was especially good for ZDT-3. As has been previously observed by Purshouse and Fleming (2001), the MOEA struggled to achieve good coverage of the surface as z_1 approaches zero on ZDT-2. Note that this is a region where there is relatively little trade-off between the objectives.

As shown in Figure 3.7, the wider envelopes of attainment produced for the multi-fronted ZDT-4 signify entrapment in a locally non-dominated front. On no occasions did the MOEA converge to the global trade-off surface although coverage across the identified fronts was good.

The baseline MOEA achieved reasonable proximity to the global front on ZDT-5. Performance on this deceptive test function is depicted in Figure 3.8. Note that on no occasions was the algorithm able to identify the extreme right-hand section of the discrete trade-off

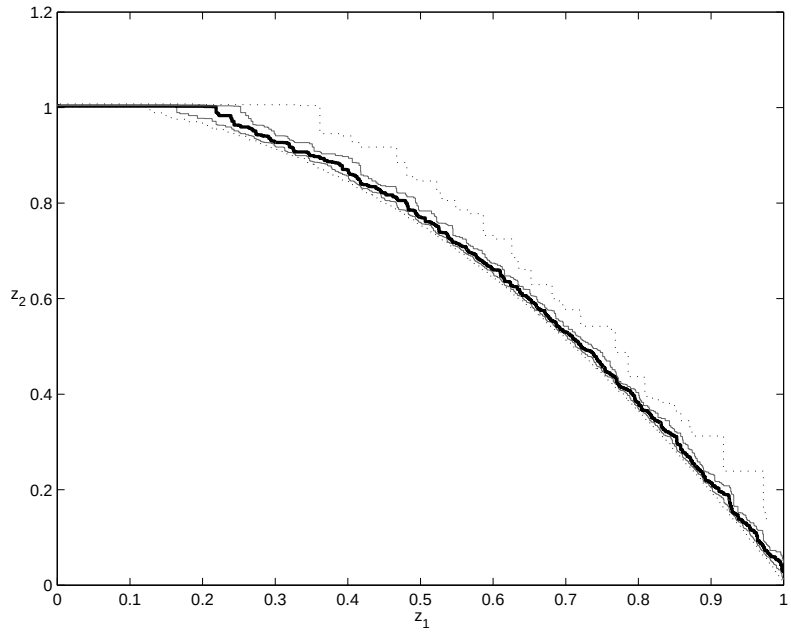


Figure 3.5: Attainment surface — baseline MOEA solving ZDT-2

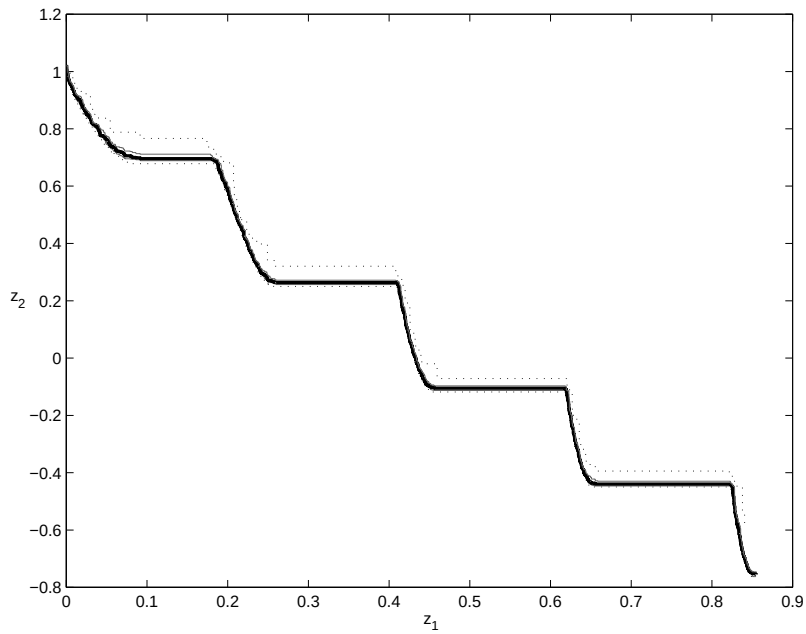


Figure 3.6: Attainment surface — baseline MOEA solving ZDT-3

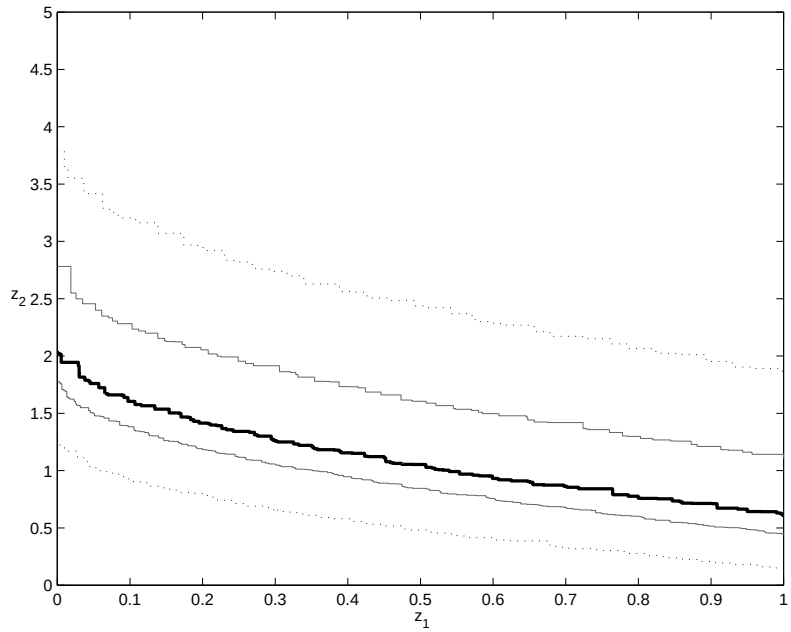


Figure 3.7: Attainment surface — baseline MOEA solving ZDT-4

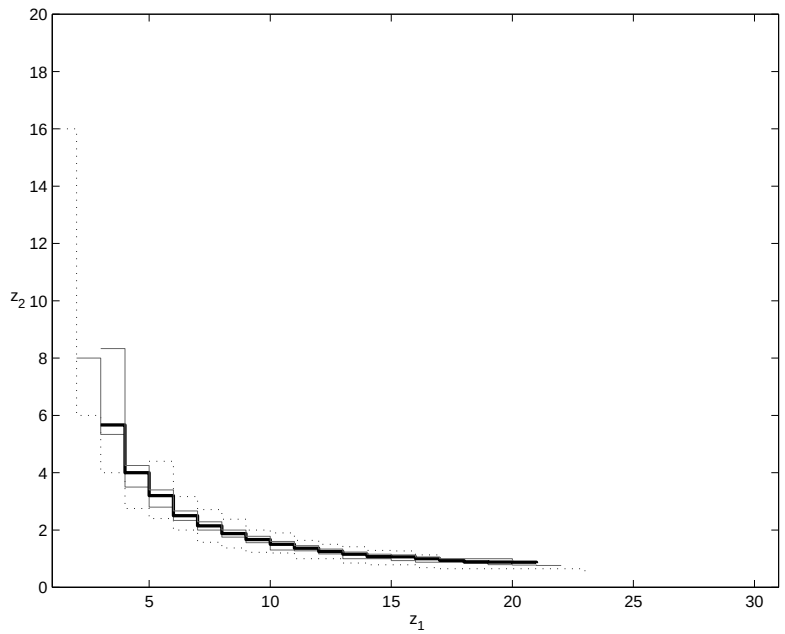


Figure 3.8: Attainment surface — baseline MOEA solving ZDT-5

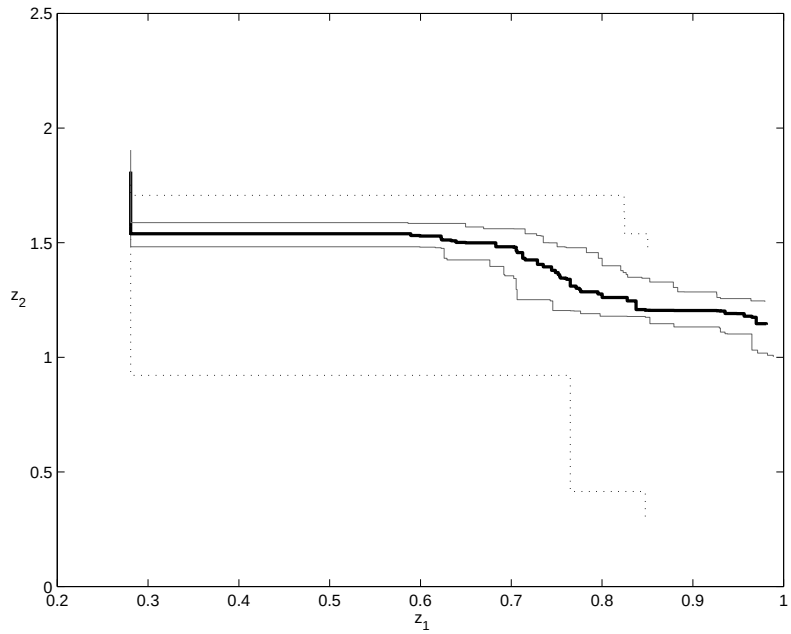


Figure 3.9: Attainment surface — baseline MOEA solving ZDT-6

surface.

Rather poor performance was observed on the non-uniform ZDT-6, as shown in Figure 3.9. Coverage is especially poor on the less naturally dense area of the front. This, together with the missing part of the ZDT-5 front, is the strongest indication that an explicit diversity enhancement mechanism would be beneficial. Proximity to the true Pareto front is also not good: only the 0%-attainment surface lies on the global front, where coverage is particularly poor. Furthermore, the position of this front with respect to the median and quartiles suggests that this result is something of an outlier.

The chapter now progresses to consider the effects of an alternative scheme for proximity promotion, the introduction of methods for diversity promotion, and a contemporary s_s scheme that incorporates elements to encourage improved proximity *and* diversity, using the baseline algorithm explored in this section.

3.4 Obtaining Good Proximity: The Ranking Strategy

3.4.1 Introduction

As explained in Section 2.4.3, many multi-objective evolutionary algorithms that are based on the concept of Pareto dominance use a derivative of one of the following multi-objective ranking procedures:

- *Dominance depth* (Goldberg 1989)
- *Dominance rank* (Fonseca and Fleming 1993)

Despite some comparative analysis in the literature, there remains much uncertainty over the relative worth of the two methodologies. In this section the empirical performance of both procedures, in the context of the baseline MOEA, is obtained for the Zitzler *et al.* (2000) test suite. Previous published comparisons have been between algorithm *brands*, in which it is difficult to decide on exactly what is responsible for the observed discrepancies in performance. By focusing solely on the ranking method, it is hoped that clearer evidence will be produced. The discussion herein is based on pure Pareto dominance, but it should be noted that it is equally applicable to other dominance measures such as *preferability* (Fonseca and Fleming 1998a).

3.4.2 Preliminary Analysis

In a review of the EMO research field, Veldhuizen and Lamont (2000) argue that there is no clear evidence to favour either ranking method overall. Dominance rank is generally regarded as the more efficient method (Coello 1999, Veldhuizen and Lamont 2000) and has been suggested to be easier to analyse (Fonseca and Fleming 1997). Dominance rank has also been found to be the simpler method to extend (Hughes 2001). In the only direct empirical comparison of the two schemes, in the context of a single real-world problem, dominance rank was shown to provide a more accurate trade-off surface (Thomas 1998).

In essence, dominance rank provides a more fine-grained ranking than dominance depth (Horn 1997). However, it is arguable whether or not this is a definite benefit. The dominance rank of a solution describes how many other solutions in the population are preferable to itself, whereas dominance depth provides only a minimum number. Thus, current population

density has more impact in the dominance rank scheme. This led Deb (2001a) to suggest that dominance rank may be sensitive to the shape of the Pareto front and to the density of solutions in the search space.

Both methods meet the fundamental aims of a multi-objective ranking strategy: (i) that all preferred individuals are assigned the same rank, and (ii) that all individuals are ranked higher than those that they are preferable to. Note that both methods produce identical rankings for a single-objective problem.

3.4.3 Evaluation

Randomisation testing results for the two ranking methodologies when integrated within the selection-for-variation stage of the baseline MOEA are displayed in Figure 3.10. Observed differences to the left of the randomisation distribution favour the dominance depth technique.

No significant evidence was found on any of the test problems for either performance indicator to suggest that one of the ranking schemes was superior to the other. However, it should be noted that the remainder of the baseline selection-for-variation algorithm is that which was originally used with dominance rank (Fonseca and Fleming 1993), so it is possible that there may be implicit bias toward this ranking procedure.

3.5 Promoting Diversity: The Sharing Strategy

3.5.1 Introduction

As explained in Chapter 2 and reinforced in Section 3.2.3, a multi-objective evolutionary algorithm is required to produce a suitable distribution of candidate solutions for presentation to the decision-maker. The baseline MOEA described in Section 3.3 does not include any special diversity-promoting mechanisms. Thus, it is a matter of some interest (and largely unvalidated to-date) whether or not the incorporation of explicit diversity-promotion techniques actually produces approximation sets with a superior diversity. In this section, this matter is explored for the classic fitness sharing methodology and a variant on two contemporary approaches. It is assumed throughout that the DM requires a good distribution in objective-space and has no interest in the underlying solution distribution.

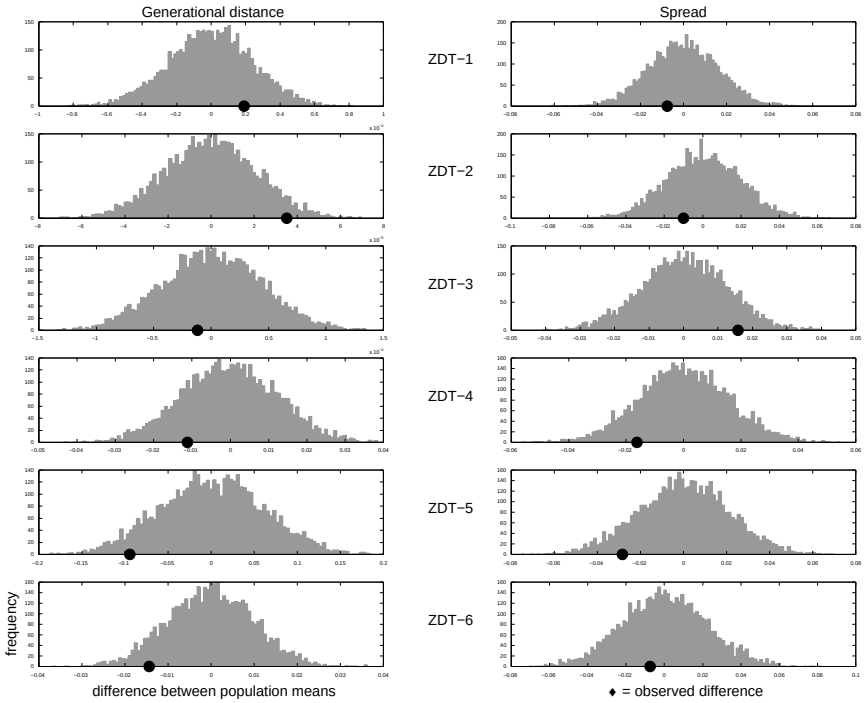


Figure 3.10: Randomisation test: dominance depth MOEA versus dominance rank MOEA (baseline)

3.5.2 Epanechnikov Fitness Sharing

Further to the brief review of the technique in Section 2.3.3, fitness sharing was originally proposed by Goldberg and Richardson (1987) to mitigate the problem of *genetic drift* in single-objective optimisation. It was then proposed as a possible method for encouraging niche formation across a multi-objective trade-off surface (Goldberg 1989). The main issue with the technique is the criticality of the niche size, σ_{share}, setting for the achievement of good performance. Appropriate values for σ_{share} can be difficult to obtain in practice.

Fonseca and Fleming (1995a) noted the similarity between the power law sharing function, previously given in Equation 2.3, and the Epanechnikov kernel density estimator K_e used by statisticians and described in Equation 3.3, where M is the dimension of the data (or the number of objectives for estimation in objective-space), ζ_M is the volume of the unit M-dimensional hypersphere, h is a smoothing parameter, and $\frac{d}{h}$ is the normalised Euclidean distance between individual vectors.

$$K_e\left(\frac{d}{h}\right) = \begin{cases} \frac{1}{2}\zeta_M^{-1}\left(M+2\right)\left(1-\left(\frac{d}{h}\right)^{\alpha}\right) & \text{if } \frac{d}{h} < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

The kernel smoothing parameter used in the estimator was shown by Fonseca and Fleming (1995a) to be directly analogous to the fitness sharing niche size parameter. The key benefit of this insight is that statisticians have developed successful techniques for estimating the value of h (and, thus correspondingly, σ_{share}) (Silverman 1986). An example approach, which is approximately optimal in a mean-squared error sense if the data follows a multivariate normal distribution, is given in Equation 3.4 in which N is the number of samples (objective vectors). Fonseca and Fleming (1995a) explain that a population with an arbitrary sample covariance matrix $\boldsymbol{T}$ can be transformed to a form suitable for this estimator by multiplying by a matrix $\boldsymbol{R}$, such that $\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}} = \boldsymbol{T}^{-1}$. This approach is amenable to update at each iteration of the evolutionary optimiser and can be regarded as fitness sharing with automatic tuning of σ_{share}. Note that absolute fitness values are modified using Equation 2.4 as usual.

$$h = \left(8\zeta_M^{-1}\left(M+4\right)\left(2\sqrt{\pi}\right)^{M}/N\right)^{\frac{1}{M+4}} \tag{3.4}$$

Epanechnikov fitness sharing has been incorporated within the selection-for-variation mechanism of the baseline MOEA and has been applied to the benchmark problems. Sharing is performed using the Euclidean distance metric in objective-space. The results of a randomisation comparison with the baseline algorithm are shown in Figure 3.11. Observed values that favour the sharing scheme will lie to the left of the randomisation distribution.

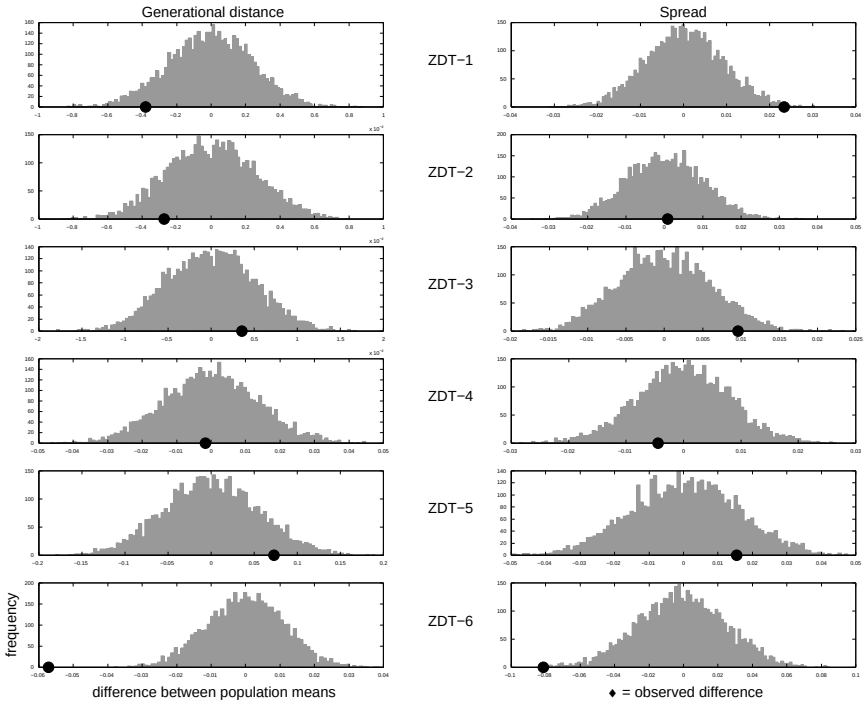


Figure 3.11: Randomisation test: Epanechnikov versus baseline

The inclusion of Epanechnikov sharing has improved both aspects of performance on ZDT-6. The non-uniform nature of this problem should particularly highlight the benefits of such a sharing scheme. Note in particular that a method designed to improve diversity has also helped to improve proximity, thus suggesting the strong interaction between the two optimiser requirements. However, no improvements in either diversity or proximity have been achieved for any other test function. Indeed there is some evidence to suggest a deterioration in diversity on ZDT-1, although this is not significant at the 1%-level.

3.5.3 Rank-Based Methods

Discussion

The difficulty and inconvenience involved in determining the niche size value has encouraged many researchers to investigate alternative methods for diversity promotion. A review of techniques has been provided in Section 2.4.4. Deb, Pratap, Agarwal and Meyarivan (2002) used solution density estimates as a secondary discriminator in binary tournament selection-for-variation when the two candidate solutions were of equal dominance depth. Zitzler, Laumanns and Thiele (2001) added the inverse density estimate to the raw strength value in SPEA2. The maximum amount of fitness that can be added is limited to ensure that the resulting fitness hierarchy respects the original inter-equivalence-class, dominance-based results. The strength is then used in binary tournament selection-for-variation. The density enhancement mechanisms in these two schemes are identical. All performance differences are completely attributable to the actual underlying process used for density estimation. In the NSGA-II, this is the *crowded distance*, whilst in SPEA2 it is kth NN. An appropriate value for k is obtained via statistical heuristics (Silverman 1986).

In this section, the above mechanism is generalised as a methodology that increases the resolution of an existing ranking obtained through Pareto considerations. The primary difference between this approach and fitness sharing is that the density estimate is only used for ranking purposes rather than to make an absolute modification to fitness. This may help to reduce the sensitivity of the technique to the actual density estimate computed.

In terms of the actual density estimators note that there is no difference, in terms of the number of free parameters, between Deb, Pratap, Agarwal and Meyarivan's (2002) and Zitzler, Laumanns and Thiele's (2001) schemes and fitness sharing: a value of k in the kth NN is effectively required in both of the former estimators, whilst σ_{share} is needed in the power-law sharing equation. However, the NN approach exploits relative distances in the current population distribution whilst the sharing approach requires an absolute distance to be ascertained. This self-adaptation characteristic can also be hypothesised to benefit performance. The rank-based diversity enhancement scheme is implemented within the baseline MOEA as described below.

In Fonseca and Fleming's (1998a) approach, a candidate solution is ranked according to how many other solutions in the current population are preferred to it. Given the minimum

amount of preference information (a direction of monotonically increasing preference in each objective), the comparison is made in terms of pure Pareto dominance. This concept is illustrated in Figure 3.12. In this simple example, both objectives are to be minimised. Objective vectors for five candidate solutions $\{A, \ldots, E\}$ are shown.

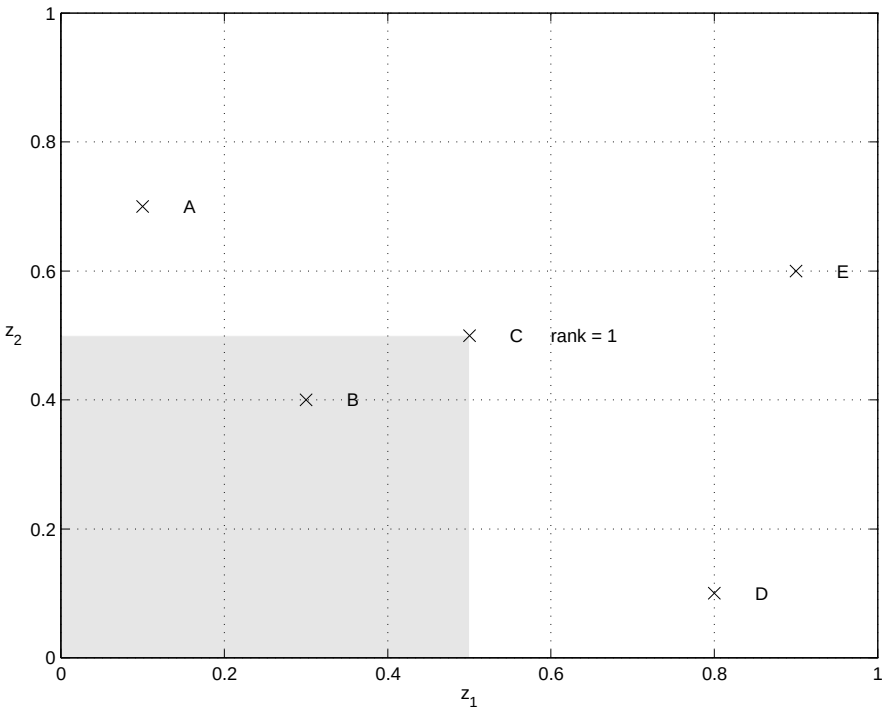


Figure 3.12: Example of ranking based on the dominance rank approach

A solution is dominated by all other solutions within the hypercube defined by its own objective vector and the Utopian point (in this case $\{0, 0\}$). In this example, the domination hypercube for solution C is indicated by the grey rectangle. Solution C is seen to be dominated by solution B alone, and thus receives a multi-objective ranking of 1. The rankings for all solutions are shown in Table 3.3.

The original multi-objective genetic algorithm (Fonseca and Fleming 1993) uses the stochastic universal sampling selection-for-variation mechanism because of its low stochastic error properties (Baker 1987). This technique requires a mapping between ranking and fitness value (in contrast to tournament selection, where this is implicit to the tournament size). This is achieved by sorting the population according to rank, assigning fitness according to

Table 3.3: Multi-objective ranking for the solutions in Figure 3.12

Candidate	Objectives $\{z_1, z_2\}$	Rank, ρ
A	$\{0.1, 0.7\}$	0
B	$\{0.3, 0.4\}$	0
C	$\{0.5, 0.5\}$	1
D	$\{0.8, 0.1\}$	0
E	$\{0.9, 0.6\}$	3

some function, and then averaging the fitnesses for solutions of the same rank. This process is illustrated in Figure 3.13. The narrower bars show the pre-averaged fitness values, whilst the wider bars indicate the post-averaged fitnesses.

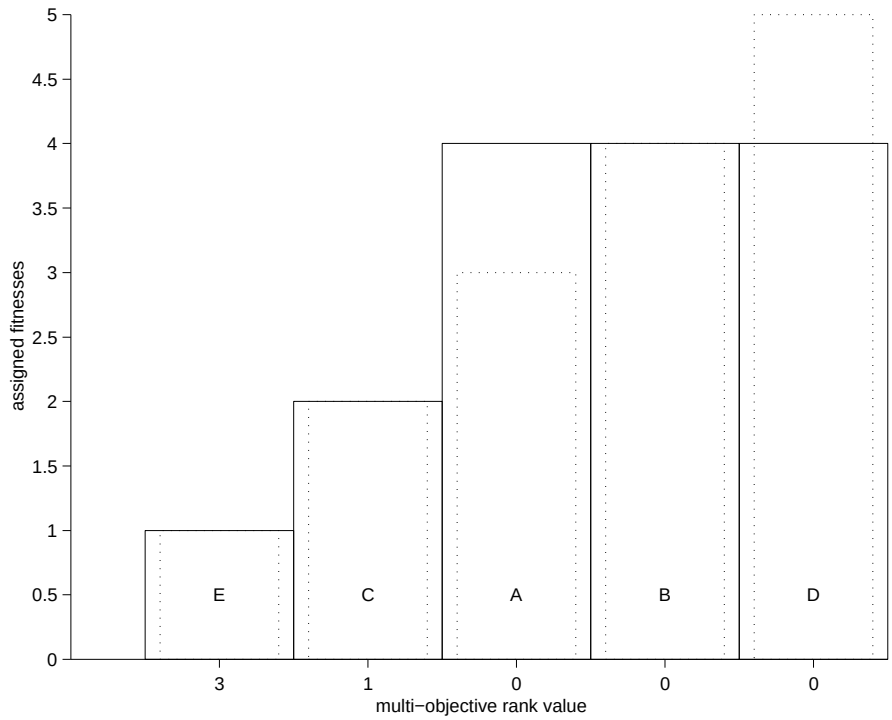


Figure 3.13: Rank-to-fitness assignment procedure

The functional mapping between the partially-ordered list and fitness is often either linear or exponential, although other forms are possible. It generally includes a parameter related to selective pressure that can be used to vary position on the exploration-exploitation trade-off (although this may be implicit to the procedure). A linear mapping is used in this chapter,

where fitness is assigned in unit increments to a sorting of the population based on the obtained multi-objective ranking. This process is shown in Equation 3.5, where f is fitness, ρ is the index in the sorting of the population (with position zero corresponding to a solution with a multi-objective rank of zero), and N is the number of candidate solutions.

$$f(\rho) = N - \rho \tag{3.5}$$

The results in Figure 3.13 have been computed using Equation 3.5, where $N = 5$. Note that the selective pressure of this scheme is implicit to the number of solutions in the population. For standard population sizes, the selective pressure is close to 2. Note that, in any linear scheme, the selective pressure is always bounded between 1 and 2.

The niching approach presented here increases the resolution of the above ranking procedure through the inclusion of population density information. An *intra-ranking* is performed on candidate solutions of identical dominance-based rank, discriminating on the basis of population density. Solutions in less dense areas receive a superior intra-ranking to their counterparts in denser regions. The distance metric is likely to be problem dependent and could conceivably contain decision-maker preference information. Following the new fine-grained ranking, the fitness assignment procedure remains unchanged. This new diversity preserving measure is illustrated in Figure 3.14.

The density measure selected is the Euclidean distance to the first-nearest neighbour in objective-space (see Figure 3.12). Solutions A, B, and D all have the same Pareto rank, but solution D is the remotest and thus receives the highest fitness. Solutions A and B are an identical distance apart and thus share the next two available fitnesses equally. Note that A and B still receive a higher fitness than the solutions that they dominate (C and E). The associated fine-grained ranks are shown in Table 3.4, to the right of the original coarse-grained equivalents.

The proposed new scheme has a number of important properties, namely:

- If one candidate solution is preferred over another, then the former is guaranteed to have a superior fitness value. This was not the case under the original fitness sharing scheme of Fonseca and Fleming (1993).

- The cumulative fitness assigned to each ranking group remains unchanged.

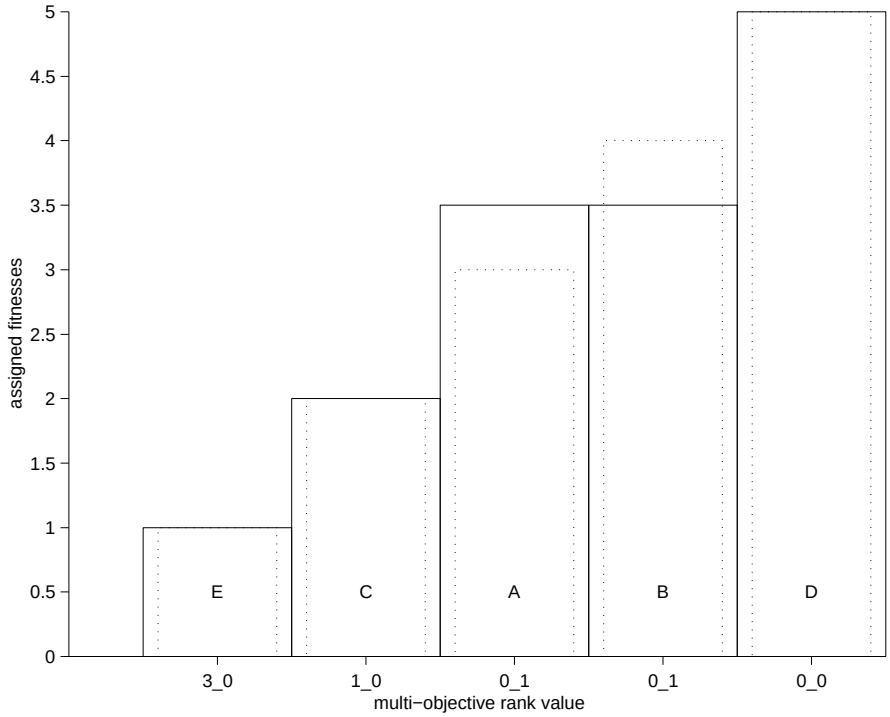


Figure 3.14: Fitness based on multi-objective ranking followed by diversity intra-ranking

Table 3.4: Multi-objective and diversity-based ranking for Figure 3.14

Candidate	Distance to nearest neighbour	Coarse rank	Fine rank
A	0.361	0	0_1
B	0.361	0	0_1
C	∞	1	1_0
D	0.583	0	0_0
E	∞	3	3_0

- When all solutions are currently non-dominated, discrimination is based purely on density.
- When all solutions are currently non-dominated and the population density measure is globally uniform, all fitness values are identical.

With any type of ranking scheme, information content is lost. The ranking indicates that one solution lies in a more densely packed region than another solution but the actual difference in density between the two is lost. This limits the amount of information available to the search procedure but protects against premature convergence to locally *superfit* solutions and reduces the sensitivity to any free parameters of the density estimator.

Results

The results of randomisation testing, when comparing the contemporary diversity enhancement method to the baseline (non-sharing) system are shown in Figure 3.15. Observed differences between sample means to the left of the randomisation distribution provide evidence in favour of the new scheme.

The central aim of sharing is to improve the distribution of solutions in objective-space and this should be primarily evident in the spread results. As shown in Figure 3.15, there is strong evidence to suggest that the new method improved spread on ZDT-3 and ZDT-4. The use of the Epanechnikov kernel, by contrast, did not improve results on these problems. In no cases, was the absence of the contemporary sharing mechanism shown to be preferable (whereas there was some evidence in Section 3.5.2 to suggest that the Epanechnikov kernel could cause a deterioration in diversity). However, there is no evidence to suggest that the use of sharing made any difference to the results for ZDT-6. This is particularly surprising because of the naturally non-uniform distribution across the trade-off surface for this problem — a situation in which the positive effect of explicit diversity enhancement should be most evident. In terms of closeness of the approximation set to the global Pareto front, the contemporary mechanism was found to be associated with improved proximity on ZDT-1, 2, and 4.

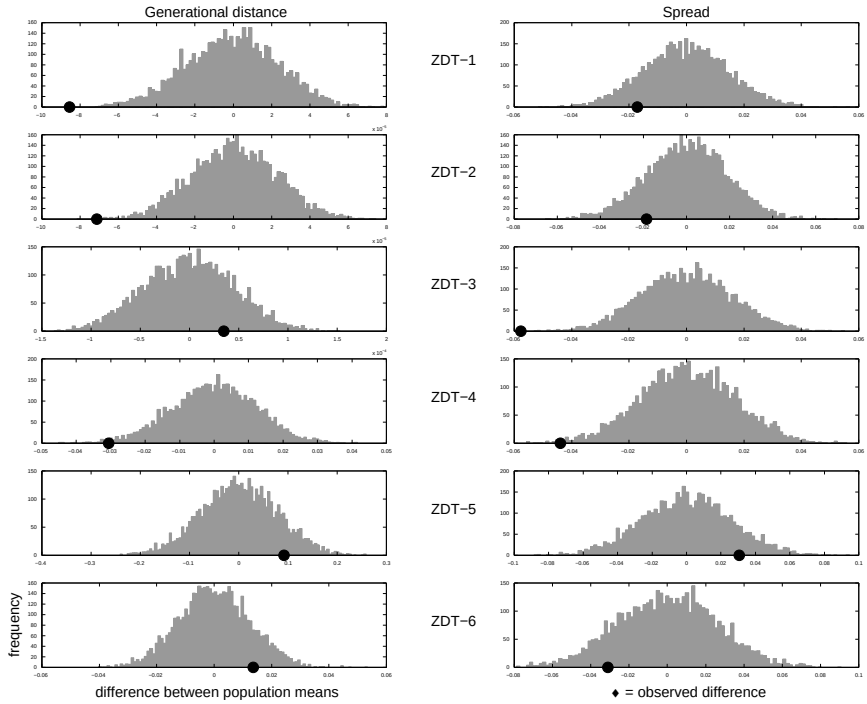


Figure 3.15: Randomisation test: Rank-based sharing versus no sharing

3.6 Obtaining Good Proximity *and* Good Diversity: an Elitist Strategy

Elitism is the process of actively preserving previous high-performance solutions during the selection-for-survival stage of an MOEA. Elitism has long been considered an effective method for improving the efficiency (in terms of convergence speed to a good approximate solution) of an EA (De Jong 1975). Various studies in the EMO community have indicated that the inclusion of an elitist element can considerably improve the performance of an MOEA (Zitzler *et al.* 2000, Deb, Pratap, Agarwal and Meyarivan 2002). In particular, the success of the SPEA algorithm (which contains an elitist element) across a diverse set of bi-objective problems has led to the widespread adoption of elitist schemes in the EMO community (Zitzler *et al.* 2000). Refer to Section 2.4 for a review of elitist strategies in MOEAs.

This study will consider the effect of the elitism technique proposed by Zitzler and Thiele (1999) which was implemented in the SPEA. The method is based on the generational gap approach to selection-for-survival, where a clustering technique is used to maintain a representative subset of good solutions in $P[t]$ to be preserved. Generational gap techniques are common in the genetic algorithm community, from which the baseline MOEA used in this study has its origins. For this reason, and because of the impressive results reported in the literature for SPEA, it has been chosen for analysis herein.

The SPEA (conceptually) maintains an on-line archive of currently non-dominated solutions and uses this in the processes that generate new candidate solutions. The archive should be a representative subset of all the non-dominated solutions found thus far. Note that MOEAs developed prior to SPEA generally maintain an up-to-date off-line archive of *all* non-dominated solutions found, but these results are not explicitly used in the generation of new candidate solutions.

The on-line archive requires a clustering mechanism in order to control the number of elite solutions. This set of solutions should represent the characteristics of the underlying off-line archive: generally referring to the objective vectors, although decision-space discrimination is also possible. The *truncation procedure* described in Zitzler, Laumanns and Thiele (2001) is an effective means of elitism control for bi-objective problems. In these cases, this method can reduce an over-sized archive without losing boundary solutions. This attribute is desirable

in the search for diverse trade-off solutions. However, for problems with more than two objectives, extreme trade-off solutions can be lost by this method. It is also possible for the procedure to remove globally non-dominated solutions whilst retaining currently non-dominated, yet sub-optimal, solutions (both solutions would be non-dominated from the perspective of the truncation process, but the truly sub-optimal solution would be in a less-dense area of objective-space) (Laumanns, Thiele, Deb and Zitzler 2002). The truncation procedure is outlined below:

To remove one member of the over-sized archive follow the subsequent procedure with k initialised to 1:

1. Find the set of solutions, $\boldsymbol{X_S}$, with the shortest (objective-space) distance between themselves and their kth nearest neighbours.

2. If the size of this set is greater than one, increment k and repeat Step 1 using only solutions in $\boldsymbol{X_S}$, otherwise select the individual in $\boldsymbol{X_S}$ for removal.

If nearest neighbour information is exhausted, select randomly from the current set $\boldsymbol{X_S}$. The above procedure should be repeated until the archive has been reduced to an acceptable size.

The elitist strategy adopted in this study is a variant on the *universal elitism* approach developed by Zitzler (1999) and is illustrated by the schematic in Figure 3.16. The key difference is that the archive size is allowed to vary within pre-defined limits, whilst the number of newly generated candidate solutions is varied such that the total population size (elites plus new solutions) is held constant.

The on-line archive is initialised to the empty set, whilst the initial population is initialised to a random set of candidate solutions (possibly seeded with information provided by the decision-maker). The populations at subsequent iterations of the algorithm are the combination of new solutions and current elite solutions. The currently non-dominated solutions in the population are identified and are stored as the new, potentially over-sized, archive. Over-represented solutions are then eliminated from the archive, if necessary, using the truncation procedure defined above. For the test problems used in this study, the neighbourhood distance measure is defined as the Euclidean distance between two objective vectors.

When the new elite set has been finalised, the size of this set is known, and thus the number

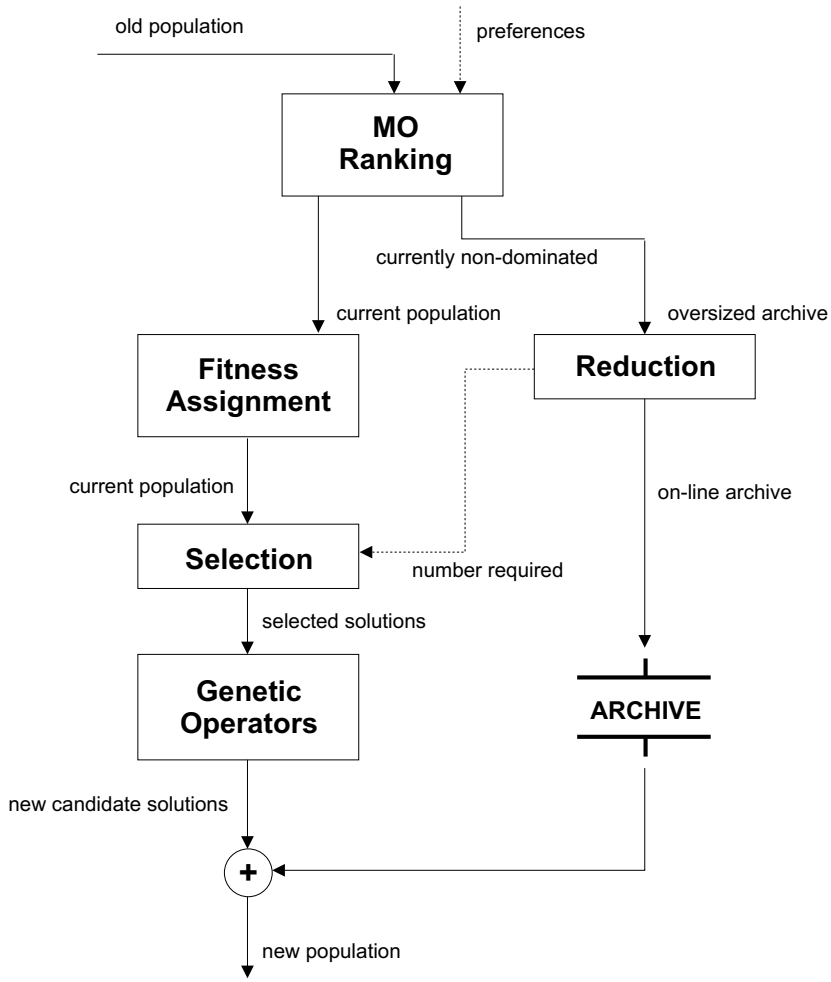


Figure 3.16: Elitist strategy employed in this study

of new candidate solutions required to fill the population can be calculated. These solutions are created through the selection and genetic manipulation of members of the current population. The new solutions are then combined with the elite set to form the subsequent total population, which completely replaces the old population.

This elitist strategy has been integrated within the baseline MOEA described in the previous section and has been applied to the six benchmark problems. The results of the randomisation testing between the elitist algorithm and the baseline algorithm are shown in Figure 3.17. Observed differences to the left of the randomisation distribution offer evidence in favour of the elitist version outperforming the baseline case.

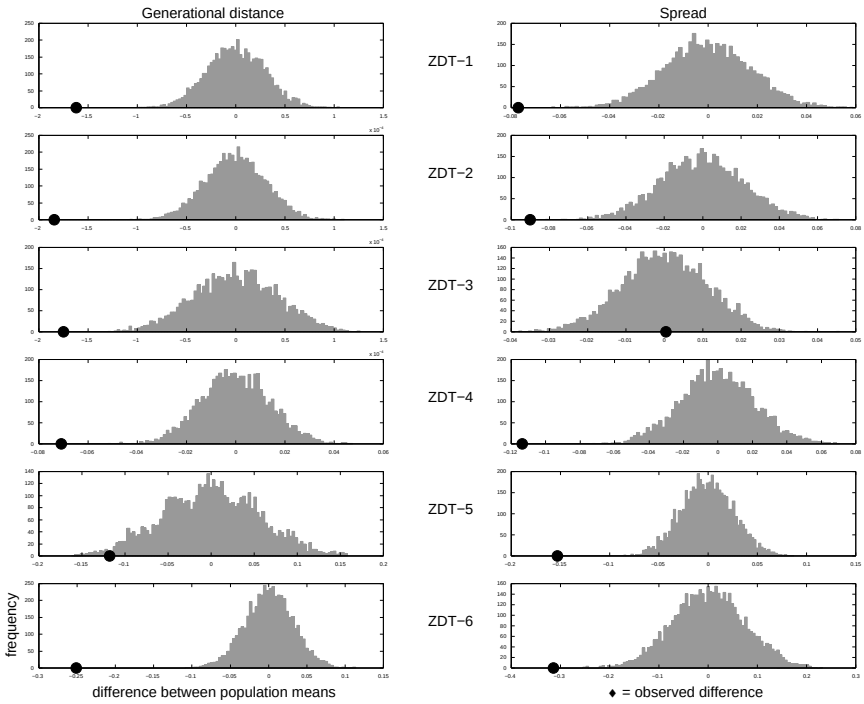


Figure 3.17: Randomisation test: elitist versus baseline

There is considerable evidence, clearly shown by the results in Figure 3.17, that the elitist algorithm produces results closer to the true front than the baseline for ZDT-1, 2, 3, 4, and 6. The observed result for ZDT-5 is not significant at the 1%-level, although it would have been significant at the 5%-level. Superior performance in terms of diversity is strongly suggested for ZDT-1, 2, 4, 5, and 6.

The inclusion of elitism increases the convergence speed of the algorithm. The danger of sub-optimal convergence is somewhat reconciled by the distributed nature of the elite set. Variation operators with the potential for a high degree of exploration, such as the Gaussian mutation operator used in this work, can also reduce the risk of premature convergence. Hence, the increased successful convergence exhibited in this study could perhaps be expected. The elitism scheme also maintains the characteristics of the currently identified trade-off surface within the on-line population. Thus, diversity of non-dominated solutions in the population is maintained and encouraged (through the thinning of similar objective vectors) by the truncation mechanism. This helps to explain the improvement in diversity seen in the results. However, the truncation process only represents the current distribution: it does not, directly, promote improved diversity through multiple selection-for-variations of remote solutions. Despite this, the inclusion of elitism did lead to improved diversity on the non-uniformly distributed ZDT-6. The extra inclusion of diversity-promotion mechanisms during selection-for-variation (as previously detailed in Section 3.5) may assist further in improving diversity across the trade-off surface. The dual use of diversity enhancement mechanisms for s_s and s_v receives further consideration in the next section.

3.7 State-of-the-Art MOEA: Enhanced s_s and s_v Mechanisms

The use of an elitist strategy or a sharing strategy in isolation has been shown to offer improved performance in terms of both proximity and diversity. It is instructive to now consider the effect of these schemes in combination. A schematic of the resulting algorithm is shown in Figure 3.18.

This optimiser has been applied to the problems in the ZDT test suite. The resulting attainment surfaces are shown in Figures 3.19 through 3.24.

The envelopes of attainment are generally very tight, indicating good consistency in the results. As evident from Figure 3.22, proximity has been greatly improved on ZDT-4: indeed the 25%-attainment surface lies very close to the global front of this difficult test problem. Complete coverage of the right-hand section of the trade-off surface has been achieved for ZDT-5, as shown in Figure 3.23. Finally, proximity and diversity have been much improved on ZDT-6 (see Figure 3.24).

A comparison with the baseline algorithm is made, via randomisation testing, in Fig-

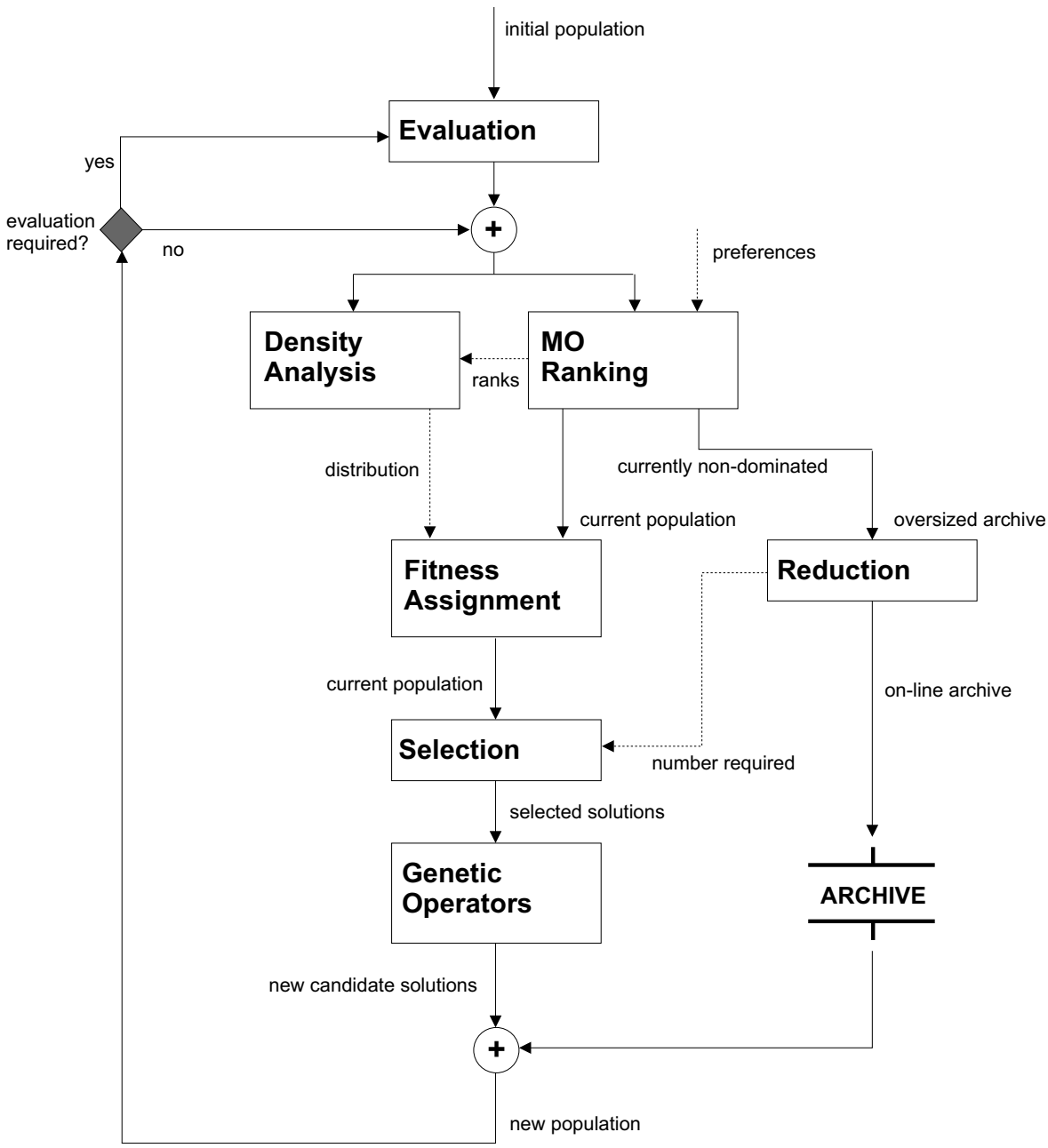


Figure 3.18: Elitist sharing MOEA schematic

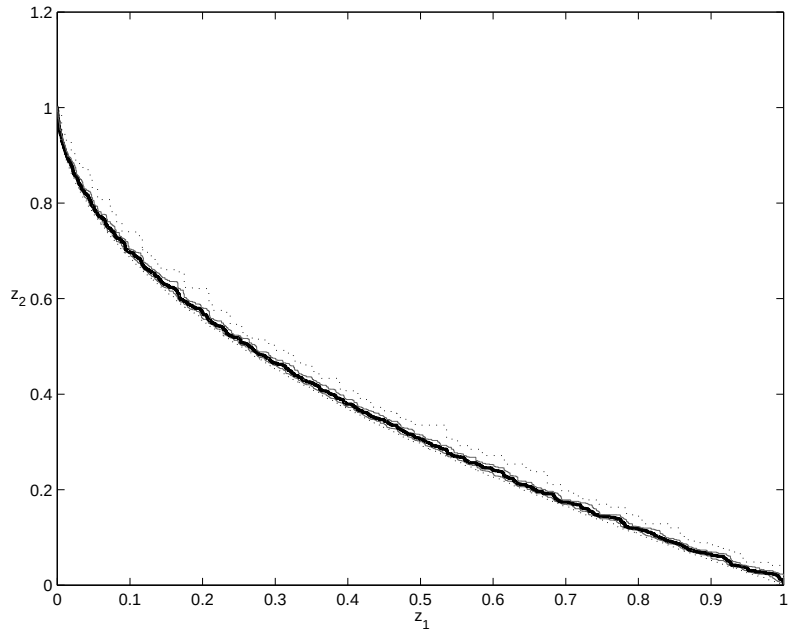


Figure 3.19: Attainment surface — contemporary MOEA solving ZDT-1

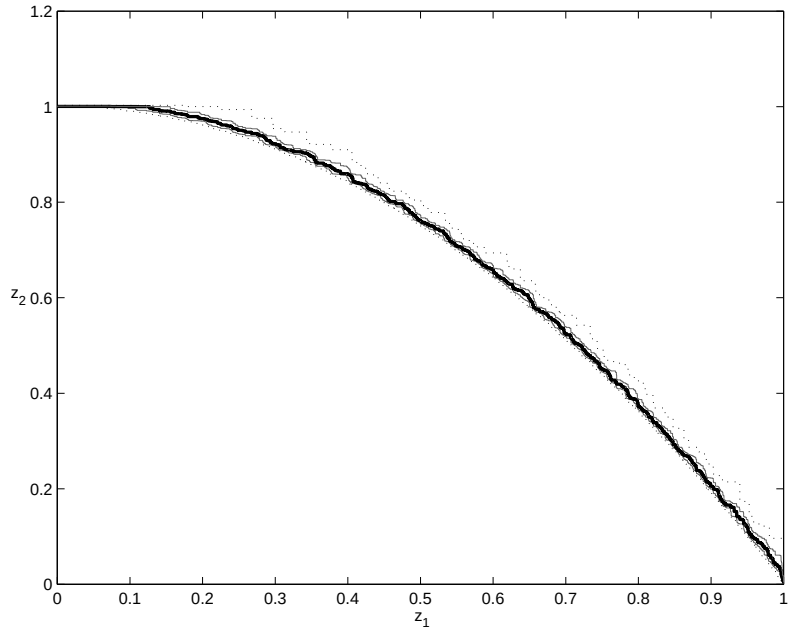


Figure 3.20: Attainment surface — contemporary MOEA solving ZDT-2

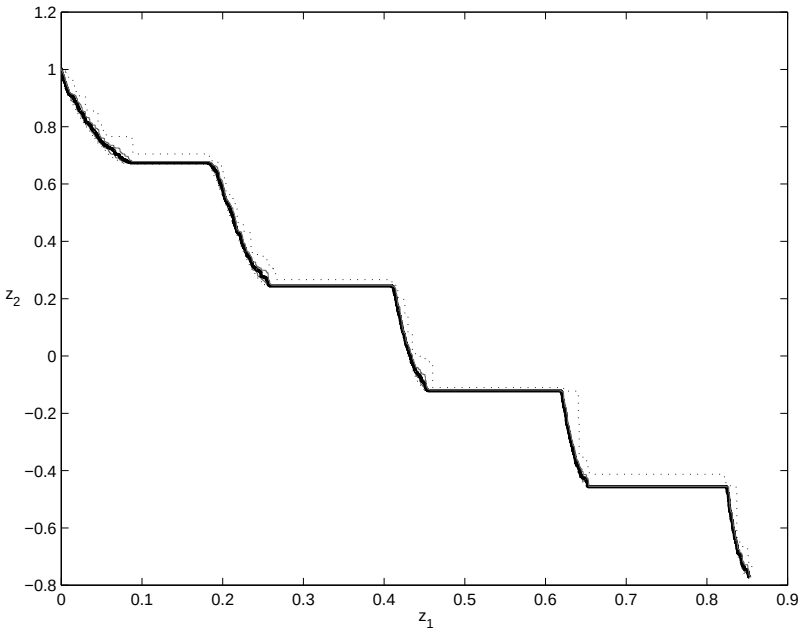


Figure 3.21: Attainment surface — contemporary MOEA solving ZDT-3

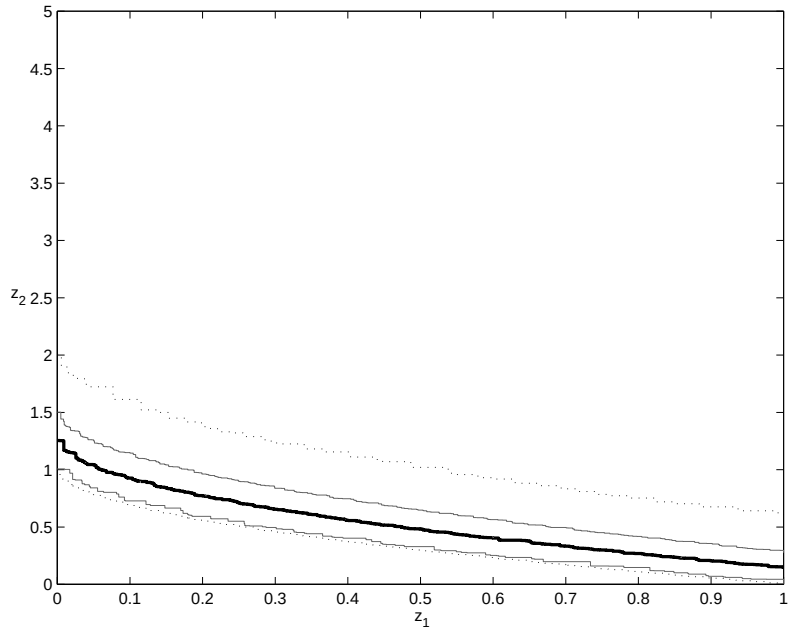


Figure 3.22: Attainment surface — contemporary MOEA solving ZDT-4

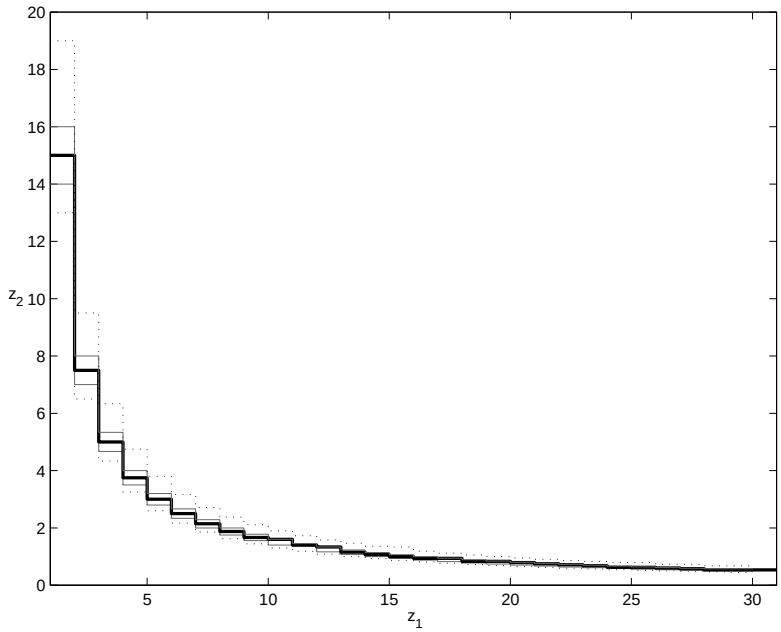


Figure 3.23: Attainment surface — contemporary MOEA solving ZDT-5

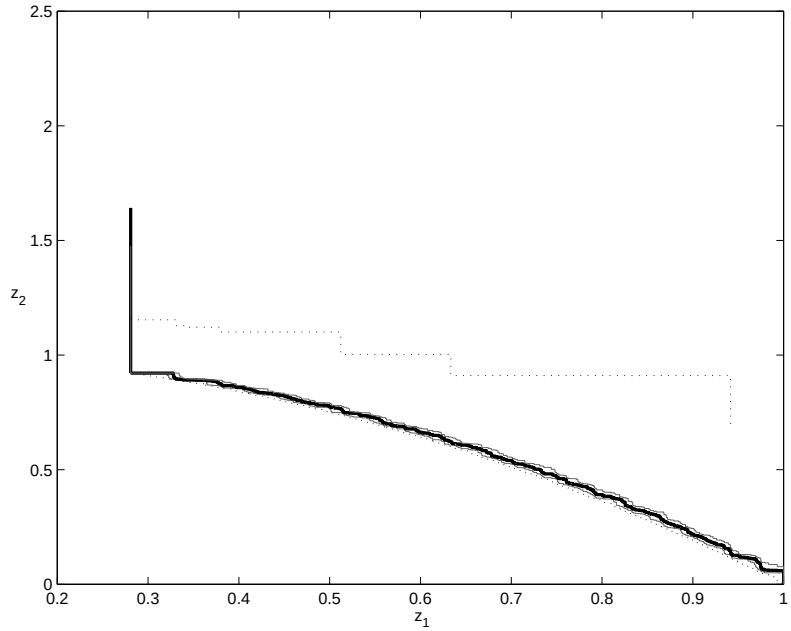


Figure 3.24: Attainment surface — contemporary MOEA solving ZDT-6

ure 3.25. Observed differences between the means of each indicator that lie to the left of the randomisation distribution favour the contemporary algorithm.

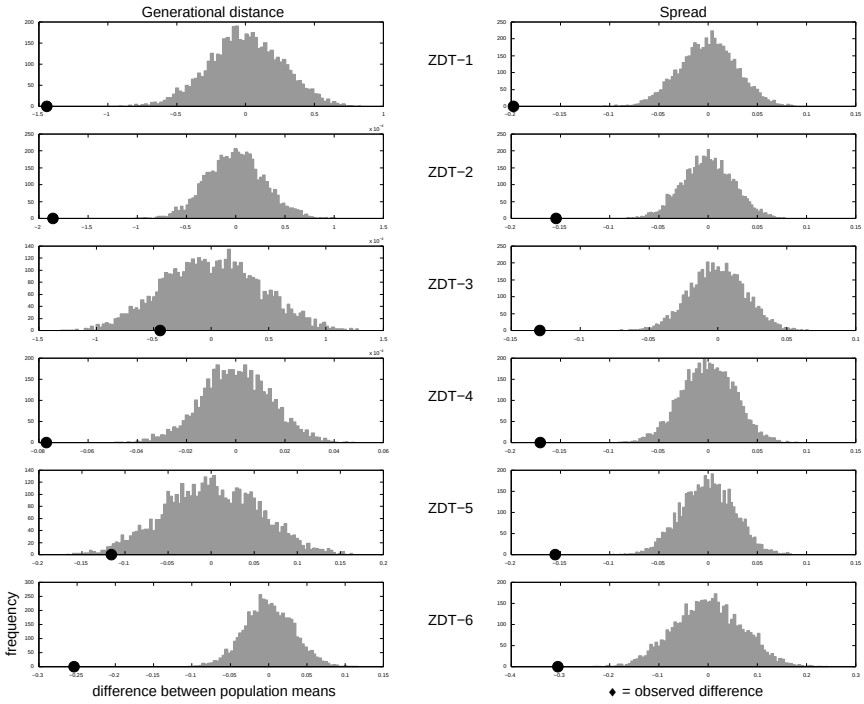


Figure 3.25: Randomisation test: contemporary MOEA versus baseline MOEA

Compelling evidence points to the new algorithm substantially outperforming the baseline in terms of diversity across all six benchmark problems. The *combination* of diversity enhancement schemes at both selection-for-variation and selection-for-survival was required in order to achieve this improvement. Neither scheme by itself was capable of producing this result. Interestingly, this benefit has been accompanied by a reduction in the improvement to the proximity result for ZDT-3 when compared to what can be achieved by the enhanced selection-for-survival mechanisms in isolation (described in Section 3.6). However, improved proximity was observed for ZDT-1, 2, 4, and 6 (the result for ZDT-5 is not significant at the 1%-level).

A direct comparison of the combined scheme with the MOEA from Section 3.6 is shown in Figure 3.26. Observed differences to the left of the randomisation distribution favour the combined scheme. There is substantial evidence that the incorporation of explicit diversity

promotion at the s_v stage has improved diversity still further on ZDT-1, 2, 3, and 4. As mentioned above, the improved diversity on ZDT-3 has been accompanied by attenuation of proximity. The reason for this is somewhat unclear.

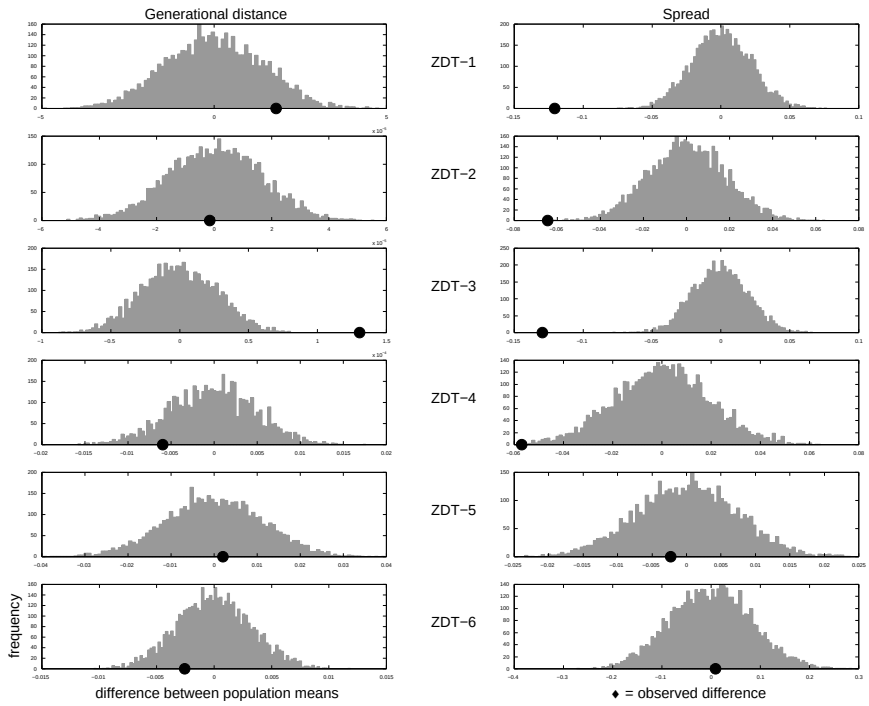


Figure 3.26: Randomisation test: contemporary MOEA versus elitist MOEA

3.8 Summary

In this chapter, a progressive and tractable experimental approach to the analysis of MOEA performance has been presented. This methodology stresses the importance of baselining as central to developing an understanding of the behaviour of various algorithmic components and the interactions between them.

The experimental framework has been used to explore some of the benefits of using advanced selection strategies in EMO. The results apply primarily to the class of real-parameter, bi-objective function optimisation problems. It has been demonstrated that performance is largely invariant under either of the main Pareto ranking methods in the EMO field.

Some evidence has been found to suggest that explicit diversity enhancement schemes during selection-for-variation can offer a degree of improved diversity in the resulting approximation sets. Much more obvious benefits were obtained for both diversity and proximity through the use of an elitist selection-for-survival scheme. When advanced s_s and s_v schemes were fused together, the improvements over the baseline were almost universal.

The fundamental limitation of these conclusions is that they — like the vast majority of such conclusions in the literature — only apply to bi-objective problems. But how do the extensions to the original EMO philosophy expounded by Goldberg (1989) perform when applied to optimisation tasks with greater than two objectives? Indeed, how does the original formulation perform? These are essential questions that face the theoretical EMO community. Timely answers are required since many EMO applications have already considered many objectives simultaneously. There is little theoretical support for this work at present.

In the next chapter, and the remainder of this thesis, the foundations are laid for research into the simultaneous optimisation of more than two objectives.

Chapter 4

Evolutionary Many-Objective Optimisation: An Introduction

4.1 Introduction

Theoretical evolutionary multi-objective optimisation studies, such as that detailed in Chapter 3, generally consider a small number of objectives or criteria. The bi-objective case is by far the most heavily studied. EMO applications, by contrast, are frequently more ambitious, with the number of treated objectives reaching double figures in some cases (Coello *et al.* 2002). Hence, there is a very clear need to develop an understanding of the effects of increasing numbers of objectives on EMO. The phrase *many-objective* has been suggested in the OR community to refer to optimisation problems with more than the standard two or three objectives (Farina and Amato 2002). Following this, the terminology *evolutionary many-objective optimisation* (E$\mathbb{M}$O) is proposed herein to refer to EMO problems of increased scale.

This chapter establishes a platform for research into E$\mathbb{M}$O via consideration of the different types of pair-wise relationships between the objectives. A classification of possible relationships is offered in Section 4.2, and the notation used to establish the concepts is introduced. *Conflict* between objectives is discussed in Section 4.3, whilst Section 4.4 considers *harmonious* objectives. The aim of a multi-objective evolutionary algorithm is generally regarded as to generate a sample-based representation of the Pareto optimal front, where the samples lie close to the true front and are well distributed across the front. The effects of

increasing numbers of each type of objective on both aspects of the quality of the trade-off surfaces produced are described, together with a review of methods for dealing with the difficulties that arise. The case where the global EMO problem can be decomposed into sub-problems is introduced in Section 4.5. Qualitative studies of pair-wise relationships between objectives are not uncommon in the EMO community, especially in the case of real-world applications. These are discussed in Section 4.6, alongside similar quantitative methodologies from the multi-criterion decision-making discipline.

Some of the concepts described in this chapter are illustrated using an example result from the contemporary multi-objective genetic algorithm (MOGA), developed in Chapter 3, solving the 3-objective DTLZ2 benchmark problem (Deb, Thiele, Laumanns and Zitzler 2002). This test function is described in Equation 4.1. Note that all objectives are to be minimised.

$$\left.\begin{array}{lrcl}
\text{min.} & z_1(\boldsymbol{x}) & = & [1+g(x_3,\ldots,x_{12})]\cos(x_1\pi/2)\cos(x_2\pi/2)\,, \\
\text{min.} & z_2(\boldsymbol{x}) & = & [1+g(x_3,\ldots,x_{12})]\cos(x_1\pi/2)\sin(x_2\pi/2)\,, \\
\text{min.} & z_3(\boldsymbol{x}) & = & [1+g(x_3,\ldots,x_{12})]\sin(x_1\pi/2)\,, \\
\text{w.r.t.} & \boldsymbol{x} & = & [x_1,\ldots,x_{12}]\,, \\
\text{where} & & & g(x_3,\ldots,x_{12}) = \sum_{i=3}^{12}(x_i-0.5)^2\,, \\
\text{and} & & & 0 \le x_i \le 1, \text{ for } i=1,\ldots,12.
\end{array}\right\} \tag{4.1}$$

4.2 Relationships Between Objectives

4.2.1 Classification

In theoretical EMO studies, the objectives are generally considered to be in some form of conflict with each other. Thus, in the bi-objective case, the optimal solution is a one-dimensional (parametrically speaking) trade-off surface upon which conflict is always observed between the two objectives. However, other relationships can exist between objectives and these may vary within the search environment. A basic classification of possible relationships is offered in Figure 4.1. These relationships are explained in the remainder of the chapter.

The dependency classifications are not necessarily mutually exclusive. For example, in the case of three conflicting objectives, there may be regions where two objectives can be improved simultaneously at the expense of the third. This is illustrated in Figure 4.2 for the final approximation set of a MOGA solving the 3-objective DTLZ2 problem (see Equation 4.1).

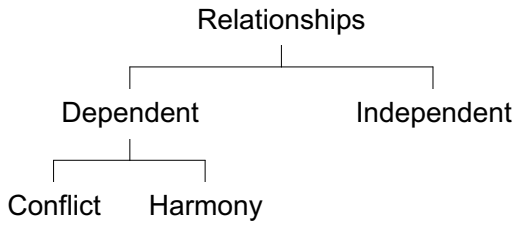


Figure 4.1: Classification of relationships between objectives

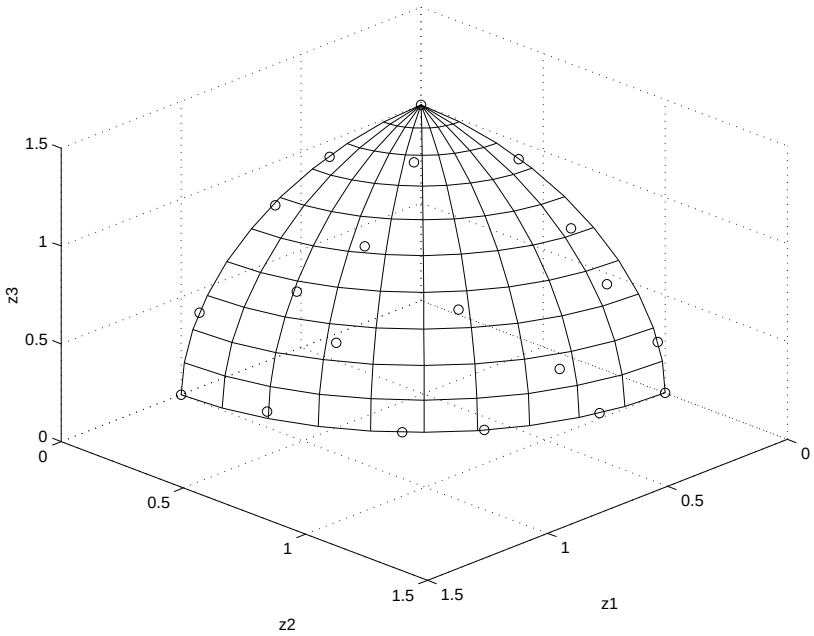


Figure 4.2: Final on-line archive of MOGA (depicted as circles) solving DTLZ2, superimposed on the global trade-off surface

For example, ideal performance in z_2 and z_3 (evidence of harmony) can be achieved at the expense of nadir performance in z_1 (evidence of conflict), as indicated by the left-most objective vector in the figure. However, on the far right of the figure, z_1 and z_3 are now in harmony and are both in conflict with z_2. Thus, the nature of the relationships change across the Pareto front.

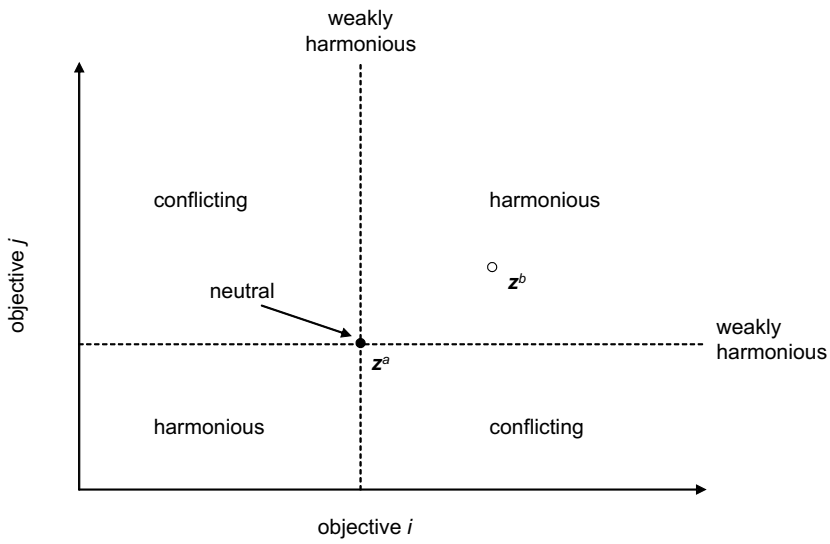


Figure 4.3: Dependency relationship regions between a pair of objectives, i and j, identified using the location of sample vector $\boldsymbol{z}^b$ relative to that of $\boldsymbol{z}^a$

4.2.2 Notation

The following notation is used in the remainder of the chapter: M is the number of objectives to be considered in the optimisation procedure, $\boldsymbol{Z}$ is the set of all realisable objective vectors $\boldsymbol{z} \in \mathbb{R}^M$, and $\boldsymbol{Z_R}$ is a particular region of interest in objective-space, $\boldsymbol{Z_R} \subseteq \boldsymbol{Z}$. If $\boldsymbol{Z_R} = \boldsymbol{Z}$ then the relationship is said to be *global*, otherwise it is described as *local*. The case $\boldsymbol{Z_R} = \boldsymbol{Z_*}$, where $\boldsymbol{Z_*}$ is the Pareto optimal region, may be of particular interest since these are typically the relationships that will be presented to the decision-maker.

Let i and j be indices to particular objectives: $i, j \in [1, \ldots, M]$. Let a and b be indices to individual objective vector instances: $a, b \in [1, \ldots, |\boldsymbol{Z_R}|]$. Also let (a, b) denote a pair of instances for which $a \neq b$. Minimisation is assumed throughout the discussions without loss of generality.

The dependency relationships that can be identified via pair-wise analysis are summarised in Figure 4.3. They are based on the position of objective vector $\boldsymbol{z}^b$ relative to the position of $\boldsymbol{z}^a$. These relationships are explored in more detail in Section 4.3 and Section 4.4 to follow.

4.3 Conflicting Objectives

4.3.1 Definitions of Conflict

A relationship in which performance in one objective is seen to deteriorate as performance in another is improved is described as *conflicting*. This is summarised by Definition 4.1 below and can be related to the $\boldsymbol{z}^b$-relative-to-$\boldsymbol{z}^a$ regions marked as conflicting in Figure 4.3.

Definition 4.1 (Conflict) *Objectives i and j exhibit evidence of conflict according to the condition $(\boldsymbol{z}_i^a < \boldsymbol{z}_i^b) \wedge (\boldsymbol{z}_j^a > \boldsymbol{z}_j^b)$. If $\neg\exists(a,b)$ for which the condition holds then there is* no conflict*, if $\exists(a,b)$ then there is* conflict*, whilst if the condition holds $\forall(a,b)$ then there is* total conflict*.*

Note that no attempt has been made to define intermediate levels of conflict (or harmony, as discussed in Section 4.4) since this requires DM preference information beyond that required for pure Pareto optimality.

4.3.2 Effect on EMO

For M conflicting objectives, an $(M-1)$-dimensional trade-off hypersurface exists in objective-space. The number of samples required to achieve an adequate representation of the surface is exponential in M. Given a finite population, an evolutionary optimiser will encounter intractable difficulties in representing the surface when large numbers of conflicting objectives are considered. Even if such a representation were possible, the value to the DM of such a large number of candidate solutions is questionable.

Deb (2001a) has shown that the proportion of locally non-dominated objective vectors in a finite randomly-generated sample becomes very large as the number of objectives increases. Similar results were reported in Fonseca and Fleming (1998b) for the final non-dominated set of a real-world MOEA application. Since dominance-based discrimination is used to determine the relative worth of current solutions, insufficient selective pressure may exist to make progress toward the global Pareto front. The use of a large population can help to reduce the proportion of locally non-dominated solutions, but this approach is impractical for the many real-world problems in which evaluation of a single candidate solution is very time-consuming. Also, the reduction in the proportion of locally non-dominated solutions

that is achievable through the use of larger populations becomes progressively less as the number of objectives increases.

Many MOEAs use some method of population density estimation, either to modify the selection-for-variation probability of an individual or as part of the selection-for-survival acceptance procedure, to achieve a good distribution of solutions. Density estimation also becomes increasingly difficult as the dimensionality of the problem-space in which the estimates are to be computed is increased. This dimensionality is the number of objectives for density estimation in objective-space. Due to the 'curse of dimensionality' (the sparseness of data in high dimensions), the ability to fully explore surfaces in greater than five dimensions is regarded as highly limited (Scott 1992). Statisticians generally use dimensionality reduction techniques prior to application of the estimator. This assumes that the 'true' structure of the surface is of lower dimension, but the potential for reduction may be limited for a trade-off surface in which all objectives are in conflict with each other.

The effect of conflict in EMO, in the absence of any remedial measures, is explored empirically in Chapter 5 to follow. Possible measures that have been considered previously by the EMO community are described below.

4.3.3 Remedial Measures

Preferences

The exploitation of DM preferences, either *a priori*, *a posteriori*, or progressively, is arguably the current best technique for handling large numbers of conflicting objectives. In the *a priori* and progressive cases, the optimiser is required to represent only a limited DM region of interest in objective-space. Thus, good proximity and good distribution of solutions is only required in a sub-space of the global trade-off hypersurface. The classic *weighted-sum* approach to objective aggregation has been applied extensively in EA applications, but has considerable limitations. The aggregation of objectives is described further below, together with a more powerful preference-based scheme based on dominance-like concepts. Supplementary information on the use of preferences with EMO can be found in Section 2.4.5.

Aggregation. One popular method for reducing the number of conflicting performance objectives is to combine several of them into a single optimisation objective. Aggregation

may be achieved by means of a weighted-sum, or a more complicated function. In this approach, the DM pre-specifies the trade-offs between the combined subset of objectives. This eliminates the requirement for the optimiser to represent this portion of the global trade-off surface. The inherent disadvantage of the approach is that the DM must be able to specify the required trade-off *a priori*. Also, large adjustments to the preferences may require a complete re-run of the optimisation. Nevertheless, this may be an appropriate technique, especially when faced with very large numbers of objectives. Note that optimisers based on the weighted-sum method of aggregation are unable to identify non-convex regions of the trade-off surface (Censor 1977).

Goals and Priorities. Greater flexibility can be achieved in terms of objective reduction by exploiting goal values and priorities for various objectives, if these can be elicited from the DM. The preferability relation developed by Fonseca and Fleming (1998a) unifies various classical operations research schemes based on goals and priorities and applies them within the context of EMO. In essence, the method adaptively switches on or off different objectives, from the perspective of the dominance relation, for each pair of vectors considered. The iterative nature of the EA paradigm can be exploited to update the preferences as information becomes progressively available to the DM. Similar schemes to that proposed by Fonseca and Fleming (1998a) exist in the EMO literature, as documented in Chapter 2.

Dimension Reduction

Existing dimensionality reduction techniques could be used to transform objective-space into a lower dimension. This could be done prior to the optimisation, based on some preliminary analysis, or could be updated on-line as the MOEA evolves. The key benefit of the latter approach is that, as the MOEA progressively identifies the trade-off surface, the reduction is performed on a space more relevant to both the EA and the DM. If the reduction is to be performed iteratively then the balance between capability and complexity of the applied technique must be considered. For example, *curvilinear component analysis* (Demartines and Hérault 1997) has good general applicability but a significant computational overhead, whilst *principal components analysis* (Joliffe 1986) has the opposite features.

Dimension reduction methods can be applied directly to the density estimation process to preserve trade-off diversity in information-rich spaces. However, since the methods do not

respect the dominance relation, they cannot be used directly in the Pareto ranking process without modification (Collette, Siarry and Wong 2000).

Visualisation

Note that the ability to visualise the developing trade-off surface becomes increasingly difficult as the number of objectives increases. The method of parallel coordinates is a popular countermeasure for large numbers of objectives. Scatter-plots with brushing and glyph approaches, such as Chernoff faces (Chernoff 1973), are amongst the possible alternatives (Scott 1992, Cleveland 1993). Parallel coordinates and scatter-plots are both closely linked to the concepts of conflict and harmony described in this paper, and are discussed further in Section 4.6

Obayashi and Sasaki (2003) have used *self-organising maps* (SOMs) for visualisation purposes in EMO. A SOM is an unsupervised neural network that performs a mapping of data from a high dimension on to a lower dimensional structured lattice, whilst attempting to preserve the topology of the data (Kohonen 1995). Obayashi and Sasaki (2003) used two-dimensional neural lattices to highlight the relationships between the objectives and also to cluster decision-space to discover overall qualitative categories of solution that correspond to particular trade-off choices.

4.4 Harmonious Objectives

4.4.1 Definitions of Harmony

A relationship in which enhancement of performance in an objective is witnessed as another objective is improved can be described as harmonious. If performance in the objective is unaffected, the relationship is described as weakly harmonious. Complete definitions are provided below and can be related to the relevant $\boldsymbol{z}^b$-relative-to-$\boldsymbol{z}^a$ regions and lines in Figure 4.3.

Definition 4.2 (Harmony) *Levels of harmony are determined by the condition* $(\boldsymbol{z}_i^a < \boldsymbol{z}_i^b) \wedge (\boldsymbol{z}_j^a < \boldsymbol{z}_j^b)$. *If* $\neg\exists(a,b)$ *for which the condition holds then there is* no harmony, *if* $\exists(a,b)$ *then there is* harmony, *whilst if the condition holds* $\forall(a,b)$ *then there is* total harmony.

Definition 4.3 (Weak Harmony) *Levels of weak harmony are determined by the condition* $\Big[(\boldsymbol{z}_i^a < \boldsymbol{z}_i^b) \wedge (\boldsymbol{z}_j^a = \boldsymbol{z}_j^b)\Big] \vee \Big[(\boldsymbol{z}_i^a = \boldsymbol{z}_i^b) \wedge (\boldsymbol{z}_j^a < \boldsymbol{z}_j^b)\Big]$. *If* $\neg\exists(a,b)$ *for which the condition holds*

then there is no weak harmony*, if* $\exists(a,b)$ *then there is* weak harmony*, whilst if the condition holds* $\forall(a,b)$ *then there is* total weak harmony*.*

Definition 4.4 (Neutrality) *Neutrality is determined by the condition* $\left(\boldsymbol{z}_i^a = \boldsymbol{z}_i^b\right) \wedge \left(\boldsymbol{z}_j^a = \boldsymbol{z}_j^b\right)$*. If* $\neg\exists(a,b)$ *for which the condition holds then there is* no neutrality*, if* $\exists(a,b)$ *then there is* neutrality*, whilst if the condition holds* $\forall(a,b)$ *then there is* total neutrality*.*

Harmonious relationships have been observed in several EMO application papers, where they are indicated by non-crossing lines between pairs of objectives on a parallel coordinates plot (see Section 4.6), including the following:

- passenger cabin acceleration versus control voltage in electromagnetic suspension controller design for a *maglev* vehicle (Dakev, Whidborne, Chipperfield and Fleming 1997),
- gain margin versus phase margin, and 70% rise time versus 10% settling time, in the design of a *Rolls-Royce Pegasus* low-pressure spool speed governor (Fonseca and Fleming 1998b).

4.4.2 Effect on EMO

In either form of total harmony, one of the objectives can be removed without affecting the partial ordering imposed by the Pareto dominance relation on the set $\boldsymbol{Z_R}$ of candidate solutions. This type of relationship has received some consideration in the classical OR community, usually for $\boldsymbol{Z_R} = \boldsymbol{Z_*}$, where one member of the objective pair is known variously as *redundant*, *supportive*, or *nonessential* (Agrell 1997, Carlsson and Fullér 1995, Gal and Hanne 1999). It remains an open question whether or not to include redundant objectives in the optimisation process. Reasons to keep such objectives include:

- knowledge of the relationship may be of interest to the DM, especially if the rate of harmonious behaviour changes over the course of the search space,
- the relationship may not be apparent from a random finite sample of the search space,
- inclusion does not, necessarily, harm the search,
- the DM may be more comfortable with the inclusion of the objective.

Reasons to remove redundant objectives include:

- to eliminate the extra burden on the DM, who must inspect and make decisions on matters that do not affect the search and may be misleading,
- to reduce the computational load, in terms of both performance evaluations and comparisons.

The inclusion of a redundant objective does not affect the partial ordering of candidate solutions imposed by the Pareto dominance operator. Thus, EMO progress toward the global Pareto front is unaffected. It is, however, possible that such an inclusion could affect the diversity in the representation of the trade-off hypersurface. This depends on the definition of distance between objective vectors used by the density estimator. For example, any procedure using Euclidean distances or the *NSGA-II crowding* algorithm (Deb, Pratap, Agarwal and Meyarivan 2002) could suffer from potential bias. Consider the case of three objectives: where z_1 and z_2 totally conflict, z_1 and z_3 totally conflict, and z_2 and z_3 are in total harmony. The resulting trade-off surface is one-dimensional, and can be represented by the conflict between z_1 and z_2. A uniform distribution in the Euclidean sense may not be arrived at across the normalised trade-off surface, even if such a distribution is achievable, because the Euclidean distance calculation is biased in favour of z_2 since $\{z_2\ z_3\}$ has greater influence on the Euclidean distance measure than z_1. Thus, a diversity preservation technique would bias in favour of diversity in z_2 on the trade-off surface. The overall effect of this depends on the trade-off surface in question: sometimes, good diversity in z_2 will naturally lead to good diversity in z_1 but this is not guaranteed to be the case.

4.4.3 Remedial Measures

Redundant objectives may be identified by (i) careful use of *a priori* DM knowledge and (ii) using the sample set contained within the EA population for each objective and looking for large positive correlations between the data sets for each pair of objectives. Redundant objectives may be removed if this is felt appropriate for the problem in-hand. Alternatively, the objectives may be selectively ignored in the density estimation process (and also the ranking process in order to reduce the number of unnecessary comparisons) and yet still be presented to the DM.

Caution should be exercised when considering the removal of an objective identified as redundant through the analysis of sample data. This is because the spatial relationships between objectives can vary through objective-space and, thus, the relationship identified for one set of sample data may not be representative of the relationship in the actual DM region of interest ($\boldsymbol{Z}_*$).

4.5 Independent Objectives

4.5.1 Independence in the Context of EMO

In this thesis, *independence* refers to the ability to decompose the global optimisation problem into a group of sub-problems that can be solved separately from each other. Thus, different objectives and decision variables will be allocated to different sub-problems.

In the context of the relationship between a pair of objectives, independence means that the objectives can, in theory, be optimised completely separately from each other. As with a harmonious relationship, it is possible to make improvements to both objectives simultaneously (from the perspective of the complete solution). The difference between independence and harmony is that appropriate adjustments must be made to two distinct parts of the complete solution in the former case, whilst in the latter case a single good decision modification for one of the objectives will naturally produce improvement in the second objective.

If two objectives are independent then they do not form part of the same trade-off surface. Thus multiple, distinct, trade-off surfaces exist, each of which should be represented separately for inspection by the DM.

4.5.2 Effects of Independence on EMO

Consider a global problem, $\mathcal{P}$, comprised of ψ independent sub-problems $[\mathcal{P}_1, \ldots, \mathcal{P}_\psi]$ with associated independent collections of objectives $[\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\psi]$ and independent collections of decision variables $[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\psi]$. If advance knowledge of these collections is available then the global problem can be decomposed into the groups of sub-problems prior to optimisation. Then a proportion of the total available resources (candidate solution evaluations) could be exclusively allocated to the optimisation of each sub-problem. Both a global approach and the aforementioned *divide-and-conquer* method should yield the same solution of ψ independent

trade-off surfaces. From an EMO perspective, it then becomes a matter of interest as to which technique produces superior results in terms of trade-off surface quality. Is the effort expended identifying and exploiting the correct decompositions rewarded with improved results?

In the first study of its kind, Purshouse and Fleming (2003a) attempted to answer this question. The study demonstrated that, for a simple test problem, a divide-and-conquer strategy could substantially improve EMO performance. *A priori* decompositions were evaluated in objective-space, decision-space, and both spaces simultaneously. Parallel EA models were applied to each sub-problem. All three methods led to significantly higher-quality trade-off surfaces than the global approach, with both-space decomposition proving the most attractive. Given that it may not be possible to accurately identify the sub-problems in advance of the optimisation, an on-line adaptive divide-and-conquer strategy for MOEAs has also been proposed and evaluated by Purshouse and Fleming (2003a). Bivariate statistical tests for independence are applied to the population sample data in order to identify the independence relationships. The study is documented in Chapter 6 of this thesis.

4.6 Existing Methods for Identifying Pair-Wise Relationships

This chapter considers the relationships that exist between pairs of objectives, by comparing pairs of objective vectors. In this approach, composite relationships must be inferred from these simpler relations. However, the pair-wise methodology is very popular in multivariate studies and forms a good foundation for analysis, with many qualitative and quantitative techniques based on this approach. Methods that are closely linked to the definitions of conflict and harmony described earlier are discussed in the remainder of this section.

4.6.1 Qualitative Methods

The method of parallel coordinates, first described by Inselberg (1985) and subsequently applied to EMO by Fonseca and Fleming (1993), reduces an arbitrary high-dimensional space to two-dimensions. The parallel coordinates representation of the approximation set of Figure 4.2 is shown in Figure 4.4. Objective labels are located at discrete intervals along the horizontal axis (and these should be interchangeable). Normalised performance in each objective is indicated on the vertical axis. A particular objective vector is displayed by joining the performance levels in all adjacent objectives by straight lines. Then, considering two

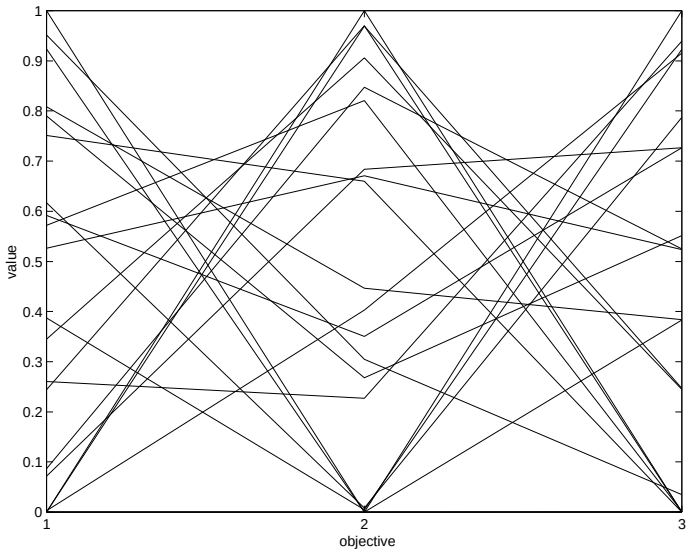


Figure 4.4: Parallel coordinates representation of the data shown in Figure 4.2

objective vector instances for a pair of objectives, the lines representing the two instances will cross if conflict is exhibited according to Definition 4.1 or will fail to cross if harmony is observed according to Definitions 4.2 or 4.3 (in the case of Definition 4.4, the lines will be superimposed). Thus, the magnitude of conflict is heuristically visualised as 'many' crossing lines.

Wegman (1990) presents some valuable insights and extensions toward using the parallel coordinates representation as a high-dimensional data analysis tool. Statistical interpretations of the plots are possible, with features such as marginal densities, correlations, clusters, and modes proving readily identifiable. Parallel coordinates plots can suffer from over-plotting for large data sets and thus a density plot variant is also presented by Wegman (1990) to overcome this.

Another popular method of pair-wise visualisation, which in its full form presents more simultaneous comparisons than the standard parallel coordinates plot, is the scatterplot matrix (Cleveland 1993). Such a plot for the MOGA approximation set of Figure 4.2 is shown in Figure 4.5. Each element of the matrix of plots shows a particular bi-objective section of the trade-off surface. For example, the upper central plot shows z_2 on the horizontal axis and z_1 on the vertical axis. It can sometimes be difficult to extract information from these plots,

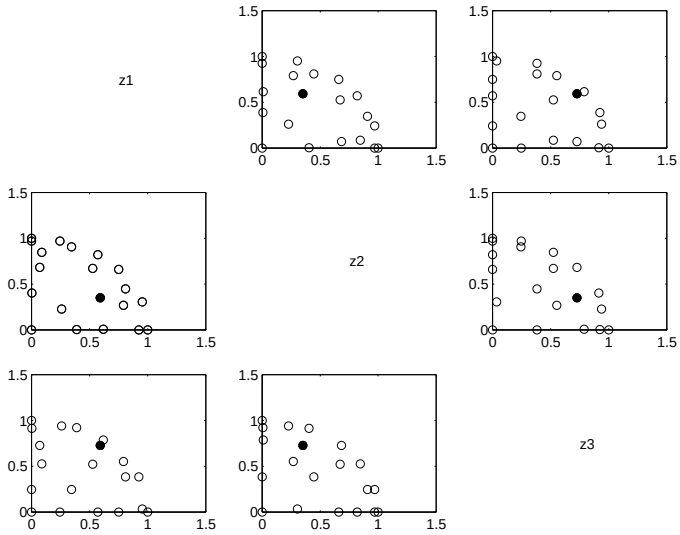


Figure 4.5: Scatterplot matrix representation of the data shown in Figure 4.2

especially as the number of objectives increases. Highlighting of a particular objective vector instance or group of instances — a technique known as *brushing* — can often aid higher-order understanding. The filled circle in Figure 4.5 indicates one particular vector.

4.6.2 Quantitative Methods

Several pair-wise methods exist for quantifying conflict between objectives that use similar concepts to the parallel coordinates notion of crossing lines. The *Kendall sample correlation statistic* measures the difference between the number of *concordant* samples (as one variable increases/decreases, the other follows suit) and the number of *discordant* samples (as one variable increases/decreases, the other does the opposite) (Kendall 1938). Thus, discordance produces crossing lines whilst concordance does not. The link to the multi-objective concepts of conflict and harmony is clear. Fuzzy measures of conflict also use this type of approach: see, for example, Lee and Kuo (1998).

Schroder (1998) developed a technique based directly on the method of parallel coordinates. In this approach, each objective range is partitioned into a number of equally sized regions. The level of conflict is then defined as a weighted-sum of the crossings between pairs of regions (rather than actual solutions), where the weights are based on the separation

between the regions. Crossings between distant regions are argued to be indicative of strong conflict. The method also normalises conflict levels with respect to population density. This may be appropriate if one particular region is thought to be over-sampled by the search, but may not in general be correct because if vectors are similar in two objectives this does not necessarily mean they are from the same region of the global trade-off surface. The method also requires additional preference information, unlike the previous techniques, since it is not based purely on ordinal data. However, more information can, potentially, be extracted using this method.

4.7 Summary

EMO applications have long considered the simultaneous optimisation of large numbers of objectives. However, EMO algorithm developers have tended to concentrate almost entirely on the bi-objective domain. Thus, there is a present lack of understanding of how MOEAs cope with larger numbers of objectives, and what functionality is required from an EMO algorithm. This chapter has sought to lay foundations for future work in this direction by considering how increasing numbers of objectives affect MOEA search performance.

Three relationships — conflict, harmony, and independence — have been identified. It has been demonstrated how the relationship between two objectives can contain elements of both conflict and harmony, resulting from interaction with other objectives.

It has been argued that increasing numbers of conflicting objectives will severely hamper the ability of an MOEA to represent the global trade-off surface in a form that is of use to a DM. Many candidate solution evaluations will be required to represent the naturally high-dimensional trade-off surface of an EMO environment at a resolution considered the norm for bi-objective tasks. Even if such a resolution were achievable, the resulting amount of information is likely to cognitively overwhelm the DM, who must ultimately select a single solution. Lower resolution approximations of the entire surface can be equally unsatisfactory because the spaces between neighbouring solutions may be much greater than the DM indifference values. Also, as argued in Chapter 2, for a many-objective problem the DM generally has a limited region of interest within the overall trade-off space. Thus, the 'pure' EMO aim of representation across the entire span of the surface is inefficient for many practical applications. Furthermore, the dimension of the problem must be kept reasonably low for

human DM analysis to remain tractable: twelve objectives has been suggested as an upper limit (Bouysson 1990). Aggregation of objectives, whilst anathematic to many EMO researchers, may thus be necessary. DM preferences are an inescapable part of evolutionary many-objective optimisation.

Even if the number of objectives is limited by DM considerations, it may still be prudent to consider further dimensionality reduction in the context of the MOEA search. This can be achieved by removing some objectives from certain comparisons (and thus preserving the dominance relation) or by applying some form of transformation to a new set of coordinates (the standard dimension reduction approach). The utility of methods based on the latter approach is limited because they do not respect the dominance relation. They may, however, be used to help achieve good diversity in information-rich spaces (where 'information' is defined according to the chosen method).

Interactive preference articulation schemes, such as that described by Fonseca and Fleming (1998a), are particularly valuable in problems with large numbers of conflicting objectives and fit very nicely within the iterative EA framework. In such schemes, the attention of the optimiser is focused on various sub-regions of the trade-off surface as the search progresses. This is beneficial to the DM who may only be interested in learning about certain trade-offs within the global problem. The progressive nature of the scheme suits the often changing aspirations of the DM as more knowledge is uncovered. The main drawback of this approach is that it can be rather DM-intensive.

To summarise, increasing numbers of conflicting objectives in a problem transforms the aim of EMO from identification of a globally optimal solution set toward assisting the DM in learning about the trade-offs between objectives and finding an acceptable solution. Harmonious objectives, and the special case of redundancy, do not have the same severe impact on EMO as does conflict. Convergence to the Pareto front is unaffected by increasing numbers of totally harmonious objectives. Issues surrounding distribution of solutions across the surface do require some care however. The decision on whether to eliminate any identified redundant objectives from the search is perhaps best left to the discretion of the analyst and the DM. The existence of independence within the global problem leads to multiple, separate, trade-off surfaces. If independence can be identified then the deployment of a divide-and-conquer strategy could potentially improve EMO performance. In conclusion, the simultaneous con-

sideration of many objectives is arguably the greatest challenge facing the EMO community at the present time. This challenge is met in the following two chapters of the thesis. Chapter 5 seeks to reveal and explain contemporary MOEA performance in an environment of many conflicting objectives. Chapter 6 considers the benefit of a divide-and-conquer strategy for decomposable E$\mathbb{M}$O tasks.

Chapter 5

The Effect of Increased Conflict in $\mathbb{EMO}$

5.1 Introduction

The previous chapter has identified the requirement for more theoretical work into the optimisation of many objectives using evolutionary algorithms. Consideration of the different relationships between objectives — classified as conflict, harmony, and independence — has been suggested as a route to understanding the behaviour of $\mathbb{EMO}$ algorithms.

In this chapter, the effect of many **conflicting** objectives on the performance of MOEAs is explored. Recall that conflict is experienced between two objectives when improved performance in the first objective can only be achieved at the expense of deterioration in the performance of the second objective.

In Section 4.3.3 it was argued that the exploitation of decision-maker preferences is a key technique for managing the optimisation of many conflicting objectives. The inherent difficulties in solving many-objective problems have already lead EMO researchers to incorporate preference-based schemes within their algorithms. These methods have been introduced in Section 2.4.5 and a comprehensive review can be found in the work of Coello (2002). The fundamental aim of these methods is to limit the search requirements of the optimiser to a sub-region of overall objective-space. However, as argued by Knowles (2002), the potential for an exclusively Pareto-based solution to the many-objective optimisation problem remains a matter of some interest. Indeed, if the resolution of the obtained approximation set is

regarded as a function of some maximum limit imposed on the size of the set (such as the population size of an MOEA), then there is no *a priori* restriction such that the achieved set should not be globally non-dominated and optimally distributed across the trade-off surface. But is it possible to design an evolutionary algorithm that is capable of generating such results, given finite resources, when faced with many conflicting objectives?

A family of tractable, real-parameter optimisation tasks that are scalable to any number of conflicting objectives was proposed by Deb, Thiele, Laumanns and Zitzler (2002) to stimulate research into many-objective optimisation. In the first known study of its kind, this test suite was used by Khare, Yao and Deb (2003) to investigate the scalability of three contemporary MOEAS: Deb, Pratap, Agarwal and Meyarivan's (2002) NSGA-II, Corne *et al.*'s (2000) PESA, and Zitzler, Laumanns and Thiele's (2001) SPEA2. The implementation of PESA was found to produce approximation sets with good overall proximity of solutions to the global Pareto front but with a poor distribution when solving for many conflicting objectives. By contrast, the approximation sets obtained for the implementations of NSGA-II and SPEA2 exhibited a good distribution of solutions but with poor proximity. However, since a single design-space instance of each algorithm was used, and each algorithm is itself a complicated structure of basic EMO components, it is not immediately clear from the study which components and processes are critical from the many-objective optimisation perspective. In particular, it is not clear why PESA should produce approximation sets of a fundamentally different character to NSGA-II and SPEA2.

In this chapter, in a manner consistent with the approach developed in Chapter 3, MOEA processes are considered at a component level rather than the brand-based approach adopted by Khare *et al.* (2003). The observed EMO behaviour can subsequently be explained in terms of fundamental search components and processes. In addition, results are generated for a *map* of variation operator configuration settings. This permits analysis to be made in terms of the exploration-exploitation trade-offs in EMO (Bosman and Thierens 2003) and for regions of high performance configurations (*sweet-spots*) to be identified (Goldberg 1998).

The class of MOEAs studied in this chapter is introduced in Section 5.2, for which the associated selection and variation processes are considered in detail. The empirical framework of the inquiry is developed in Section 5.3, which draws on concepts from Chapter 3 and the work of Laumanns *et al.* (2001). The optimisation task used in simulations is described at

this stage, together with the indicators used to measure approximation set performance. The presentation of results is also considered. Simulation results for mutation-based algorithms are presented in Section 5.4. The results for recombination-based optimisers are included in Section 5.5. The combined results from these two sections are then analysed in Section 5.6, with respect to the underlying processes described in Section 5.2. Deb (2001a) suggests that a potential technique for obtaining good EMO results is to increase the population size of an MOEA as an, ideally exponential, function of the number of objectives. This is unlikely to be a feasible approach in practice because of the computational resources demanded. However, the potential benefits of this technique remain a matter of interest and a population sizing study is thus explored in Section 5.7. Note that a more intensive investigation of conflict in EMO to that recorded in this chapter is provided by Purshouse and Fleming (2003d).

5.2 Algorithms Considered

5.2.1 Introduction

The different multi-objective evolutionary optimisers proposed in the literature can generally be categorised according to the manner by which selection-for-variation and selection-for-survival are performed. A taxonomy of the various techniques is offered by Laumanns *et al.* (2001). The representation and variation operators are problem-specific but interact heavily with the selection operators (mainly in terms of the e-e trade-off) to form the complete algorithm.

Each MOEA is comprised of a number of different selection mechanisms. As argued in Section 3.1, it can often prove challenging to correctly determine which mechanisms and combinations of mechanisms are chiefly responsible for the observed algorithm performance. Thus, as was the case in Chapter 3, deconstructed algorithms are again used, but this time to determine the fundamental processes that are responsible for the observed *many-objective* behaviour of MOEAs.

NSGA-II components have specifically been chosen for this inquiry (Deb, Pratap, Agarwal and Meyarivan 2002). These components tend to be computationally and conceptually simple when compared to other algorithm processes in the literature. Also, the NSGA-II has been heavily studied by the EMO community. As evident from the review of EMO in Section 2.4,

the NSGA family of processes has strong similarities with other popular MOEAs, such as the SPEA (Zitzler and Thiele 1999) and MOGA (Fonseca and Fleming 1993) families.

NSGA-II, like many other MOEAs, uses Pareto dominance and density estimation considerations in its selection processes. Further to the review in Chapter 2, these are discussed in detail in Section 5.2.2 to follow. Two algorithms are considered: NSGA-II and a decomposition of NSGA-II.

The key issues that are identified from the many-objective analysis of these two algorithms should be readily extendible to other MOEAs that share the following properties:

1. The selection mechanisms operate on a global population model, for which s_s respects a arbitrarily chosen specific integer upper bound on the population size.

2. Selective bias is primarily based on Pareto dominance.

3. Selective bias *may* be secondarily based on density estimation.

4. If two solutions have equal properties from both a dominance and density perspective then selection should not favour one solution over the other.

Many of the selection mechanisms proposed in the literature, and described in Chapter 2, share the above four properties. These include early algorithms, such as MOGA, NSGA, and NPGA, and also contemporary MOEAs such as SPEA, SPEA2, and the state-of-the-art MOGA described in Chapter 3.

An example of a methodology that is *not* represented by the processes considered in this chapter is the ϵ-dominance selection-for-survival concept developed by Laumanns, Thiele, Deb and Zitzler (2002). In this scheme, all locally non-dominated solutions can be retained (within the restrictions imposed by ϵ), and the population size is increased dynamically to account for this. The theoretical upper bound on population size is known but a specific size cannot be specified. Thus, property 1 listed above is different for this scheme.

Properties 2 and 3 are met by most of the Pareto-based MOEAs. Clearly, approaches based on classical OR methods (such as weighted-sum and target vector approaches) do not have these properties and are not considered in this chapter.

The selection-for-survival method in the PAES algorithm does not respect property 4 (Knowles and Corne 2000). When a new locally non-dominated solution is considered for

inclusion in $P[t+1]$, if the population size limit is reached then the solution can only be included if it has a density estimate that is lower than that of any solutions in the partially updated $P[t+1]$. If these latter solutions all have the minimum possible density estimate (each resides in a separate hyperbox) then the new solution cannot be included even if it also unique to its associated hyperbox. Thus, despite being equal in terms of both dominance and density, the new solution is maximally biased against (the probability of selection is zero) when compared to the solutions in the partially updated $P[t+1]$. Note that the PAES method can be made to respect property 4 by changing the $<$ requirement on density to $\leq$.

If the four properties of selection are shared between two algorithms this does not imply that the selection methods are directly equivalent. The properties can be implemented in a different manner: for example, within either tournament selection or proportional selection. Also, multiple levels of selective bias are possible: dominance and density discrimination only occur at s_v in NSGA but are implemented at both s_v and s_s in NSGA-II. Ultimately, the differences between the mechanisms result in different selective pressures. Thus, even allowing for stochastic errors, different mechanisms that share commonality via the four properties listed above can still produce different results through the interaction of selective pressure with the variation operator e-e configuration and the representation of the problem landscape.

5.2.2 Selection Mechanisms

Algorithm D1

Algorithm D1 has a selective bias that is solely attributable to Pareto dominance. Binary tournament selection is used at the s_v stage. Of two solutions selected at random from the current population (with replacement), the selected solution is the one that dominates the other. If the solutions are non-dominated with respect to each other then one of the solutions is selected at random.

At the s_s stage of D1, the $P[t]$ population and $v(s_v(P[t]))$ population are combined. This new population is then ranked according to dominance depth. The new population, $P[t+1]$ is then formed by including the best of the ranks until the population size limit is breached. Solutions that share the current rank under consideration in this situation are selected for inclusion on a random basis. This is a $(\mu + \lambda)$ strategy.

Algorithm D2

Algorithm D2 extends the selective capabilities of D1 by including supplementary density considerations in both s_v and s_s. The resulting algorithm is identical to the standard NSGA-II algorithm proposed by Deb, Pratap, Agarwal and Meyarivan (2002).

As in D1, binary tournament selection is used at the s_v stage. Of two solutions selected at random from the current population (with replacement), the selected solution is the one that dominates the other. If the solutions are non-dominated with respect to each other, then the solution with the smallest density estimate is selected. If the density estimates are also tied then a solution is chosen at random. Deb, Pratap, Agarwal and Meyarivan (2002) define this method as the *crowded-comparison operator*.

The s_s stage of D2 proceeds as for D1 until the population size limit is breached. At this point, instead of selecting solutions at random from the current rank, D2 selects solutions with the smallest densities from this rank until the new population is full. If the population size limit is again breached because of a tie on density information, then the solutions that are tied on both dominance depth and density are selected for inclusion on a random basis.

Any density estimator can potentially be used to provide the density information required at the s_v and s_s stages of D2. A review of estimators that have been proposed in the EMO community was undertaken in Section 2.4.4. This inquiry uses the *crowding distance* estimator proposed for use in the original NSGA-II (Deb, Pratap, Agarwal and Meyarivan 2002). In this estimator, density is calculated as the mean side length of the hypercube formed using the first-nearest-neighbour values in each objective as vertices. In this inquiry, the boundary condition for an objective is set to the maximum non-boundary value calculated for that objective (rather than being set to infinity as in the original method proposed by Deb, Pratap, Agarwal and Meyarivan (2002)). This ensures that the estimator is unbiased for the equilibrium condition of a perfectly distributed approximation set.

Crowding distance is a low complexity estimator with limited accuracy. It should be noted that the relative effectiveness of the method when compared to other estimators has been questioned by some researchers (Laumanns *et al.* 2001, Deb, Mohan and Mishra 2003). However it is argued that the behaviour of D2 when compared to D1 will be generally indicative of the effect of introducing any explicit diversity-promoting mechanism, *when such a mechanism is active.*

The activity — or otherwise — of the diversity promotion mechanisms is an important point. Some estimators, such as fitness sharing and simple hypergrid counting, operate within a defined local neighbourhood. If the neighbourhood is unsuitable for the current distribution of solutions then density information is lost. For example, in the hypergrid scheme, if all the solutions are contained within one hyperbox then they all have the same density estimate. This is also the case if each solution resides in its own hyperbox (regardless of the relative distances between occupied hyperboxes). This behaviour requires the careful selection of neighbourhood size: a task that has been automated in approaches by, for example, Fonseca and Fleming (1995a).

Estimators that are based on the nearest neighbour concept are not susceptible to information loss in the above sense, since the neighbourhood is effectively scaled automatically. However, the 1st-NN crowding distance can also fail in a different sense: if two vectors reside at the same location in objective-space then the estimates will be zero regardless of the distance to other solutions. Thus, all solutions with copies in the population will receive an identical density estimate. This is not generally the case for the schemes described above. Note that the NN estimator used in the s_s procedure of SPEA2 is more advanced than crowding distance and is not susceptible to the same problem, although it carries a higher complexity cost. In summary, it is not sufficient for a diversity mechanism to be present: it must also be *active*.

5.2.3 Representation and Variation

This inquiry will simulate the many-objective performance of the selection mechanisms described in the previous section using a real-parameter function optimisation problem (described later in Section 5.3.2). Since the decision variables in the task are real-valued, the choice of a real-valued genotype representation follows naturally for the reasons discussed in Section 3.3. Thus each element of the genotype corresponds directly to a decision variable in the problem.

Two frequently studied variation operators for real representations are described in this section. The first is a one-parent mutation operator that produces a single child solution, whilst the second is a two-parent recombination operator that produces two children. Both operators have associated parameters that allow levels of exploration and exploitation to

be controlled. The selection mechanisms in Section 5.2.2 can then be studied for a variety of e-e trade-off settings. The interest is not primarily in which settings are good for the problem considered, but rather in how the different selection methods perform with respect to each other under various e-e conditions when simultaneously optimising various numbers of conflicting objectives. The two operators are described in detail below.

Mutation

Deb and Goyal's (1996) *polynomial mutation* operator is used in the inquiry. This variation operator is popular in real-parameter multi-objective optimisation tasks, and has previously been used to solve the example problem used in this chapter (Deb, Thiele, Laumanns and Zitzler 2002, Khare *et al.* 2003). Variable-wise mutation is performed according to a probability distribution function centred over the parent value. The operator is defined in Equation 5.1, where γ_i is the parent value for the ith decision variable, u_i and l_i are the upper and lower bounds on the ith decision variable, η_m is a distribution parameter, r_i is a number generated uniformly at random from [0 1], and c_i is the resulting child value for the ith decision variable.

$$\left.\begin{aligned} c_i &= \begin{cases} \gamma_i + (\gamma_i - l_i)\,\delta_i & \text{if } r_i < 0.5, \\ \gamma_i + (u_i - \gamma_i)\,\delta_i & \text{otherwise.} \end{cases} \\ \delta_i &= \begin{cases} (2r_i)^{1/(\eta_m+1)} - 1 & \text{if } r_i < 0.5, \\ 1 - [2\,(1 - r_i)]^{1/(\eta_m+1)} & \text{otherwise.} \end{cases} \end{aligned}\right\} \tag{5.1}$$

Polynomial mutation has two controllable parameters: (i) the probability of applying mutation to a chromosome element, p_m, and (ii) a mutation distribution parameter, η_m. The latter parameter controls the magnitude of the expected mutation of the candidate solution variable. The normalised variation is likely to be of $\mathcal{O}(1/\eta_m)$. Thus, relatively speaking, small values of η_m should produce large mutations whilst large values of η_m should produce small mutations.

Mutation is applied independently to each element of each candidate solution with probability p_m. Thus, the probability of mutating a candidate solution of n decision variables is

defined as shown in Equation 5.2.

$$p(\text{mutate}) = 1 - (1 - p_m)^n \tag{5.2}$$

Recombination

Deb and Agrawal's (1995) *simulated binary crossover* (SBX) operator is also considered in the inquiry as an alternative to polynomial mutation. Unlike the latter operator, SBX is a two-parent variation operator that produces two new solutions. SBX has been considered extensively in previous EMO studies using real-parameter representations (Deb, Thiele, Laumanns and Zitzler 2002, Khare *et al.* 2003), and is defined in Equation 5.3, where $\gamma_{1,i}$ and $\gamma_{2,i}$ are the parent values for the ith decision variable, η_c is a distribution parameter, r_i is a number generated uniformly at random from [0 1], and $c_{1,i}$ and $c_{2,i}$ are the the resulting child values for the ith decision variable.

$$\left.\begin{aligned} c_{1,i} &= 0.5\left[(1+\beta_i)\,\gamma_{1,i} + (1-\beta_i)\,\gamma_{2,i}\right] \\ c_{2,i} &= 0.5\left[(1-\beta_i)\,\gamma_{1,i} + (1+\beta_i)\,\gamma_{2,i}\right] \\ \beta_i &= \begin{cases} (2r_i)^{1/(\eta_c+1)} & \text{if } r_i < 0.5, \\ \left[1/\left(1-r_i\right)\right]^{1/(\eta_c+1)} & \text{otherwise.} \end{cases} \end{aligned}\right\} \tag{5.3}$$

SBX generates child values from a probability distribution, with standard deviation derived from the distance between parent values and a distribution parameter η_c. The distance determines the overall magnitude of the distribution, whilst η_c determines the shape of the distribution. To generate child values for a decision variable, the distribution is centred over each parent and a random value is generated from the distribution to create one child. The second child is generated symmetrically to the first child about the mid-point between the parents. The child values for a decision variable are then exchanged between the complete child solutions with probability p_e. For simplicity $p_e = 0$ in the study. Note that child values that are generated outside the range of a decision variable are cropped to the nearest feasible value.

SBX is applied to a pair of parent solutions with probability p_c. If a uniform recombi-

nation scheme is chosen, in which each individual decision variable is selected for variation independently of any other (given that variation is to be applied to the selected parents), then the probability of applying recombination to a pair of candidate solutions (each comprised of n decision variables) can be expressed as shown in Equation 5.4

$$p(\text{recombine}) = p_c \left[1 - (1 - p_{ic})^n\right] \tag{5.4}$$

p_{ic} is the probability of applying variation to a decision variable, given that recombination is to be applied in general to the complete solution pair. In standard uniform recombination schemes in the literature, p_{ic} is usually set to 0.5. However, by allowing this probability to vary and setting p_c to unity, the probability of applying two-parent SBX to a solution can be viewed as equivalent to that expressed for polynomial mutation in Equation 5.2. This factor will assist in making comparisons between the results obtained using the two variation operators.

5.3 Inquiry Design

5.3.1 Introduction

The inquiry into the effect of increasing numbers of conflicting objectives in EMO combines process analysis with simulations of algorithm performance. The optimisation task used in the simulations is described in Section 5.3.2. The performance indicators chosen to reduce an approximation set to a more tractable number of summary statistics are presented in Section 5.3.3. The experimental framework that has been developed to facilitate new understanding of many-objective optimiser behaviour is introduced in Section 5.3.4. In Section 5.3.5, an introduction is made to the empirical results detailed in the remainder of this chapter.

5.3.2 Scalable Optimisation Task

This inquiry considers a real-parameter function optimisation task known as *DTLZ2*, which is defined in Equation 5.5. This problem was previously used, in a three-objective instance, to illustrate EMO concepts in Chapter 4. The task is taken from a highly tractable set of problems developed by Deb, Thiele, Laumanns and Zitzler (2002) specifically for studies

into many-objective optimisation. The global Pareto front is continuous and non-convex. Distance from the front is determined by a single, unimodal cost function, g. M is the number of objectives, $n = M + \kappa - 1$ is the number of decision variables, and κ is a difficulty parameter ($\kappa = 10$ in this study).

$$\left.\begin{array}{lrcl}
\text{min.} & z_1(\boldsymbol{x}) & = & [1 + g(\boldsymbol{x_M})]\cos(x_1\pi/2)\cos(x_2\pi/2)\ldots\cos(x_{M-2}\pi/2)\cos(x_{M-1}\pi/2), \\
\text{min.} & z_2(\boldsymbol{x}) & = & [1 + g(\boldsymbol{x_M})]\cos(x_1\pi/2)\cos(x_2\pi/2)\ldots\cos(x_{M-2}\pi/2)\sin(x_{M-1}\pi/2), \\
\text{min.} & z_3(\boldsymbol{x}) & = & [1 + g(\boldsymbol{x_M})]\cos(x_1\pi/2)\cos(x_2\pi/2)\ldots\sin(x_{M-2}\pi/2), \\
\vdots & \vdots & & \vdots \\
\text{min.} & z_{M-1}(\boldsymbol{x}) & = & [1 + g(\boldsymbol{x_M})]\cos(x_1\pi/2)\sin(x_2\pi/2), \\
\text{min.} & z_M(\boldsymbol{x}) & = & [1 + g(\boldsymbol{x_M})]\sin(x_1\pi/2), \\
\text{w.r.t} & \boldsymbol{x} & = & [x_1, \ldots, x_n], \\
\text{where} & g(\boldsymbol{x}_M) & = & \sum_{x_i \in \boldsymbol{x}_M} (x_i - 0.5)^2, \text{ with } \boldsymbol{x}_M = [x_M, \ldots, x_n], \\
\text{and} & 0 \leq x_i \leq 1, & & \text{for } i = 1, 2, \ldots, n, \text{ with } n = M + \kappa - 1.
\end{array}\right\} \tag{5.5}$$

DTLZ2 is comprised of decision variables of two distinct functional types: those that control convergence toward the globally optimal surface $(x_1, \ldots, x_{M-1})$ and those that control distribution in objective-space $(x_M, \ldots, x_n)$. The convergence variables define the distance of the solution vector from the true front via a κ-dimensional quadratic bowl, g, with global minimum $x_{M,\ldots,n} = 0.5$. The distribution variables describe position on the positive quadrant of the unit hypersphere. An M-objective instance of DTLZ2 is denoted by DTLZ2(M).

The DTLZ test suite encompasses many problem characteristics, such as multimodality, discontinuity, and distributional bias. For small numbers of objectives, such as three, it is a straightforward task to obtain a good sample-based approximation of the DTLZ2 global Pareto front using an EMO optimiser (Deb, Thiele, Laumanns and Zitzler 2002). This is not necessarily true for some of the other functions in the test suite, such as DTLZ4. However, as will be demonstrated in Sections 5.4 and 5.5, the generation of a good quality approximation set becomes significantly more challenging as the number of objectives is increased. Thus, the DTLZ2 function in isolation is a sufficiently interesting example of real-parameter, many-objective optimisation problems for the purposes of this study.

5.3.3 Performance Indicators

In the context of this inquiry, as was also the case in Chapter 3, performance is regarded as the quality of the trade-off surface discovered by an optimiser, given a finite number of candidate solution evaluations. As explained in Section 2.2.3, quality is generally expressed in terms of (i) the proximity of the obtained locally non-dominated vectors to the true Pareto surface and (ii) the distribution of those vectors across the surface (Bosman and Thierens 2003). Ideally, the optimiser should obtain solutions that are Pareto optimal (are of distance zero from the global front), that extend across the full range of optimal objective values (assuming that the DM is interested in complete trade-off information for the problem), and that are as near uniformly distributed as the true surface permits.

Many performance indicators have been proposed for EMO and a discussion of these was previously given in Section 3.2.3. Again, a functional approach is taken in this chapter, in which specific unary indicators are used to evaluate specific aspects of performance.

Proximity Indicator

The proximity indicator measures a median level of proximity of the approximation set, $\boldsymbol{Z_A}$, to the global surface. In terms of attainment across the objectives, an objective vector for DTLZ2 will respect Equation 5.6. The equality condition will only hold for a globally optimal vector. Thus, a specialised proximity indicator, I_P, for DTLZ2 can naturally be described by Equation 5.7. This is essentially the same as Veldhuizen's (1999) generational distance metric (previously used in Chapter 3 and described in Equation 3.1), for the case of a continuous globally optimal reference set, $\boldsymbol{Z_*}$.

$$1 \leq \left[\sum_{m=1}^{M} (z_m)^2\right]^{1/2} \tag{5.6}$$

$$I_P = \text{median}_{\boldsymbol{z_A} \in \boldsymbol{Z_A}} \left\{ \left[\sum_{m=1}^{M} (\boldsymbol{z}_{\boldsymbol{A}_m})^2\right]^{1/2} - 1 \right\} \tag{5.7}$$

Distribution Indicators

A multi-objective optimiser is required to find a good distribution of candidate solutions across the trade-off surface to present to the decision-maker: the identified non-dominated

vectors should span the complete surface, with appropriate distances between each. Given a direction of monotonically increasing DM preference in each objective and a continuous trade-off surface, the vectors should be equal distances apart.

To achieve high quantisation of the non-dominated set, it would be advantageous to express both aspects of distribution within a single indicator. This approach has been implemented in Deb, Pratap, Agarwal and Meyarivan's (2002) Δ metric and was used in Chapter 3. Using this approach it can be somewhat unclear which aspect of the distribution — extent or uniformity — is responsible for the observed indicator value. Thus in this study, in order to manage the complexity of the inquiry, only the extent of solutions is considered further.

Spread Indicator The study uses a variant of Zitzler's (1999) maximum spread indicator. This metric measures the length of the diagonal of the hypercube with vertices set to the extreme objective values observed in the achieved approximation set, as defined in Equation 5.8.

$$\lambda = \left[\sum_{m=1}^{M} \left(\max_{\boldsymbol{z_A} \in \boldsymbol{Z_A}} \{\boldsymbol{z}_{\boldsymbol{A}_m}\} - \min_{\boldsymbol{z_A} \in \boldsymbol{Z_A}} \{\boldsymbol{z}_{\boldsymbol{A}_m}\} \right)^2 \right]^{1/2} \tag{5.8}$$

It is possible to achieve too much or too little spread. In the former case, the vectors span regions that are not part of the global trade-off surface, (highlighting a relationship between spread and proximity). In the latter case, the optimiser has converged to a sub-region (that *may* be globally optimal). To highlight the requirement for an intermediate spread value, the indicator, I_S, is normalised with respect to the optimal spread, as indicated in Equation 5.9. Indicator values decreasing from unity to zero now represent increasing levels of population convergence to a sub-region. Thus, globally optimal regions of the surface are missing. Indicator values increasing from unity demonstrate widespread dispersal of vectors throughout non-optimal objective-space.

$$I_S = \lambda \Bigg/ \left[\sum_{m=1}^{M} \left(\max_{\boldsymbol{z}_* \in \boldsymbol{Z}_*} \{\boldsymbol{z}_{*_m}\} - \min_{\boldsymbol{z}_* \in \boldsymbol{Z}_*} \{\boldsymbol{z}_{*_m}\} \right)^2 \right]^{1/2} \tag{5.9}$$

5.3.4 Investigative Framework

The inherent high dimensionality of many-objective optimisation presents both conceptual and computational challenges to the analysis of algorithm behaviour. Thus, the inquiry

framework is aimed toward exploratory data analysis rather than statistically significant performance comparison. Following the methodology of Laumanns *et al.* (2001), single replication results are generated for a wide variety of configuration instances (each representing a particular e-e trade-off setting) to yield a response map in optimiser design-space. The use of multiple replications is still regarded as preferable, but this is computationally impractical for this inquiry. Note that spatial similarity between optimiser responses arguably provides some support for statistical confidence (or otherwise) in the observed behaviour. Laumanns *et al.* (2001) also fitted a polynomial model *response surface* to the raw data map to obtain a smooth representation. However, for the purposes of this inquiry, the raw data itself proves sufficient to expose the spatial relationships in the data. For further details on response surfaces, refer to Myers and Montgomery (2002).

The configuration of the variation operators, via both the probability and expected magnitude of variation, provides suitable control over the overall algorithm e-e trade-off. Optimiser responses have been obtained for all pair-wise permutations from sample sets of probabilities and magnitudes, with elements chosen according to a heuristic, pseudo-logarithmic scale that helps to show relativity within and between different response maps. The maps themselves portray scalar summary statistics for each overall response, such as proximity and spread indicator values.

Responses for algorithms incorporating mutation are presented in Section 5.4. These are measured over 1000 generations and are generated for a variety of values of M with the population size fixed at 100. Similar results for recombination are presented in Section 5.5. Results for both the variation operators are subsequently analysed in Section 5.6. Deb (2001a) has suggested that a key method for coping with large M is to increase the population size, since this will tend to reduce the proportion of the population that is non-dominated and thus provide improved dominance-based discrimination. Whilst this approach is unlikely to be practical in many real-world applications, where the computational cost of evaluating a candidate solution may be very high, results for various population sizes are considered in Section 5.7.

5.3.5 Introduction to Results

The types of response map shown in the results and analysis sections are introduced below, together with an example of how the results are presented.

Types of Response Map

Two performance response maps are considered in the results, together with various other process variable maps as described below:

Proximity is measured by subtracting the initial proximity indicator value, computed as described in Equation 5.7, from the median of the values obtained from the final 100 iterations of the optimisation process. Thus, a proximity value of zero indicates no progress from the initial population, a negative value indicates convergence toward the global trade-off surface, and a positive value indicates divergence away from the true surface.

Spread is measured as the median value of the spread indicator, described in Equation 5.9, taken over the final 100 iterations of the optimiser. The optimal value of spread is unity. Values less than unity indicate approximation set convergence to a region of objective-space that is smaller than the true trade-off surface. Values larger than unity indicate that the approximation set extends over a greater region than that described by the optimal surface.

Process measurements of other system variables assist the analysis of the observed optimiser performance behaviour. Variation operator outcomes are particularly important. Response maps for the outcomes listed below, measured over the first 10 generations of the optimiser are presented during the analysis in Section 5.6. Values for system variables tend to be dynamic over the course of the optimisation. Experience gained during simulation data collection suggests that early measurements can be highly indicative of the general trend of optimiser behaviour. For example, variation success rates may be initially high in an algorithm that exhibits good performance, but these rates may decrease as the algorithm converges close to the global surface. In an unsuccessful algorithm, early rates may be lower but do not continue to decrease because successful convergence does not occur. It is important to be able to draw a distinction between the summary success rates for these two algorithms. Early measurements assist in meeting this aim.

- **Copy rate.** The proportion of child solutions that are direct copies of (at least) one of their parents.

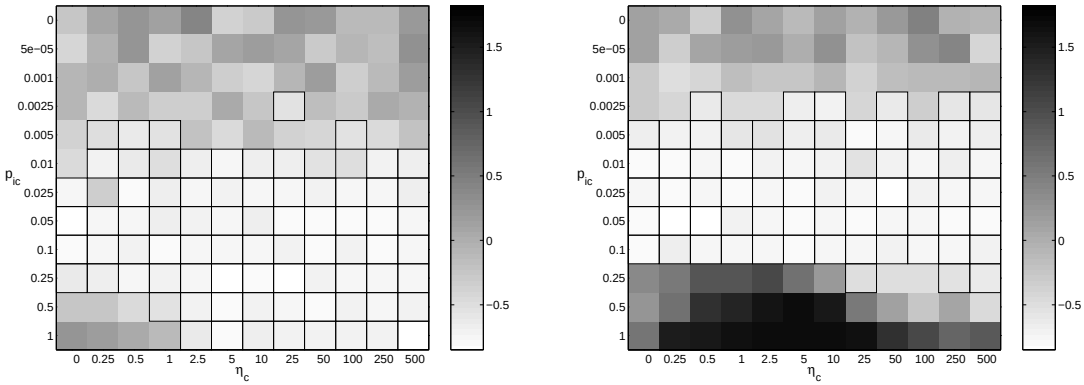
- **Resistance rate.** The proportion of child solutions that are non-dominated with respect to all of their parents.
- **Success rate.** The proportion of child solutions that dominate at least one parent and are not copies of any parent.

Presentation of Results

An example presentation of response maps for the proximity indicator, I_P is shown in Figure 5.1. Part (a) of the figure, on the left, shows the response map for the algorithm D1 that does not include any specific diversity promotion mechanisms. Part (b) of the figure, on the right, shows the corresponding map for the algorithm D2 that does include such mechanisms. The results have been collected for algorithms that use recombination: the range of expected recombination magnitudes, η_c, is shown on the horizontal axis. The vertical axis shows the range of recombination application probabilities, p_{ic}. If the algorithm had been using mutation, the same ranges would be shown for η_m and p_m respectively. Note that variation magnitude is related to the inverse of the distribution parameter for both recombination and mutation, and thus discussions relating to this will generally refer to large values of $1/\eta_{c,m}$ rather than small values of $\eta_{c,m}$.

Performance for each $\{p_{ic}, \eta_c\}$ setting is indicated by a grey-scale square at the appropriate location. Lighter shades of grey indicate better proximity, as shown by the colour-bar of indicator values to the right of each map. Considering Figure 5.1b, a range of good proximity is thus evident for intermediate values of p_{ic} in the range [0.005 0.1]. Conversely, a region of poor proximity is evident in the region of intermediate $1/\eta_c$ coupled with high p_{ic}, such as $\{p_{ic} = 0.5, \eta_c = 5\}$.

The grid squares highlighted by a solid boundary correspond to configurations that exceed a pre-defined performance threshold. For proximity, the performance threshold is -0.5. The configuration $\{p_{ic} = 0.0025, \eta_c = 25\}$ is one such example in Figure 5.1a. For the spread response maps, performance is highlighted in the range [0.75 1.25] (within 25% of optimal).



(a) D1 (b) D2

Figure 5.1: Example D-class proximity response maps

5.4 Results: Mutation

The three-objective proximity map for algorithm D1 is shown in Figure 5.2a. The corresponding results for algorithm D2 are shown in Figure 5.2b. Both algorithms produce approximation sets with good proximity for a large region of mutation configuration choices. However, in the region of high p_m and large $1/\eta_m$, the obtained proximity values are essentially unchanged from those obtained for the initial population. This behaviour is slightly more extensive for D2 than D1.

As the number of objectives is increased, the proximity sweet-spots contract for both D1 and D2. Proximity results for DTLZ2(6) are presented for D1 in Figure 5.2c, and for D2 in Figure 5.2d. Some reduction in the sweet-spot is observed for D1, particularly for high p_m and large $1/\eta_m$ configurations. The behaviour is more pronounced in the case of D2: the band of good proximity is noticeably thinner in regions of (i) low p_m coupled with small $1/\eta_m$ and (ii) high p_m coupled with large $1/\eta_m$. In these latter configurations, there is evidence that D2 is producing approximation sets with a worse proximity than that obtained for the initial population of the optimiser.

Further contraction in the proximity sweet-spot is observed as the number of conflicting objectives is continued to be increased. Results for a 12-objective instance of DTLZ2 are shown for D1 and D2 in Figure 5.2e and Figure 5.2f respectively. There is only a small reduction in the sweet-spot for D1, but the sweet-spot for D2 has continued to decrease

substantially. For this latter algorithm, in regions of high p_m and large $1/\eta_m$, the obtained proximity is now considerably worse than that which would be expected from a random sample of 100 solutions to DTLZ2(12).

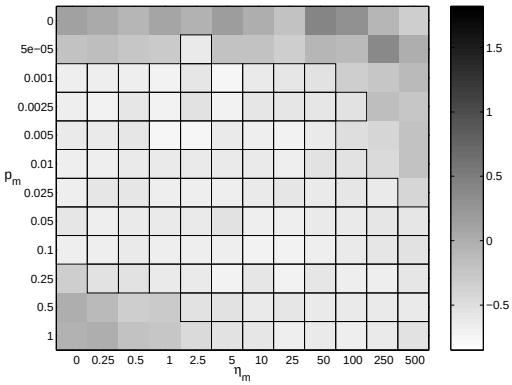
Spread results for the three-objective instance of DTLZ2 are shown for algorithm D1 in Figure 5.3a and for algorithm D2 in Figure 5.3b. The sweet-spot for D1 is limited to a band stretching from configurations of intermediate p_m combined with large $1/\eta_m$, such as $\{p_m = 0.05, \eta_m = 0\}$, to configurations of high p_m coupled with intermediate $1/\eta_m$, such as $\{p_m = 1, \eta_m = 10\}$. In the region where high p_m is combined with large $1/\eta_m$, spread values are larger than unity. This indicates that the approximation set is spread widely through objective-space. In regions of low p_m and small $1/\eta_m$, spread values are close to zero. This indicates that the approximation set represents a region of objective-space that is much smaller than optimal. The sweet-spot is considerably larger for D2, although the latter two observations for D1 also hold in this case. Good spread values are identified in larger regions at each end of the band previously identified for D1.

As the number of conflicting objectives is increased, the regions of good spread contract for both D1 and D2. Results for DTLZ2(6) are shown for D1 and D2 in Figure 5.3c and Figure 5.3d respectively. General thinning of the bands identified for DTLZ2(3) is evident for both algorithms. Large magnitudes of spread can be seen for D2 in regions of high p_m coupled with large $1/\eta_m$.

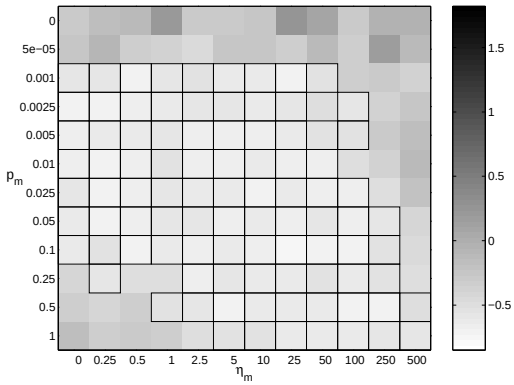
Spread results for the 12-objective instance of DTLZ2 are presented for D1 in Figure 5.3e and for D2 in Figure 5.3f. Further contraction of the sweet-spots is shown for both algorithms, but the level of deterioration is much lower between DTLZ2(6) and DTLZ2(12) than that observed between DTLZ2(3) and DTLZ2(6).

EMO algorithms are required to produce approximation sets with both good proximity and good diversity. From the results in Figure 5.2 and Figure 5.3 it is evident that mutation operator configurations that produce approximation sets of this quality can be identified for both D1 and D2 for all values of M considered. However, the number of such configurations reduces with increasing M for both D1 and D2.

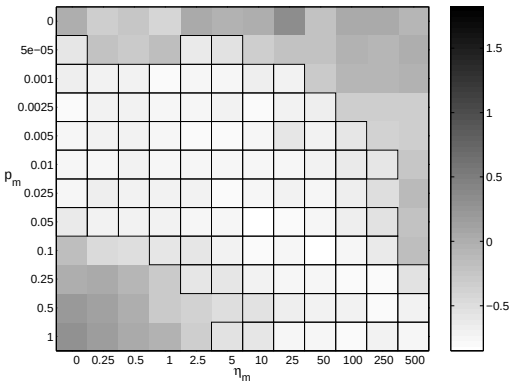
Good configurations for D1 are located in a different area of the map to those for D2. Configurations with relatively high p_m and relatively large $1/\eta_m$, such as $\{p_m = 0.25, \eta_m = 2.5\}$, produce approximation sets with both good proximity and good diversity for D1. However,



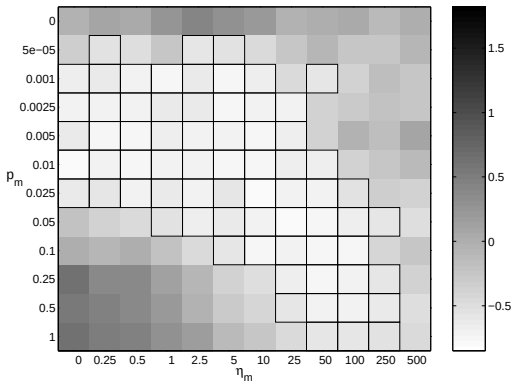
(a) D1: $M = 3$



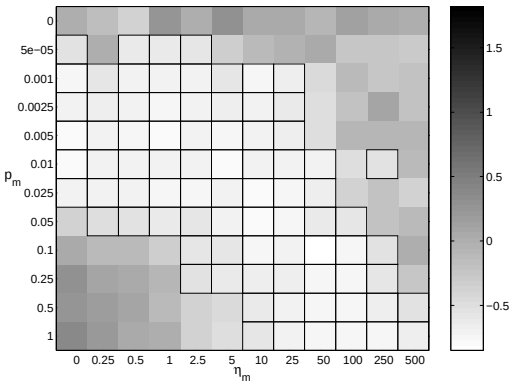
(b) D2: $M = 3$



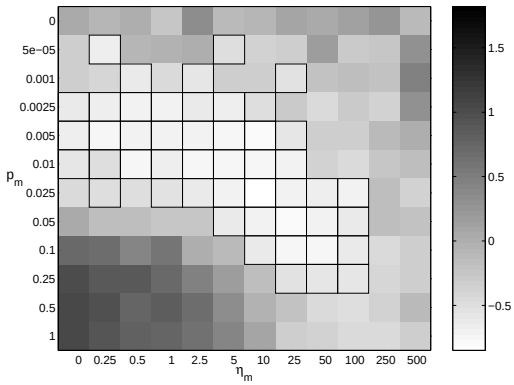
(c) D1: $M = 6$



(d) D2: $M = 6$

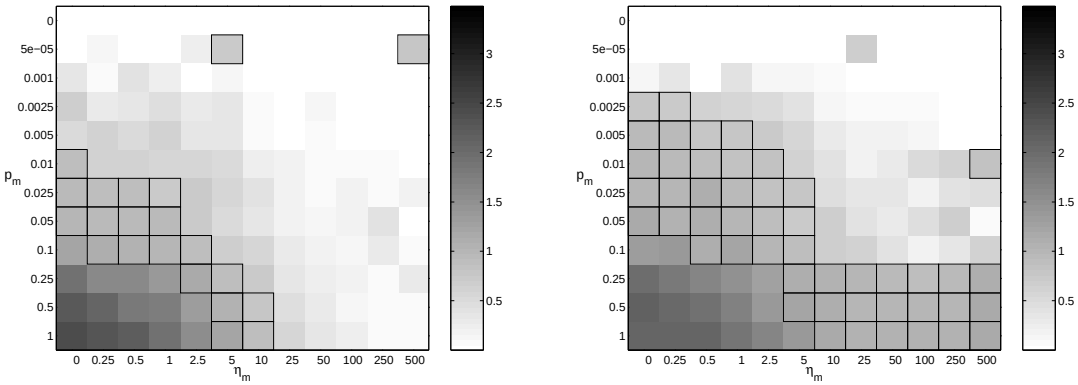


(e) D1: $M = 12$



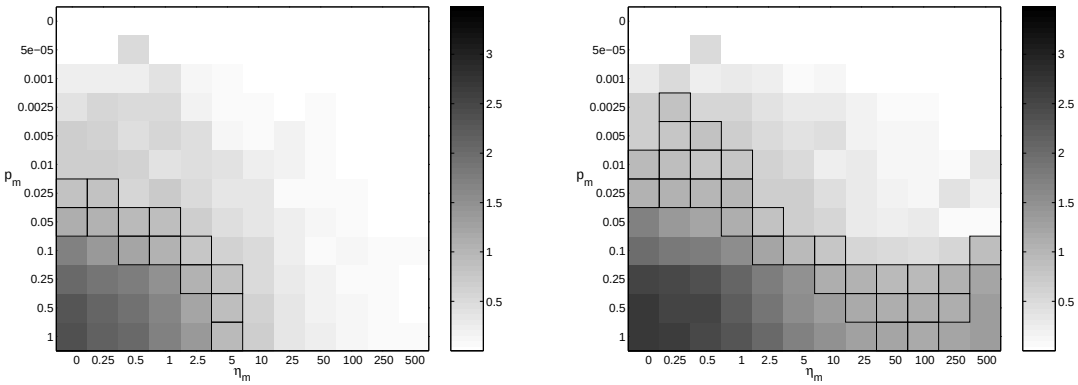
(f) D2: $M = 12$

Figure 5.2: Proximity response maps for mutation



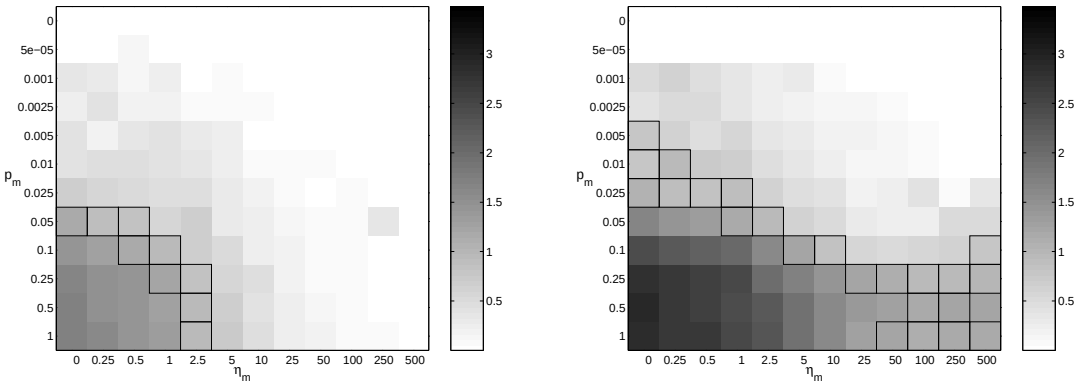
(a) D1: $M = 3$

(b) D2: $M = 3$



(c) D1: $M = 6$

(d) D2: $M = 6$



(e) D1: $M = 12$

(b) D2: $M = 12$

Figure 5.3: Spread response maps for mutation

these settings are associated with poor proximity and large spread values for D2. Good configurations for D2 arise in two areas: (i) intermediate p_m together with large $1/\eta_m$, such as $\{p_m = 0.01, \eta_m = 0.25\}$, and (ii) relatively high p_m coupled with relatively low $1/\eta_m$, such as $\{p_m = 0.25, \eta_m = 100\}$. Note that approximation sets for D1 tend to represent only a very small fraction of objective-space for these configurations.

Polynomial mutation has been used as the variation operator when solving DTLZ2 in previous studies in the literature. Both Deb, Thiele, Laumanns and Zitzler (2002) and Khare *et al.* (2003) used an algorithm identical to D2 with $p_m = 1/n$ and $\eta_m = 20$ (in combination with SBX recombination). From Figure 5.2 and Figure 5.3, it can be seen that this configuration produces approximation sets with good proximity, but only over a very small section of the trade-off surface, for both D1 and D2 when mutation is considered alone.

5.5 Results: Recombination

The proximity response maps for the D1 and D2 algorithms incorporating the SBX recombination operator are shown in Figure 5.4. For the three-objective instance of DTLZ2, a substantial region of good proximity is evident for intermediate to high p_{ic} when combined with intermediate to large $1/\eta_c$. The map for D1 is provided in Figure 5.4a, whilst the equivalent for D2 is shown in Figure 5.4b. The sweet-spot for D2 is seen to extend further into regions of lower p_{ic} and smaller $1/\eta_c$ than that for D1.

As the number of conflicting objectives to be optimised simultaneously is increased, the proximity sweet-spots for both D1 and D2 are observed to decrease. As shown in Figure 5.4c, there is a general contraction in the sweet-spot for D1 for DTLZ2(6). In particular, configurations of high p_{ic} coupled with large $1/\eta_c$ show little improvement in proximity over the initial population. For the six-objective instance of DTLZ2, as indicated in Figure 5.4d, the proximity sweet-spot for D2 is limited to regions of intermediate p_{ic}. Configurations of high p_{ic} combined with intermediate $1/\eta_c$, such as $\{p_{ic} = 1, \eta_c = 5\}$, produce approximation sets with very poor proximity — considerably worse than that obtained for the initial population. This result contrasts sharply with that for D1, where proximity remains good.

The D1 proximity response map for DTLZ2(12) is shown in Figure 5.4e, whilst the corresponding map for D2 is depicted in Figure 5.4f. The sweet-spot for D1 has reduced still further to a region of intermediate p_{ic}. The D2 sweet-spot is slightly larger than that for D1

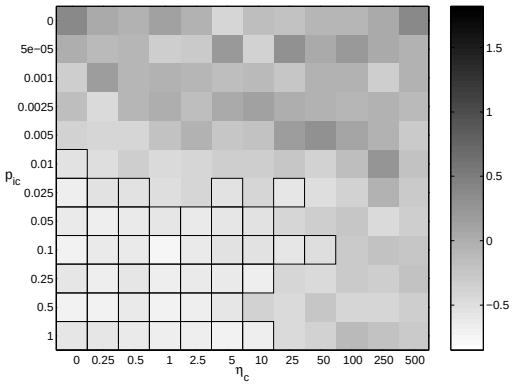
and is located in a region of slightly lower p_{ic}. In regions of high p_{ic}, the proximity results are very poor for D2. The region of divergence identified for DTLZ2(6) has expanded considerably to include expected variation magnitudes in the range from intermediate to high $1/\eta_c$.

Spread response maps for algorithms D1 and D2 optimising the three-objective instance of DTLZ2 are shown in Figure 5.5. As indicated in Figure 5.5a, the spread sweet-spot for D1 encompasses configurations of intermediate to high p_{ic} coupled with intermediate to large $1/\eta_c$. The majority of other SBX configurations produce approximation sets with very small spread measures, thus indicating that the set represents only a very small region of objective-space. D2 offers a more substantial region of good spread performance, as shown in Figure 5.5b. The D2 sweet-spot extends further into configurations of lower p_{ic} and smaller $1/\eta_c$. However, for low values of p_{ic}, spread values are close to zero.

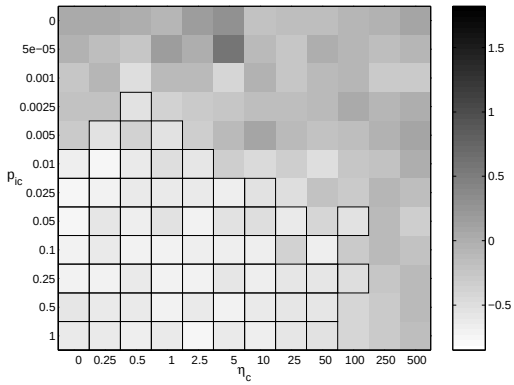
Spread sweet-spots for the D1 and D2 algorithms using recombination contract as the number of conflicting objectives is increased. The response map for D1 solving DTLZ2(6) is shown in Figure 5.5c. Good spread performance is limited to the region of configurations with large $1/\eta_c$ and $p_{ic} \approx 0.25$. Spread values for high p_{ic} together with large $1/\eta_c$ indicate approximation sets that represent a section of objective-space that is larger than optimal. The spread response map for D2 solving DTLZ2(6) is shown in Figure 5.5d. In addition to obtaining good spread for intermediate $1/\eta_c$ together with high p_{ic}, a narrow band of good spread is seen to extend across the range of expected variation perturbation magnitudes. Values of spread that are larger than optimal are observed for higher p_{ic}. Particularly large values are identified for the region represented by $\{p_{ic} = 1, \eta_c = 5\}$. In this instance the approximation set is spread widely throughout non-optimal regions of objective-space. The relationship to poor proximity is clear through comparison with Figure 5.4d.

The spread response map for the 12-objective instance of DTLZ2 is shown for D1 in Figure 5.5e, and for D2 in Figure 5.5f. The sweet-spot for D1 is now highly limited, $\{p_{ic} = 0.25, \eta_c = 2.5\}$ being one such configuration. The band of good spread identified for D2 solving DTLZ2(6) has become thinner for DTLZ2(12). A substantial range of configurations with high p_{ic} indicate approximation sets spread widely throughout non-optimal objective-space.

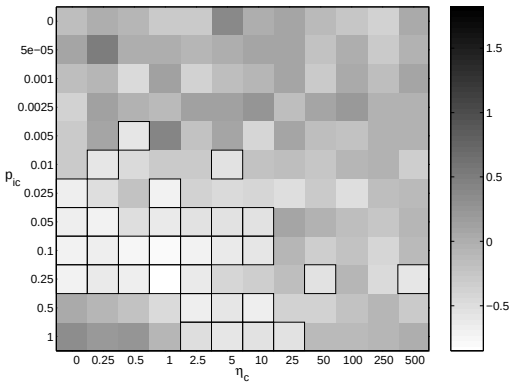
Through considering Figure 5.4 and Figure 5.5 together, recombination operator config-



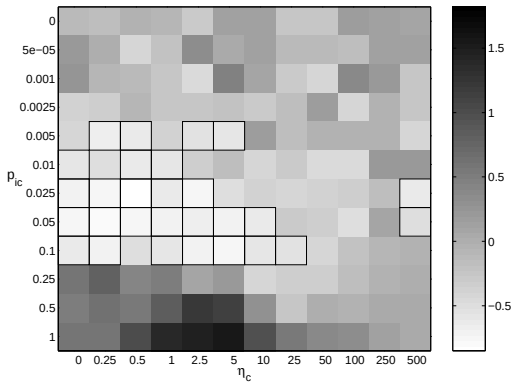
(a) D1: $M = 3$



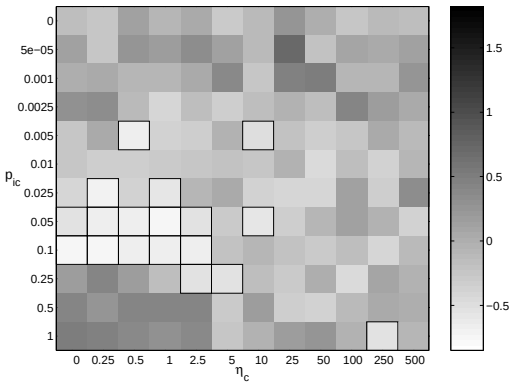
(b) D2: $M = 3$



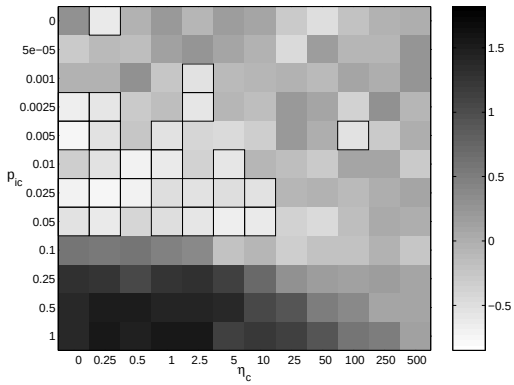
(c) D1: $M = 6$



(d) D2: $M = 6$

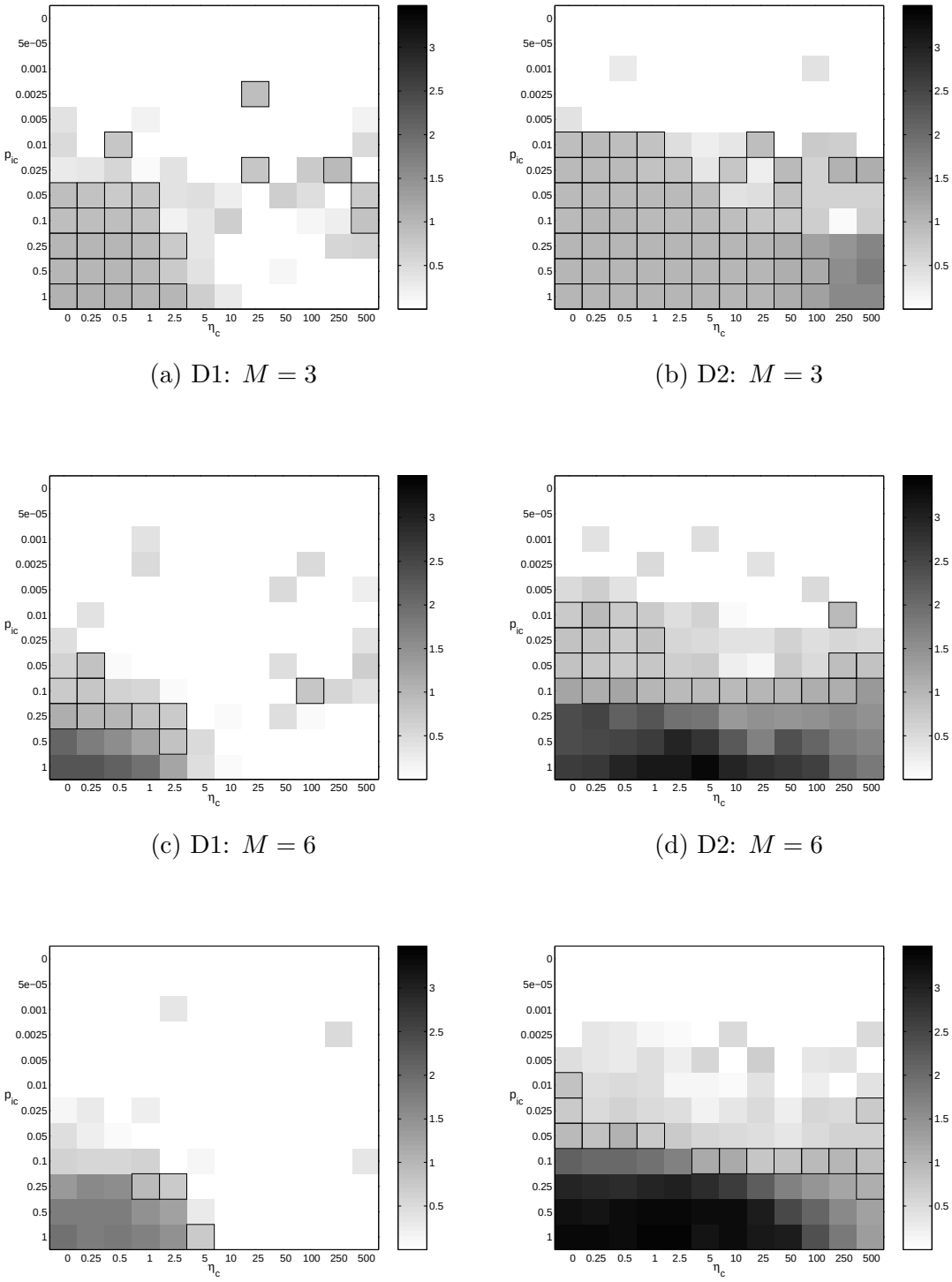


(e) D1: $M = 12$



(f) D2: $M = 12$

Figure 5.4: Proximity response maps for recombination



(a) D1: $M = 3$

(b) D2: $M = 3$

(c) D1: $M = 6$

(d) D2: $M = 6$

(e) D1: $M = 12$

(f) D2: $M = 12$

Figure 5.5: Spread response maps for recombination

urations that produce approximation sets with both good proximity and good spread can be identified for both D1 and D2 for all values of M assessed. For $M = 3$, the sweet-spots for proximity and spread co-locate nicely for both algorithms. In particular, the evidence suggests that a wide range of configuration choices for D2 will lead to approximation sets that are good from the perspective of both axes of performance. As the number of conflicting objectives is increased, the selection of configurations that produce overall good performance decreases for both D1 and D2.

For the six-objective instance of DTLZ2, the regions of overall good performance are reduced for both D1 and D2. The latter algorithm still retains a larger sweet-spot over the former, but many of the configurations are similar. For example, configuration $\{p_{ic} = 0.1, \eta_c = 0.25\}$ produces an approximation set with good proximity and good spread for both D1 and D2. For DTLZ2(12), very few configurations of this nature are evident for either D1 or D2.

Like the polynomial mutation operator, the SBX recombination operator has also been previously used in the D2 algorithm on DTLZ2 (Deb, Thiele, Laumanns and Zitzler 2002, Khare *et al.* 2003). The configuration used in this work was $\{p_{ic} = 0.5, \eta_c = 15\}$ with $p_c = 1$ (the operator was also used in conjunction with polynomial mutation). From the results shown in Figure 5.4 and Figure 5.5 this configuration is demonstrated to be appropriate for the three-objective instance of DTLZ2 for D2. However, as the number of objectives is increased, the settings appear to become increasingly unacceptable. A D2 optimisation of DTLZ2(12) would be predicted to offer an approximation set with poor proximity, with solutions spread widely throughout objective-space.

5.6 Analysis

5.6.1 Introduction

The results in Section 5.4 and Section 5.5 show that, for both algorithms D1 and D2, the region of configuration-space that is associated with high quality approximation sets contracts as the number of conflicting objectives, M, is increased. From the perspective of the proximity indicator, divergence behaviour is observed for some configurations of D2. In these cases, the proximity of the optimised approximation set is, sometimes considerably, worse than the

proximity observed for the initial population. This behaviour is particularly evident for recombination-based D2. It is not observed, in the case of either variation operator, for algorithm D1.

Recall that, as described in the introduction to the algorithms in Section 5.2.2, selection discrimination in D1 is based purely on Pareto dominance considerations. However, in D2, discrimination is based on both dominance *and* density estimation. Thus, analysis of the D2 diversity promotion mechanism provides the key to understanding the behaviour of modern MOEAs under varying M conditions.

5.6.2 Active Diversity Promotion Mechanisms

EMO researchers have previously observed that, for a fixed population size, the proportion of the population that is locally non-dominated increases rapidly with increasing M (Fonseca and Fleming 1998b, Deb 2001a). Empirical results for DTLZ2, for a variety of population sizes and M-objective instances, are shown in Figure 5.6 (where each measurement is the mean of 1000 random sample sets). For a population size of 100, the proportion increases from approximately 0.3 for $M = 3$ to approximately 0.8 for $M = 12$. Note that this is only the *initial state* for the optimisers. Future proportions, as experienced by the optimisers, depend on the effect of repeated applications of s_v, v, and s_s.

D1 and D2 both discriminate in favour of non-dominance at the s_s stage. The effect of this is to quickly increase the proportion of non-dominated solutions in the population, as shown for various M for algorithm D2 in Figure 5.7. For all values of M the proportion is observed to rises rapidly to 1.0, regardless of the initial state. Thus, from the perspective of the selection-for-variation mechanism, the majority of solutions are equivalent for D1. s_v will be random in this case. However, for D2, the secondary diversity enhancement mechanism is now activated.

The operation of the density estimator used in D2 is illustrated by the schematic in Figure 5.8. The estimator assigns larger crowding distances (or, equivalently, smaller densities) to:

- boundary solutions (which are assigned the largest distance found for that objective in the boundary direction),
- other remote solutions, and

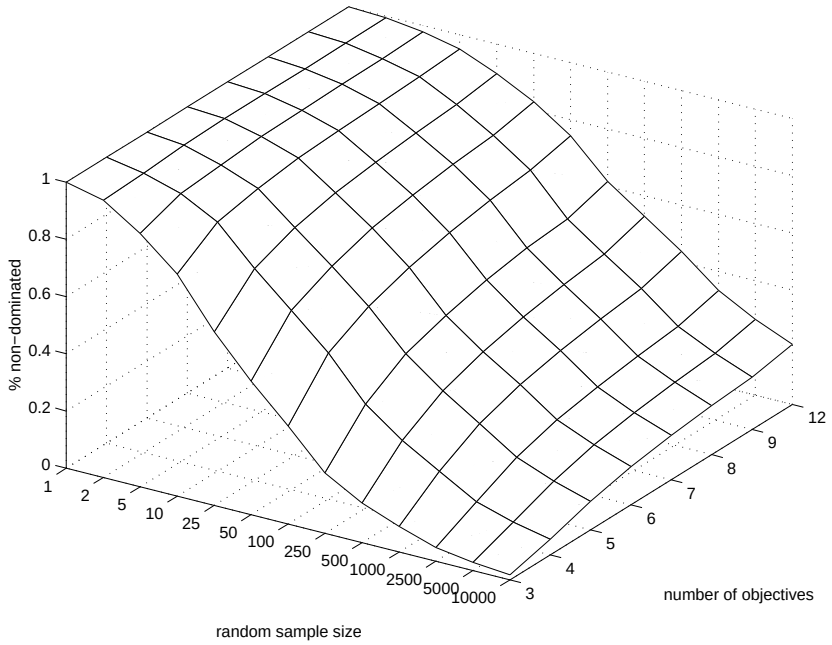


Figure 5.6: Estimated non-dominated proportion of random samples of DTLZ2(M) solutions

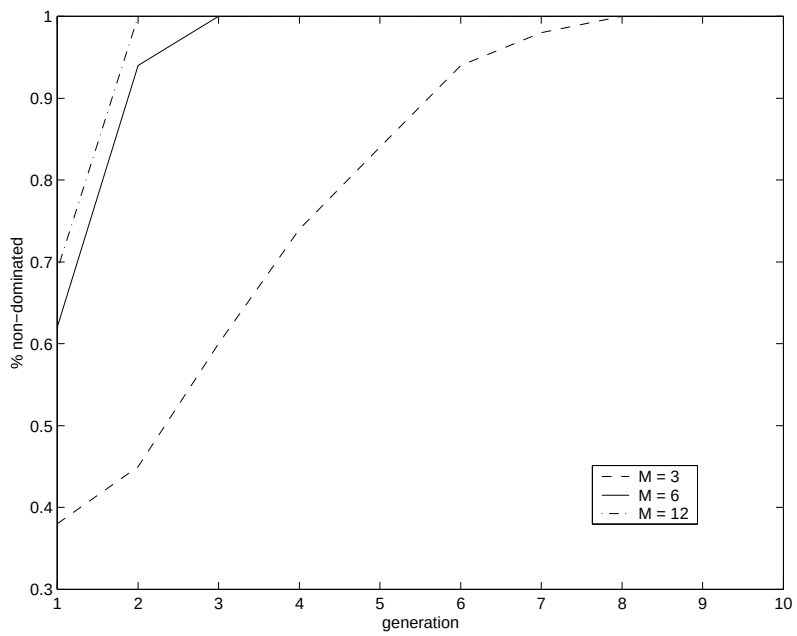


Figure 5.7: Example observed proportion of the population that is locally non-dominated during the initial stages of optimisation via mutation-based D2

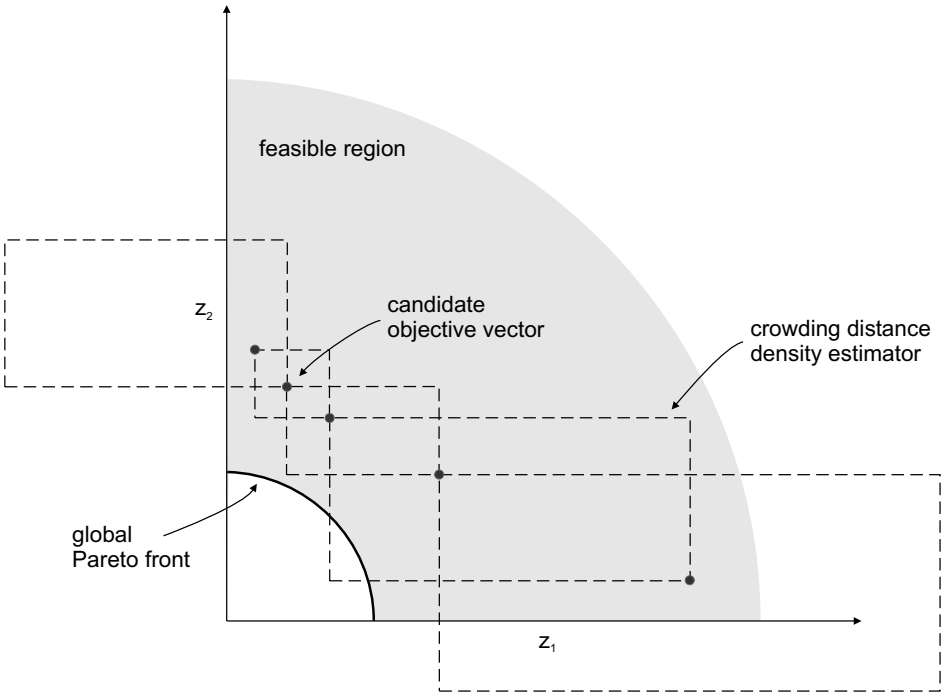


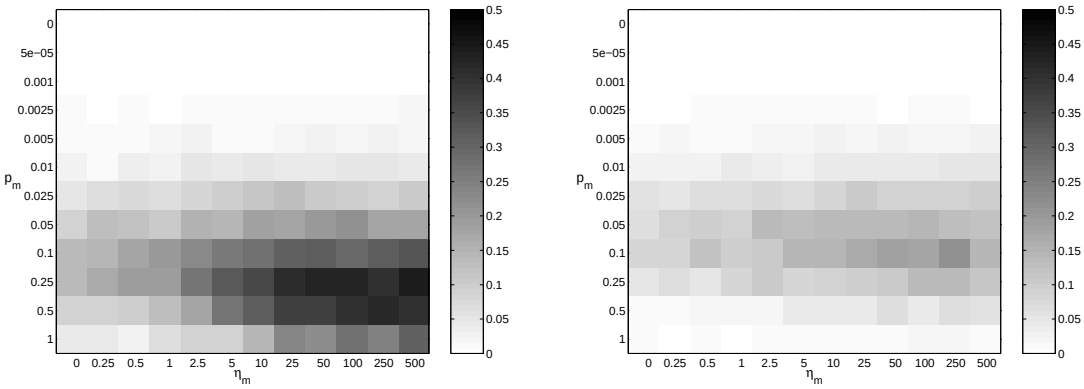
Figure 5.8: Crowding distance density estimator

- immediate neighbours of remote solutions.

The volume of objective-space increases exponentially as the number of conflicting objectives is increased linearly. For a fixed population size, this provides more opportunity for a solution to be remote from others, be distant from the global trade-off surface, and yet still be locally non-dominated. In these circumstances, active diversity promotion in D2 will bias the search toward solutions with poor proximity to the global Pareto front.

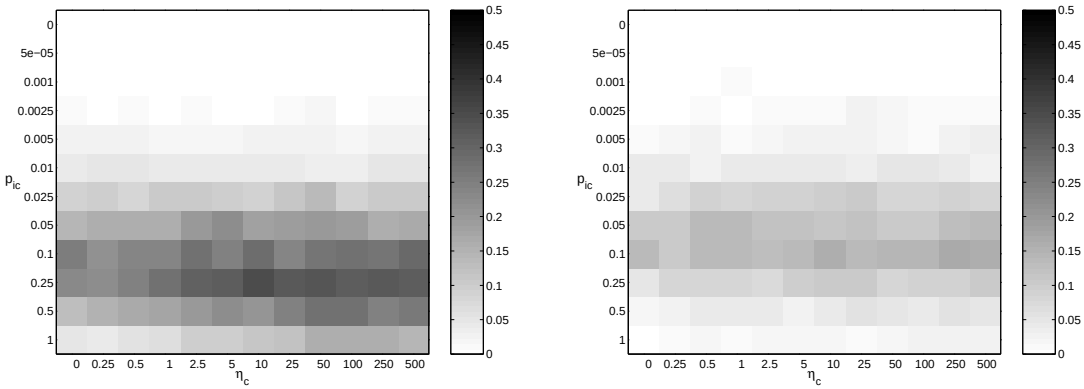
The above selective bias will present difficulties if the variation operators have a low expectation of success (defined in Section 5.3.4 as the probability of a child dominating its parent or parents). Success rates for both recombination- and mutation-based D2, for the three- and twelve-objective instances of DTLZ2, are shown in Figure 5.10. Note that the results for D1 are very similar and are thus not reproduced herein. The results for both variation operators show that success rates tend to decrease with increasing M. Therefore, one of the factors required to explain the difficulties encountered by diversity-promoting MOEAs is shown to become more apparent with increasing M.

If operator success rates are low and variation operators produce a high proportion of



(a) Mutation: $M = 3$

(b) Mutation: $M = 12$



(c) Recombination: $M = 3$

(d) Recombination: $M = 12$

Figure 5.9: Proportion of children that dominate their parents for D2

children that are dominated by their parents then the search is expected to simply stagnate to solutions already discovered. This is because the primary dominance-based comparisons will bias against the child solutions during s_s. However, the D2 observations in Section 5.4 and Section 5.5 do not suggest stagnation of the search: rather, they suggest a directed search away from regions of relatively good proximity, toward an approximation set with relatively good diversity. This behaviour can be traced to the phenomenon of *dominance resistance*.

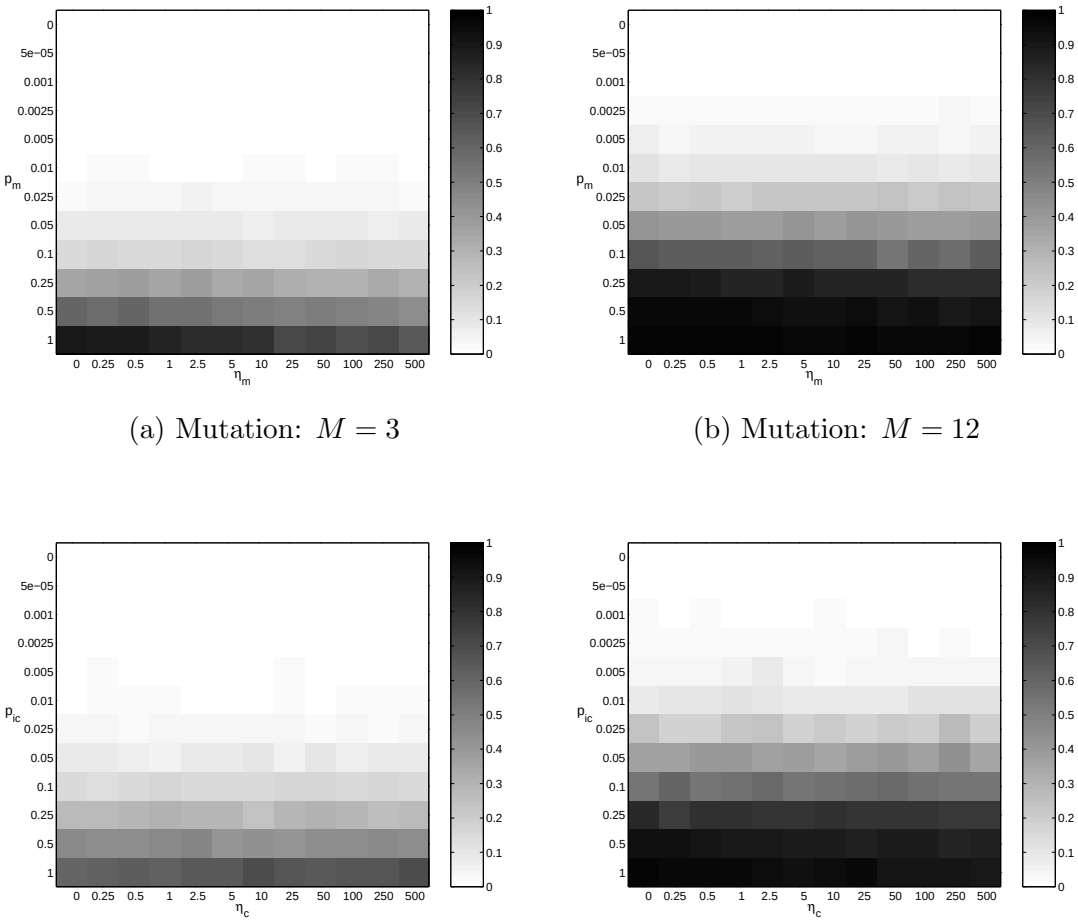
Dominance resistance was first identified, and introduced as terminology, by Ikeda, Kita and Kobayashi (2001) in the context of real-parameter representations and associated variation operators for a specific class of multi-objective problems. Deb, Thiele, Laumanns and Zitzler (2002) also encountered this behaviour for a set of real-parameter *constraint surface* tasks that are scalable in the number of conflicting objectives. These researchers also identified that the level of dominance resistance could increase with the dimension of objective-space.

Response maps that show the proportion of children that are non-dominated but are not equal to their parents are provided in Figure 5.10. For DTLZ2(3), a large proportion of such solutions is indicated for high operator application rates for both mutation and recombination (Figure 5.10a and Figure 5.10c respectively). As the number of objectives is increased, depicted in Figure 5.10b and Figure 5.10d, the high proportion of non-dominated children is observed to extend much further toward lower variation application probabilities. This empirically demonstrated dominance resistance supports the heuristic notion that a solution has more opportunity to be locally non-dominated as the number of dimensions of objective-space is increased.

5.6.3 Inactive Diversity Promotion Mechanisms

For configurations corresponding to a low probability of variation operator application (the upper regions of the response maps presented in this chapter), a large proportion of the children produced during variation are identical copies of their parents. Empirical results for D2, for both mutation and recombination, are shown in Figure 5.11. Note, again, that the results for D1 are very similar to these and are thus omitted.

Since the density estimator used in D2 is a variant of first-nearest-neighbour, and equality in decision-space corresponds to equality in objective-space for the deterministic DTLZ2,



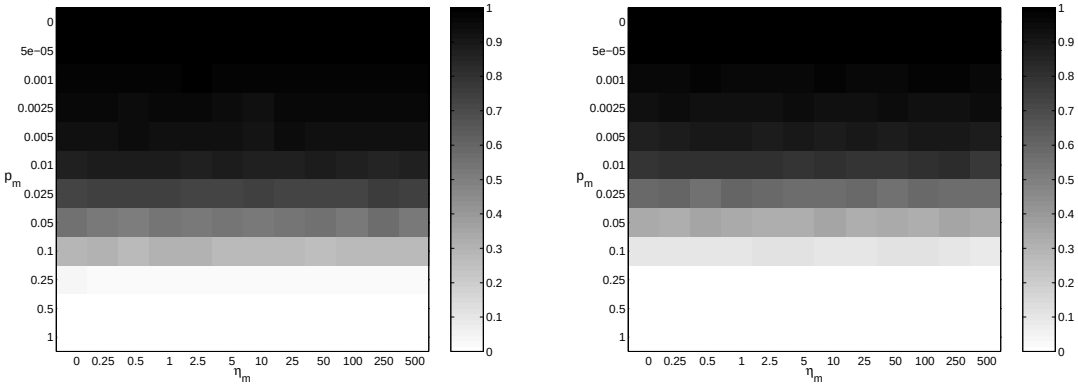
(a) Mutation: $M = 3$

(b) Mutation: $M = 12$

(c) Recombination: $M = 3$

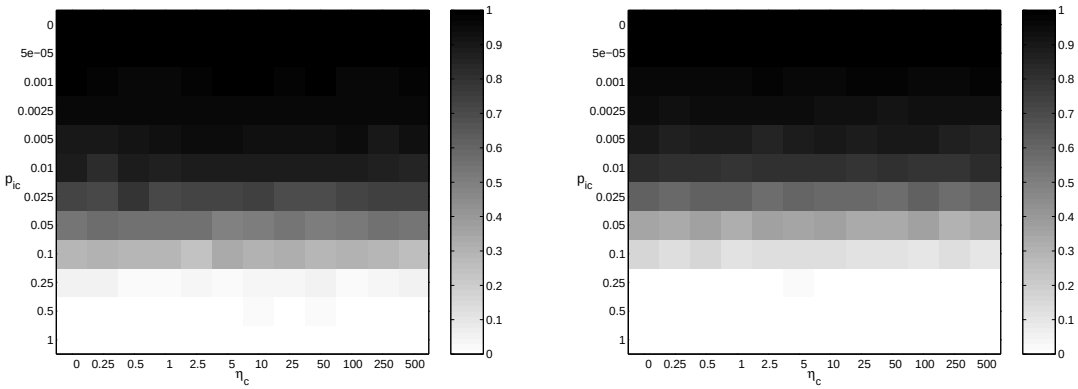
(d) Recombination: $M = 12$

Figure 5.10: Proportion of children that are non-dominated (but not equal) to their parents for D2



(a) Mutation: $M = 3$

(b) Mutation: $M = 12$



(c) Recombination: $M = 3$

(d) Recombination: $M = 12$

Figure 5.11: Proportion of children that are equal to their parents for D2

solutions with copies will have the maximum possible density estimate (corresponding to a crowding distance of zero). For configurations of low p_{ic} or p_m, this neutralises any density-dependent selection mechanisms because the densities of most solutions are identical. Thus, for the case of equivalence under dominance, selection is entirely random for both D1 and D2. Thus, little difference is evident between these two algorithms under the aforementioned configurations. This can be verified by comparing the upper region of the D1 response maps in Figures 5.2, 5.3, 5.4, and 5.5 to the corresponding D2 equivalents.

5.6.4 Comparison Between Mutation and Recombination

There are clear differences between the response maps obtained for mutation in Section 5.4 and those for recombination in Section 5.5. Many of the differences are spatial and can be attributed to subtle variations in the interaction between the operator mechanisms and the DTLZ2 problem landscape. For example, configurations with high success rates for SBX are located in a slightly different area of the map to those for polynomial mutation. However, the fundamental difference between the mutation and recombination results is that the divergence behaviour (measured in terms of proximity) is more severe in the recombination-based D2 than the mutation-based equivalent.

The key distinction between the two operators is that the e-e trade-off setting is fixed for a particular configuration of polynomial mutation, but is dynamic over the course of the optimisation for SBX recombination due to the self-adaptive nature of the SBX distribution. Different e-e settings can occur within the process of creating a single pair of child solutions for the latter operator. The fundamental difference between the mutation and recombination results can arguably be attributed to this difference between the operators.

As explained in Section 5.2.3, in the context of a single decision variable, if the decision variable values of the two parents are closer together then the expected SBX variation magnitude is smaller than if the two parent values are further apart. If selection is (effectively) biased toward poor proximity values, as discussed in Section 5.6.2, then convergence variable values (see Section 5.3.2) will be clustered in regions that are associated with poor proximity. The large proportion of parent values in these regions thus allows the self-adaptive SBX operator to focus further on these regions (since non-diverse parent material leads to more localised exploration). However, the diversity-promotion selection mechanisms in D2

also ensure that genetic diversity remains intact for the distribution variables of DTLZ2. Thus, large-scale exploration is still possible for these variables. So SBX is capable of simultaneously performing exploitation of poor proximity and exploration of diversity, leading to approximation sets spread widely in objective-space with very poor overall proximity.

5.7 Effect of Population Size

5.7.1 Introduction

Deb (2001a) proposes the use of large population sizes as a potential method for achieving good many-objective optimisation results, since this will reduce the proportion of non-dominated solutions in the population and thus provide improved Pareto-based discrimination.

In practice, the use of large population sizes is often prohibitive in real-world applications because of the computational resources required to evaluate and process potential solutions. Thus in many applications, such as that documented by Thompson, Chipperfield, Fleming and Legge (1999), the population size is generally limited to fewer than 50 individuals.

The often high cost of solution evaluation is a serious issue for computational search and optimisation techniques in general, rather than just affecting evolutionary algorithms. As a result, a research field known as *meta-modelling* has arisen that is devoted to the development and deployment of approximation models in solution evaluation (Bull 1999). A related topic that is specific to evolutionary algorithms, known as *fitness inheritance*, also considers how to reduce evaluation requirements (Smith, Dike and Stegmann 1995).

Given the above qualifications, the identification of the benefits — in terms of approximation set quality — that can be obtained for larger population sizes remains a matter of interest. In this section, the six-objective experiments in Section 5.4 and Section 5.5 are repeated for alternative population sizes (to the value of 100 used in those sections) to produce new results for algorithms D1 and D2 using both mutation and recombination. At present, population sizes for generational MOEAs tend to lie between approximately 10 and 1000 (Coello *et al.* 2002). Thus, population sizes of three orders of magnitude — 10, 100, and 1000 — are considered in this section. The results for mutation are presented in Section 5.7.2, followed by the results for recombination in Section 5.7.3. An analysis of the observations is

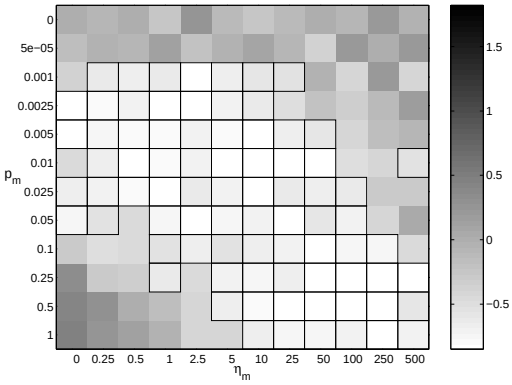
offered in Section 5.7.4.

5.7.2 Mutation Results

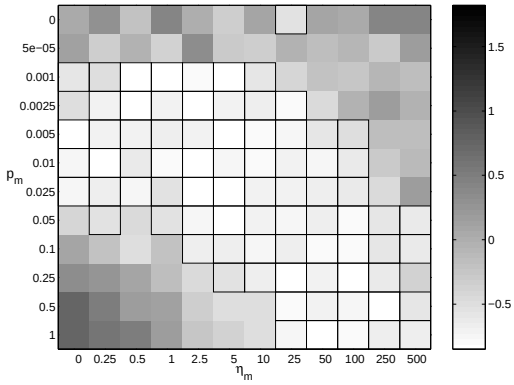
The proximity response maps for the population-sizing study are shown in Figure 5.12. The D1 response maps are presented in the right-hand column of the figure, whilst the D2 maps are shown in the left-hand column. The upper row of maps, Figure 5.12a and Figure 5.12b were obtained for a population size of 10. The maps in the middle row, Figure 5.12c and Figure 5.12d, are for a population size of 100 and have previously been presented in Section 5.4. The lower row of maps, Figure 5.12e and Figure 5.12f, show the proximity results for a population size of 1000. Little variation is evident in the results obtained for either D1 or D2 as the population size is increased. Thus, the relationship between D1 and D2 identified in Section 5.4 and analysed in Section 5.6 remains consistent across the range of population sizes generally used in a generational MOEA implementation.

The spread response maps are presented in Figure 5.13. The layout of this figure corresponds to that of the proximity results described above. Spread is generally seen to improve with increasing population size for both D1 and D2. In map regions of low p_m and low $1/\eta_m$ the spread values are close to zero (especially for D1) for small population sizes. However, the spread value is observed to increase toward the optimal value of unity for these configurations as the population size is increased. In intermediate map regions, and particularly for configurations of low p_m coupled with high $1/\eta_m$, the sweet-spot for spread is seen to dilate. However, in regions of high p_m combined with large $1/\eta_m$ the spread values become increasingly too large for both D1 and D2. The overall relationship identified in Section 5.4 between D1 and D2 is seen to be preserved: D2 has superior spread in general, but values can be too large in the aforementioned region.

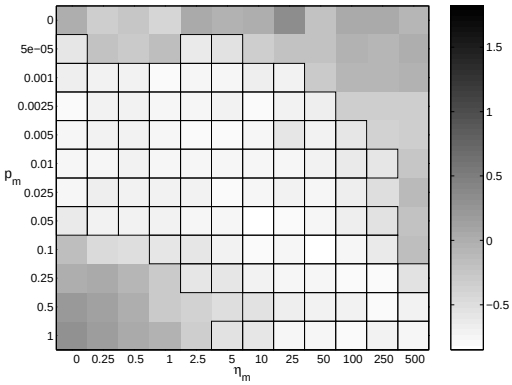
In terms of meeting the dual aims of both good proximity and good spread, the number of such configurations for mutation is seen to increase for both D1 and D2 with increasing population size. Also, more such configurations exist for D2 than D1 for each population size. Note that, even for a population size as small as 10 individuals, both algorithms are capable of producing an approximation set of good overall quality (in the terms defined by the inquiry).



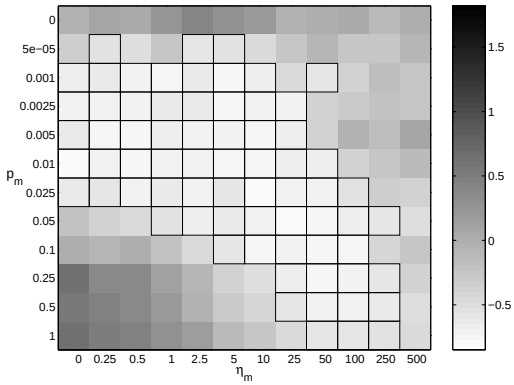
(a) D1: population size = 10



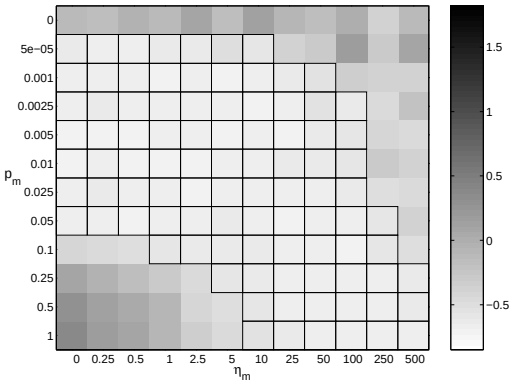
(b) D2: population size = 10



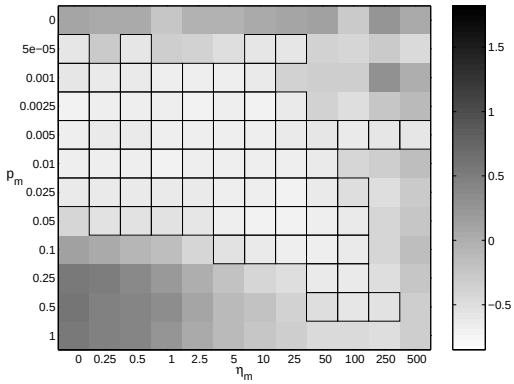
(c) D1: population size = 100



(d) D2: population size = 100

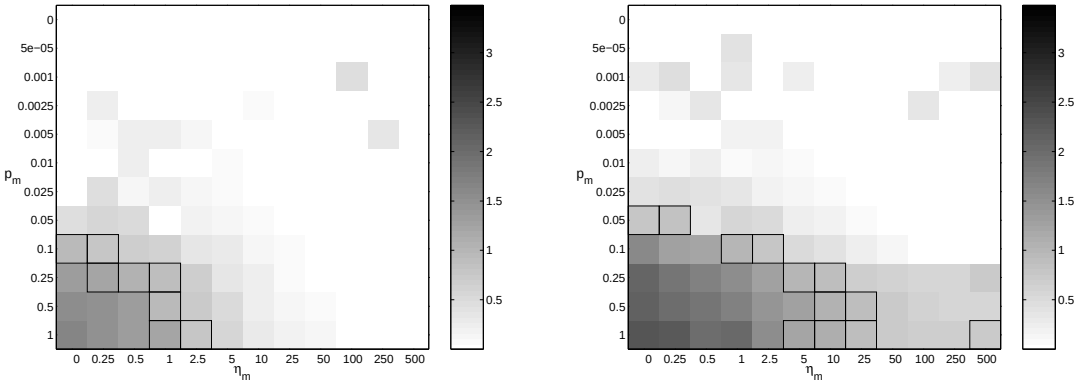


(e) D1: population size = 1000



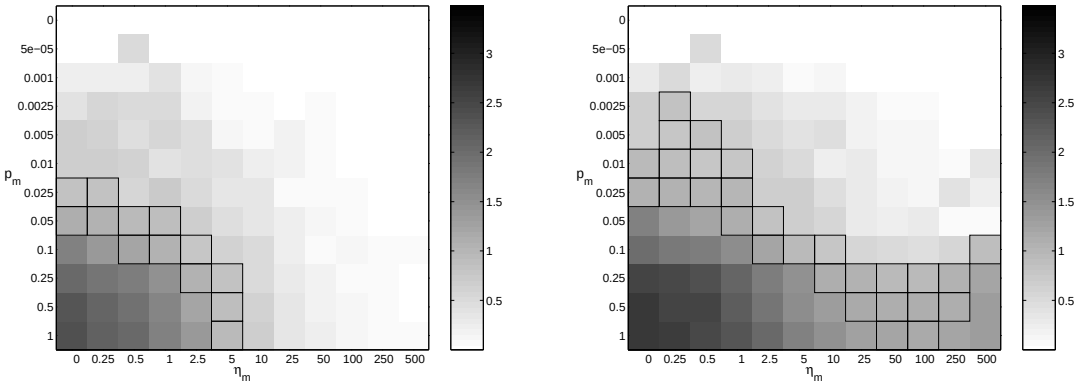
(f) D2: population size = 1000

Figure 5.12: DTLZ2(6) proximity response maps for mutation for various population sizes



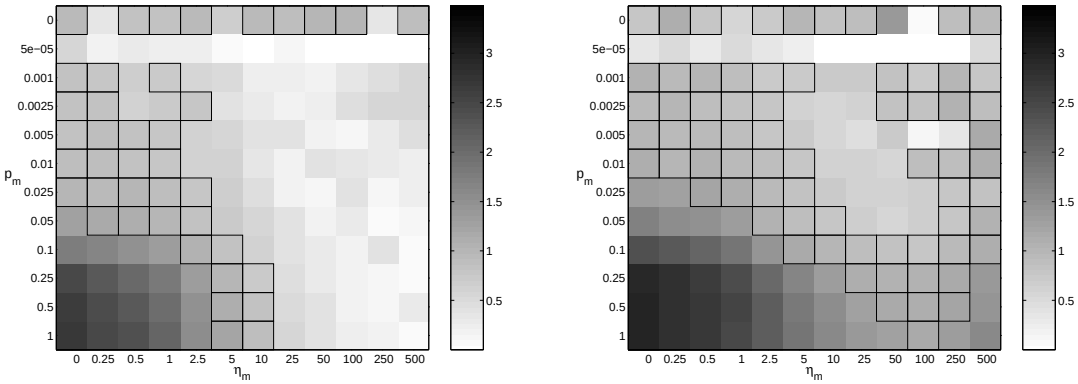
(a) D1: population size = 10

(b) D2: population size = 10



(c) D1: population size = 100

(d) D2: population size = 100



(e) D1: population size = 1000

(f) D2: population size = 1000

Figure 5.13: DTLZ2(6) spread response maps for mutation for various population sizes

5.7.3 Recombination Results

The proximity response maps for recombination-based implementations of D1 and D2 are shown in Figure 5.14. The presentation of the results is equivalent to that described for mutation in Section 5.7.2. The D1 and D2 maps for a population size of 10 are shown in Figure 5.14a and Figure 5.14b respectively. Very little convergence toward the global surface is evident for either algorithm, although a few configurations do attain the benchmark for good performance of a 0.5 unit improvement in proximity. Divergence behaviour is observed in D2 for configurations of high p_{ic}. When the population size is increased to 100, a proximity sweet-spot is generated for configurations of intermediate p_{ic} combined with intermediate to large values of $1/\eta_c$ for both algorithms as shown in Figure 5.14c and Figure 5.14d. In the region of high p_{ic} together with intermediate $1/\eta_c$, good proximity is observed for D1 but poor proximity is evident for D2. When the population size is increased still further to 1000, as shown in Figure 5.14e, the proximity sweet-spot for D1 expands further to include approximately half of the configurations considered. In the corresponding case for D2, presented in Figure 5.14f, the proximity sweet-spot is also seen to expand. The region of poor proximity for high p_{ic} is still seen to exist, but the size is more restricted than for smaller population sizes.

The spread response maps for the population sizing study are presented in Figure 5.15. Spread is close to zero for almost all configurations for D1 with a population of 10 individuals, as shown in Figure 5.15a. The D2 map shown in Figure 5.15b also indicates many such configurations, but also includes a region of larger spread for high p_{ic}. When the population size is increased to 100, more configurations produce improved spread values for both D1 and D2 as shown in Figure 5.15c and Figure 5.15d respectively. The sweet-spot for D2 is more extensive than that for D1. This relationship is retained for a population size of 1000, with general improvements for both algorithms over the 100-individual implementations. The 1000-individual maps are provided in Figure 5.15e and Figure 5.15f. The region of very large spread for high p_{ic} is somewhat more restricted for D2 with a population size of 1000, reflecting the improvements previously witnessed for proximity.

There are no recombination configurations (of those studied) for a population size of 10 for which approximation sets can be obtained with both good proximity and good spread. This result is true for both algorithm D1 and algorithm D2. As the population size is increased,

the number of configurations that relate to overall good quality increases considerably from this parlous state for both D1 and D2.

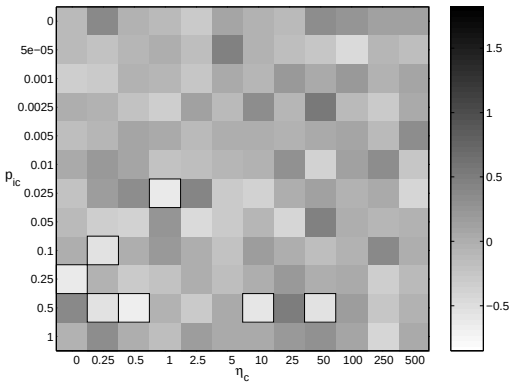
5.7.4 Analysis

The proximity indicator results were demonstrated to be largely invariant of population size for the mutation-based optimisers. However, improved results were obtained by increasing the population size for the recombination-based optimisers. This difference is argued to be attributable to the exploration-exploitation mechanisms in the two operators as described below.

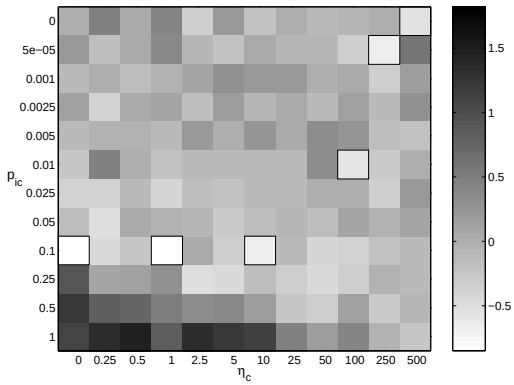
Polynomial mutation is a single-parent variation operator. Therefore its search capability, when considered in isolation from selection mechanisms, is independent of the properties of the remainder of the population from which the parent is drawn. The exploration-exploitation setting is entirely determined by parameters defined by the analyst: the probability of application, p_m, and the expected magnitude of variation (relative to the problem landscape), η_m. Hence, the convergence performance obtained for each configuration setting is largely equal for each population size setting of a particular algorithm.

SBX recombination, by contrast, is a two-parent operator in which the diversity in the genetic material of the parents is crucial to operator behaviour. If the parents contain identical genetic material then the operator has zero exploratory capabilities: the children produced are identical to the parents. Greater distinction between the genetic material of the parent solutions provides greater exploratory capabilities. The level of genetic diversity available is related closely to population size: the greater the number of candidate solutions randomly generated in the initial population, the greater will be the amount of material available. Thus, for small population sizes, there is often insufficient diversity available in decision-space for SBX exploration and consequently the search stagnates. The e-e trade-off setting in SBX is partially specified by the configuration parameters p_{ic} and η_c akin to those of polynomial mutation, but is also dependent on population diversity. Thus, the improved results for large population sizes are understandable.

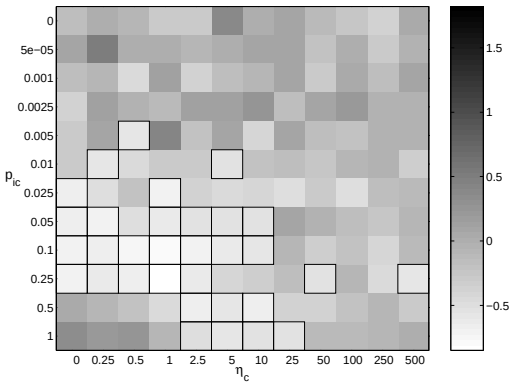
The spread indicator results can also be argued to be attributable to the richness of decision-space sampling inherent to larger population sizes. For a larger population, there is a greater probability of sampling any given area of the search space and of potentially making



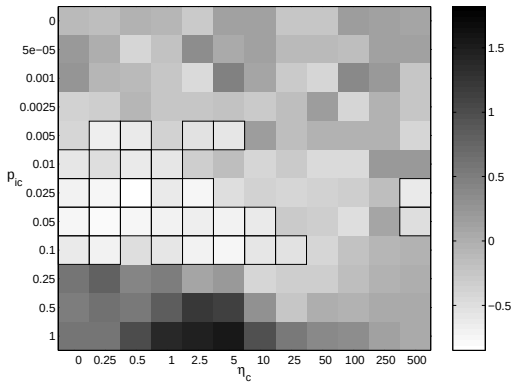
(a) D1: population size = 10



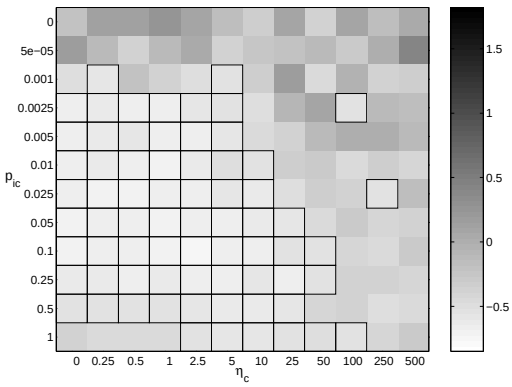
(b) D2: population size = 10



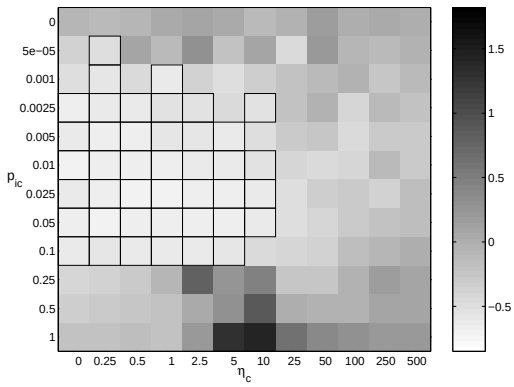
(c) D1: population size = 100



(d) D2: population size = 100

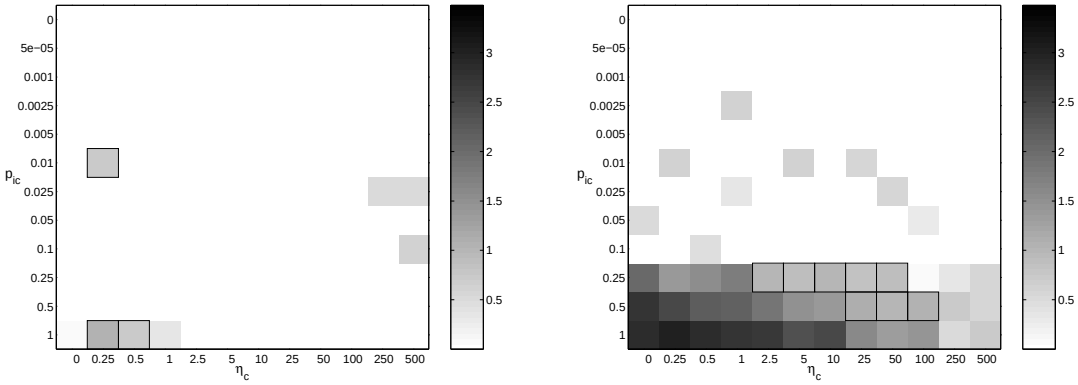


(e) D1: population size = 1000



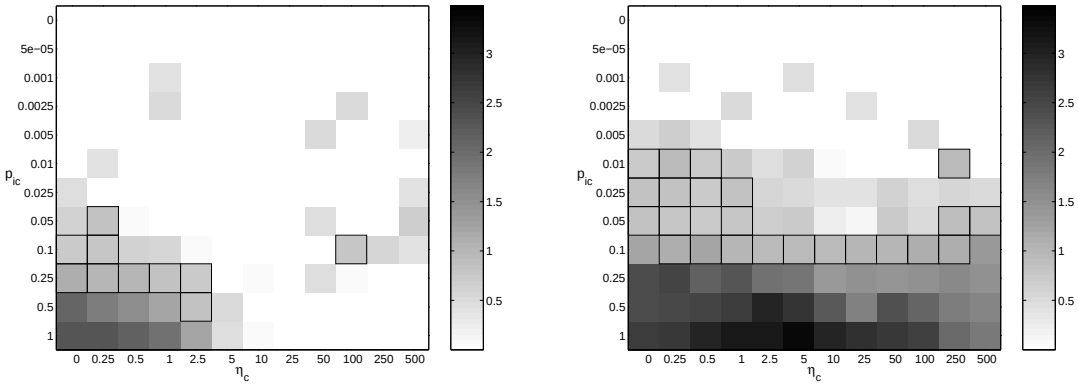
(f) D2: population size = 1000

Figure 5.14: DTLZ2(6) proximity response maps for recombination for various population sizes



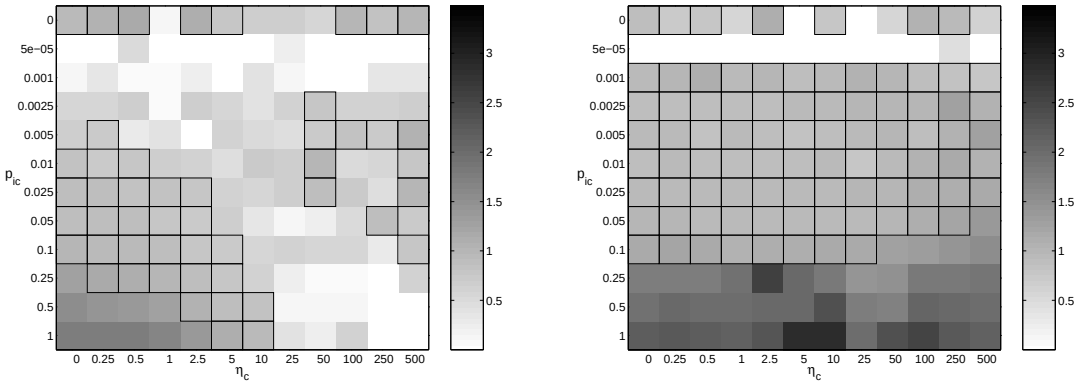
(a) D1: population size = 10

(b) D2: population size = 10



(c) D1: population size = 100

(d) D2: population size = 100



(e) D1: population size = 1000

(f) D2: population size = 1000

Figure 5.15: DTLZ2(6) spread response maps for recombination for various population sizes

multiple samples from one region. Thus, the spread is likely to be naturally superior for a larger population, and the loss of one promising solution through sampling errors is less likely to result in a region of interest becoming completely unrepresented. This argument applies to both variation operators considered in the inquiry. Also, as explained above, SBX requires diversity to facilitate exploratory behaviour. Thus, initial diversity is required in order to search for further diversity.

5.8 Summary

This chapter has shown how the behaviour of MOEAs can change dramatically with the number of conflicting objectives to be optimised. In particular, the behaviour observed for an algorithm configuration for a small number of objectives *cannot* be generalised to an arbitrary (larger) number of objectives.

Through consideration of algorithm processes at a component level, the diversity-promoting selection mechanisms have been identified as highly influential to optimiser outcome. Of the algorithms considered, only those that bias selection in favour of solutions with a low density estimate have been found to exhibit final approximation sets with a worse overall proximity value than that calculated for the initial population.

Analysis suggests that several factors are required for this divergence behaviour to occur:

- Inclusion of a diversity promotion selection mechanism.
- Activation of the diversity promotion mechanism, requiring (i) selection to be based on density estimates and (ii) the estimator to be suitably configured to provide discrimination between solutions.
- Dominance resistance solutions: the variation operators produce a large proportion of children that are non-dominated with respect to their parents.
- An objective-space of suitable volume and dimensionality, in which solutions can be spatially remote, poor in terms of absolute proximity, and yet still be non-dominated.

These results suggest that active diversity promotion can pose a serious challenge to obtaining an approximation set with good proximity to the global trade-off surface. But such a mechanism is positively required in EMO, since without it an EA will tend to experience

genetic drift and thus converge on to a smaller region of objective-space than required. Diversity promotion is a crucial process in an MOEA, but its implementation requires some care.

In terms of the performance differences observed between the optimisers when using the alternative variation operators, the behaviour under mutation appears to vary less with different numbers of objectives and population sizes than the behaviour under recombination. Thus, mutation may be favoured because performance can be more easily predicted. However, based on the limited evidence available, the best configurations for recombination appear to offer superiority in terms of both proximity and spread to their counterparts for mutation. This observation is likely to have arisen because recombination can adapt the variation step-size to the order of magnitude required for continued improvements. Note that this property has also been suggested as the factor responsible for the very poor proximity values generated for some configurations of recombination when optimising more than a small number of objectives.

In the introduction to this chapter, it was noted that a previous study found that PESA produced fundamentally different results to NSGA-II and SPEA2 but that the reasons for this discrepancy were not clear (Khare *et al.* 2003). Based on the conclusions above, it is now possible to hypothesise that the density-based selection in PESA was *inactive* due to the parameter settings chosen for the density estimator and thus the poor proximity behaviour observed for NSGA-II and SPEA2 (where such selection was active) was not replicated. The genetic drift observed for PESA in Khare *et al.*'s (2003) study provides further support for this argument. In essence, the general inability of the estimator to discriminate between solutions would cause the D2-type PESA algorithm to exhibit D1-type behaviour.

From a real-world perspective, the results of this inquiry provide some grounds for optimism. Real-world EMO applications have been undertaken since the early days of EMO development, regardless of the lack of theoretical support for this work. Now, substantial evidence has been collected to show that — for some configurations — standard MOEAs *are* capable of producing approximation sets that satisfy the dual aims of good proximity and good distribution (in an absolute sense, rather than being purely satisfactory from the perspective of the DM) when simultaneously optimising a large number of conflicting objectives.

Chapter 6

Independence in E$\mathbb{M}$O: Effects and Innovations

6.1 Introduction

6.1.1 Problem Decomposition

Decomposing a global problem into a hierarchy of smaller, tractable, sub-problems is a familiar pan-disciplinary concept. The processes of problem decomposition, solution, and recomposition are collectively known as a *divide-and-conquer* (DC) approach. Following the notation of Watson (2002), a *decomposable* problem is defined as one for which interactions arise between various sub-components of a problem. If no interactions occur between sub-problem elements then the problem can be described as *separable*. Sometimes the dependencies exist but are sufficiently insignificant for a separable approach to be acceptable, although this approach concedes that the ultimate solution attained may not be globally optimal.

In Chapter 4 it was argued that treatments for the evolutionary optimisation of large numbers of objectives can potentially be discovered through analysis of the relationships between the objectives. This chapter considers the concept of *independence* between a pair of objectives, in which performance in each objective is entirely unrelated to performance in the other. Thus, in the terminology described above, the objectives are separable.

An *independent collection* (IC) is herein defined as a grouping whose members are linked by dependencies, and for which no dependencies exist with elements external to the group.

Consider a problem with ψ ICs of objectives $[\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\psi]$ and associated ICs of decision variables $[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\psi]$. If knowledge of these collections is available then the global problem, $\mathcal{P}$, can be decomposed into a group of parallel sub-problems $[\mathcal{P}_1, \ldots, \mathcal{P}_\psi]$ that can be optimised independently of each other to ultimately yield ψ independent trade-off surfaces. The benefits of adopting such an approach are investigated in this chapter.

6.1.2 Decomposition-Based Evolutionary Algorithms

Several evolutionary computing methodologies have been proposed to exploit the possibility of problem decomposition. Watson (2002) proposes the following classification:

Implicit: The structure of the problem is addressed by the variation (or equivalent) operators. For example, in the estimation of distribution algorithms detailed in Section 2.3.3, new solutions are sampled from a structured probability model to provide a more efficient search.

Explicit: The division of available resources is explicitly defined for the family of sub-problems. Thus, each component of the problem is distinctly evolved.

The latter approach is of special interest because of its links to biological concepts (Watson 2002). Each decomposed group of decision variables can be viewed as a *species*, with the association of the species to form the complete solution emulating the concept of *symbiosis*. This notion is embedded within the EA approach known as *cooperative coevolution* (Potter and De Jong 2000). This is a multi-population approach, in which each species is represented by a particular sub-population.

Two of the key issues in cooperative coevolution are (i) the development of *collaboration* mechanisms for the *composition* of a complete solution and (ii) the method of assigning fitness to individual components (the so-called *credit assignment* problem). Many different techniques have been proposed. The most interesting, from a multi-objective perspective, is *Pareto coevolution* in which the relative performance of a component (against others from within its sub-population) is established in a dominance sense using the collaborators from other species as the performance axes (Noble and Watson 2001).

A further desirable property of a divide-and-conquer mechanism is the ability to perform the decomposition on-line with the minimum of *a priori* knowledge. This property is known

in the EA field as *emergence* (Watson 2002). The decomposition is often based on heuristics in the particular field of application. For example, Valenzuela and Jones (1994) used such techniques in an evolutionary divide-and-conquer approach to large *travelling salesman* problems (TSPs). As described in Section 2.3.3, EDAs develop a model of the dependencies (and, hence, independencies) in the structure of decision-space. This can naturally be used to indicate the required decomposition for explicit DC schemes. This type of approach is discussed and developed by Watson (2002). Tiwari and Roy (2002) considered neural networks, probability models and regression analysis to determine dependency structures in decision-space. Tree diagrams and direct analysis were also used to highlight separable decision variables.

Note that all of the above EA schemes involve the decomposition of decision-space only. Prior to the work documented in this chapter, and published as Purshouse and Fleming (2003a), no research was identified for which objective-space had been subject to decomposition. However, contemporaneously to this work, Gunawan, Farhang-Mehr and Azarm (2003) also recognised the potential for global problem decomposition (in terms of both objective-space and decision-space) and developed an EMO system to exploit this. A type of *farmer-worker* concept was suggested, in which a global MOEA (the farmer) initialises a number of individual MOEAs (the workers) to optimise sub-problems in parallel. The farmer process then performs composition and diversity promotion operations on the results of these algorithms. New worker processes are then instigated, and so forth. This methodology was implemented and compared to a global EMO solution on a benchmark problem. The results for the decomposition-based system were shown to be superior, but a rigorous test framework was unfortunately not used. Also the main limitation of the technique is that it requires a static, *a priori* decomposition of the problem-space. Both these matters are addressed by the method described in this chapter.

6.1.3 Chapter Overview

This chapter demonstrates the benefit of using a divide-and-conquer strategy for ICs in multi-objective optimisation when the correct decompositions are known in advance. It also proposes a general methodology for identifying, and subsequently exploiting, the decomposition during the optimisation process. An empirical framework is described in Section 6.2, which is then used to establish the case for divide-and-conquer in Section 6.3. An on-line

adaptive strategy is proposed in Section 6.4 that exploits the iterative, population-based nature of the evolutionary computing paradigm. Independent collections of objectives are identified using nonparametric statistical methods of independence testing. Sub-populations are assigned to the optimisation of each collection, with migration between these occurring as the decomposition is revised over the course of the optimisation. Proof-of-principle results are presented in Section 6.5, together with a discussion of issues raised by the study.

6.2 Experimental Methodology

6.2.1 Baseline Algorithm

The baseline evolutionary multi-objective optimiser chosen in this work is a variant of the elitist multi-objective genetic algorithm developed in Section 3.6. An overview is shown in Table 6.1. Parameter settings are derived from the literature and tuning has not been attempted.

Table 6.1: Baseline MOEA used in this chapter

EMO component	Strategy
General	Total population = 100ψ. Generations = 250.
Representation	Concatenation of real number decision variables. Accuracy bounded by machine precision.
Selection-for-variation	Binary tournament selection using Pareto-based ranking (Fonseca and Fleming 1993).
Variation	Uniform SBX crossover with $\eta_c = 15$, exchange probability = 0.5, and crossover probability = 1 (Deb and Agrawal 1995). Element-wise polynomial mutation with $\eta_m = 20$ and mutation probability = (chromosome length)$^{-1}$ (Deb and Goyal 1996).
Selection-for-survival	Ceiling of 20%-of-population-size of non-dominated solutions preserved. Reduction using SPEA2 clustering (Zitzler, Laumanns and Thiele 2001).

6.2.2 Test Functions

A simple way to create independent multi-objective test functions is to concatenate existing test problems from the literature, in which dependencies exist between all objectives. In this proof-of-principle study, only the test function ZDT-1 proposed by Zitzler *et al.* (2000) is used. ZDT-1 is reproduced in Equation A.2 in Appendix A. The concatenated extension to this bi-objective task for ψ sub-problems, denoted as *C-ZDT-1(ψ)*, is shown in Equation 6.1.

$\tilde{n}$ is the number of decision variables per sub-problem, such that $n = \psi\tilde{n}$ is the total number of decision variables for the global problem.

$$\left.\begin{array}{lrcl}
\text{min.} & \boldsymbol{z}(\boldsymbol{x}) & = & [z_1(x_1), z_2(x_1, \ldots, x_{\tilde{n}}), \ldots, \\
 & & & \quad z_{2\psi-1}(x_{(\psi-1)\tilde{n}+1}), z_{2\psi}(x_{(\psi-1)\tilde{n}+1}, \ldots, x_{\psi\tilde{n}})]\,, \\
\text{w.r.t.} & \boldsymbol{x} & = & [x_1, \ldots, x_{\psi\tilde{n}}]\,, \\
\text{where} & z_{2q-1}(\boldsymbol{x}) & = & x_{(q-1)\tilde{n}+1}, \\
 & z_{2q}(\boldsymbol{x}) & = & 1 - \sqrt{z_{2q-1} / \left(1 + [9/(\tilde{n}-1)] \sum_{j=(q-1)\tilde{n}+2}^{q\tilde{n}} x_j\right)}, \\
\text{and} & 0 \leq x_i \leq 1 & & \text{for } i = 1, 2, \ldots, \psi\tilde{n}.
\end{array}\right\} \tag{6.1}$$

The global solution to this problem is a set of bi-objective trade-off surfaces (z_1 versus z_2, z_3 versus z_4, and so forth). Each trade-off surface is a convex curve in the region $[0\ 1]^2$ (for which the summation in Equation 6.1 is zero). The *ideal vector* is [0 0]. The *anti-ideal vector* of worst possible performance in each objective is [1 10].

6.2.3 Performance Metrics

Hypervolume

A quantitative measure of the quality of a trade-off surface is made using the hypervolume S unary performance metric (Zitzler 1999). The hypervolume metric measures the amount of objective-space dominated by the obtained non-dominated front, and is one of the best unary measures currently available, although it has limitations (Zitzler *et al.* 2003, Knowles and Corne 2002). The anti-ideal vector is taken as the reference point. The metric is normalised using the hypervolume of the ideal vector, as illustrated in Figure 6.1.

Attainment Surfaces

An introduction to the concept of attainment surfaces is provided in Section 3.2.3. Note that, as was the case in Chapter 3, the technique is again used purely for visualisation purposes.

6.2.4 Analysis Methods

For the type of MOEA described in Table 6.1, the final population represents an appropriate data set upon which to measure performance. 35 runs of each algorithm configuration

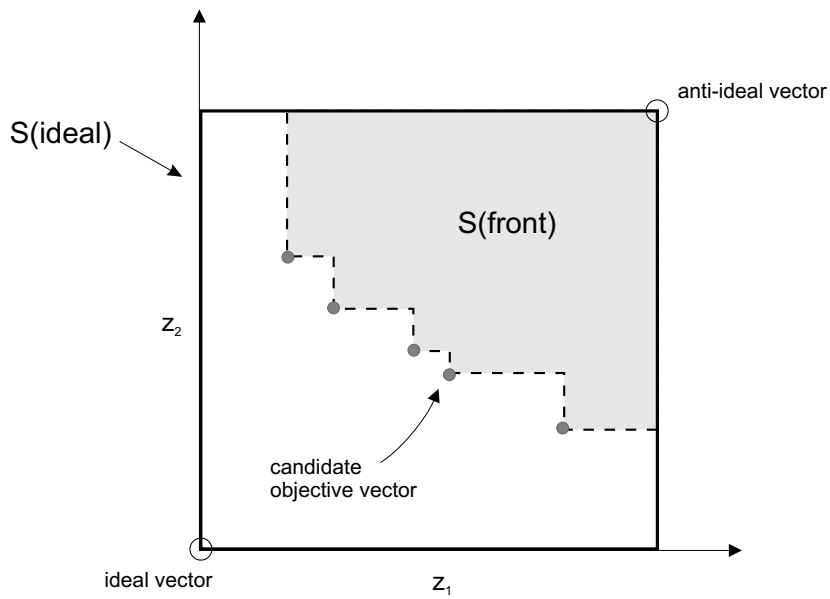


Figure 6.1: Hypervolume metric

have been conducted in order to generate statistically reliable results. Quantitative performance is then expressed in the distribution of obtained hypervolumes. A comparison between configurations is made via the difference between the means of the distributions.

The significance of the observed result is assessed using the simple, yet effective, non-parametric method of randomisation testing (Manly 1991). The methodology was used for comparative performance testing in Chapter 3 and is described in detail in Section 3.2.4. A two-tailed test at the 1%-level is used here as before.

Let $S1$ be the distribution of hypervolume metrics for `algorithm_1`, and let $S2$ be the corresponding distribution for `algorithm_2`. The observed value is then found by subtracting the mean of $S2$ from the mean of $S1$. An overall randomised distribution is then generated for this test statistic.

Since optimal performance is achieved by maximising hypervolume, if the observed value falls to the left of the distribution then there is strong evidence to suggest that `algorithm_2` has outperformed `algorithm_1`. If the observed result falls to the right, then superior performance is indicated for `algorithm_1`.

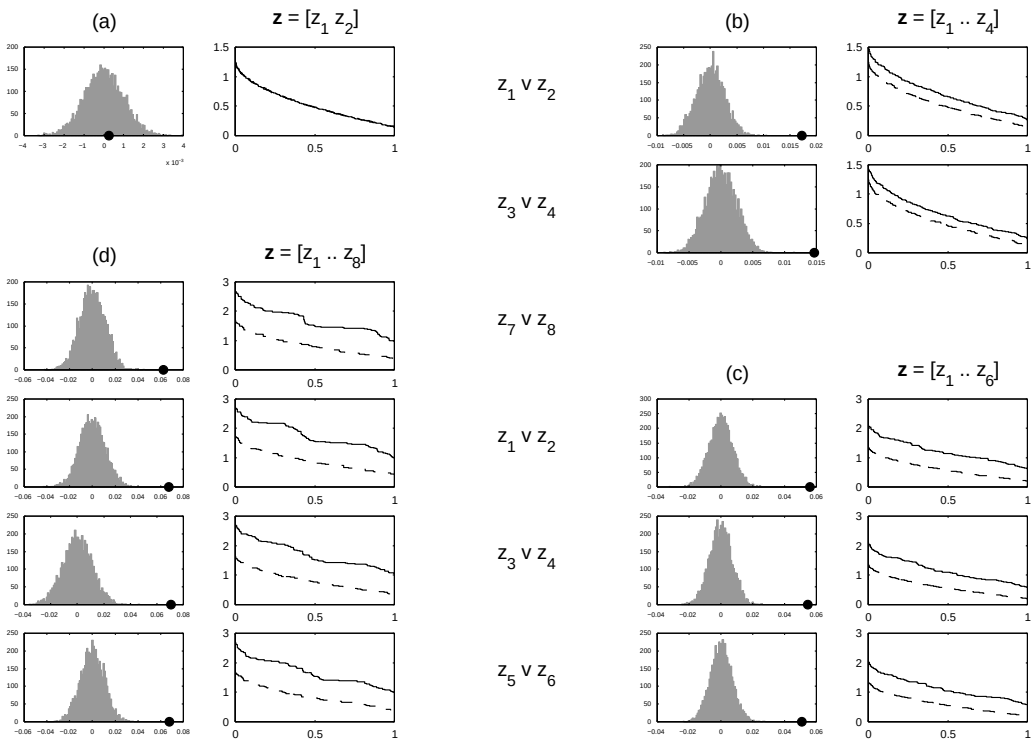


Figure 6.2: Example presentation of results: `algorithm_1` versus `algorithm_2`

6.2.5 Presentation of Results

Comparisons of `algorithm_1` versus `algorithm_2` for $\psi = [1, \ldots, 4]$ are summarised within a single figure such as Figure 6.2. Region (a) shows the validation case of one independent collection, *C-ZDT-1(1)*, whilst regions (b), (c), and (d) show two, three, and four collections respectively. Within each region, each row indicates a bi-objective comparison. The left-hand column shows the results of the randomisation test on hypervolume (if the observed value, indicated by the filled circle, lies to the right of the distribution then this favours `algorithm_1`), whilst the right-hand column shows the median attainment surfaces (the unbroken line is `algorithm_1`).

The two figures in the top-left of Figure 6.2 show the comparison between `algorithm_1` and `algorithm_2` for the bi-objective problem $\boldsymbol{z} = [z_1\ z_2]$. Considering these two figures, the left-hand figure shows the results of the randomisation test. The observed result, indicated by the filled circle, is seen to lie very close to the centre of the distribution (shown as a grey histogram). This suggests that there is no difference in performance between the two

algorithms in terms of hypervolume. This result is confirmed by the attainment surface plots in the right-hand figure. The two median surfaces are so closely located as to be indistinguishable in the figure.

As a further example, consider the bottom-right cluster of pictures in Figure 6.2. These are the results for a six-objective problem that is decomposable into three bi-objective sub-problems. Within this portion of the overall Figure 6.2, results for each sub-problem are shown on each row. The hypervolume results for each sub-problem show the observed result to lie to the right of the randomisation distribution. This indicates that `algorithm_1` exhibits better performance (in terms of the hypervolume indicator) than `algorithm_2`. The attainment surface visualisations in the left column also indicate the superiority of `algorithm_1`: recalling that all objectives are to be minimised, the dashed-line median attainment surfaces for this algorithm clearly dominate the corresponding unbroken-line surfaces for `algorithm_2`.

6.3 The Effect of Independence

The potential of a divide-and-conquer strategy can be examined by comparing a global solution to the concatenated ZDT-1 problem to *a priori* correct decompositions in terms of decision-space, objective-space, or both. Consider an island model scheme in which a sub-population of 100 individuals is evolved in isolation for each independent collection (no migration occurs). Each EA uses only the relevant objectives and decision variables. This is compared to a global approach, with a single population of size 100ψ, using all objectives and decision variables in Figure 6.3. Substantially improved performance is shown for the divide-and-conquer scheme.

To clarify which parts of the decomposition are important, sub-population schemes that decompose decision-space whilst treating objective-space globally and *vice versa* are now considered. In the decision-space scheme, each 100 individual sub-population operates on the correct subset of decision variables. However, s_s and fitness values for s_v are decided globally. During s_s, elite solutions are reinserted into the most appropriate sub-population depending on their ranking on local objective sets. Assignment is random in the case of a tie. Performance is compared to the ideal decomposition in Figure 6.4. It is evident that decision-space decomposition alone is not responsible for the results in Figure 6.3, and that the quality of the trade-off surfaces deteriorates with ψ. The attainment surfaces for cases (c) and (d)

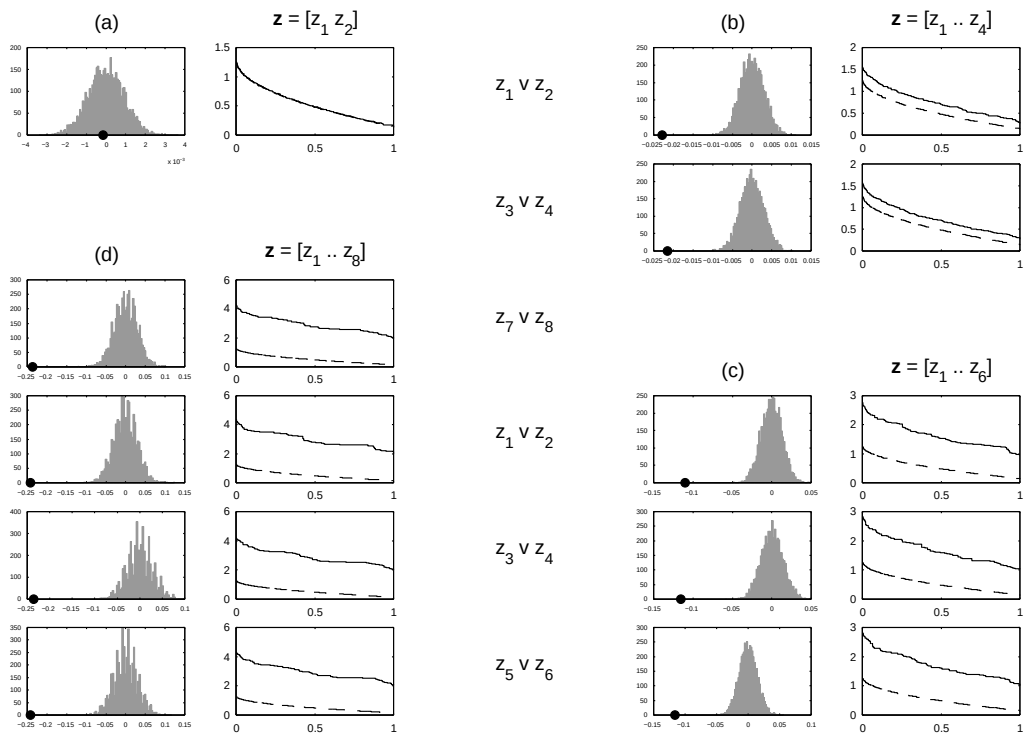


Figure 6.3: Global model versus decomposition of both objective-space and decision-space

suggest that the global treatment of objectives may be affecting the shape of the identified surface.

In the objective-space scheme, each sub-population operates on the correct subset of objectives (the s_s and s_v processes are both completely contained within the sub-population), but the EA operates on the global set of decision variables. A comparison with the decision-space method is shown in Figure 6.5. No statistically significant performance difference is evident in any of the cases. Thus, objective-space decomposition alone is also not responsible for the achievement in Figure 6.3. Note that if single-point rather than uniform crossover had been used then results would have been much worse for the global treatment of decision-space since, for the former operator, the probability of affecting any single element of the chromosome (and thus the relevant section) decreases with chromosome length.

The above results indicate that a sub-population-based decomposition of either objective-space or decision-space can significantly benefit performance. The best results are obtained when both domains are decomposed simultaneously. Given that, in general, the correct decomposition for either domain is not known in advance, the choice of domain will depend

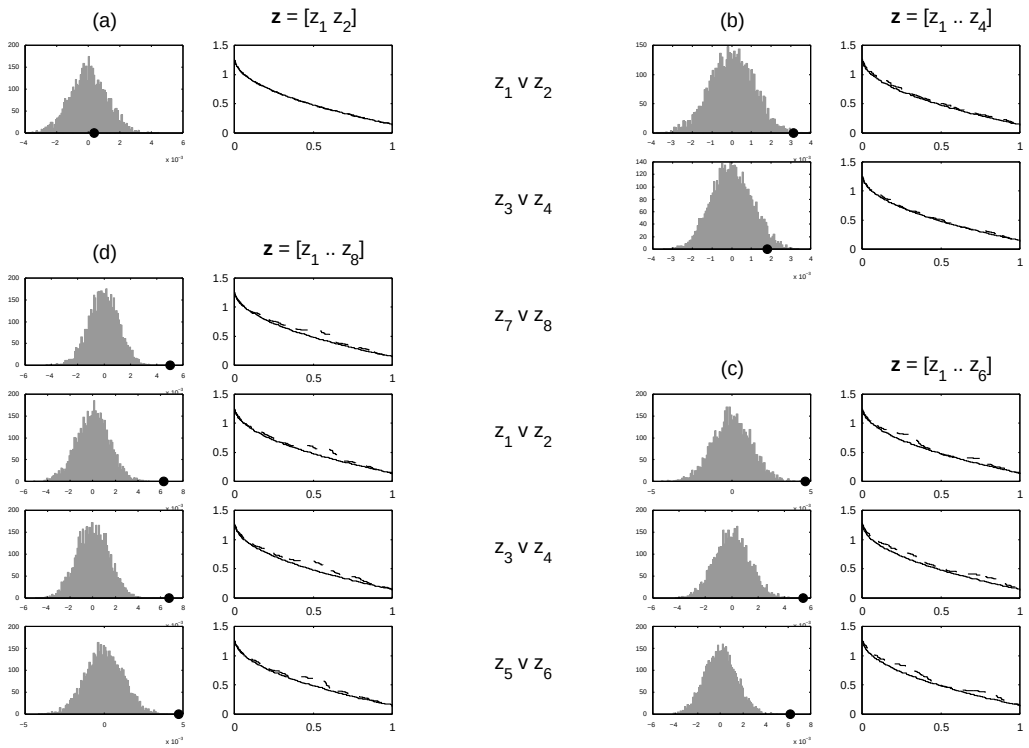


Figure 6.4: Decomposition of both objective-space and decision-space versus decision-space decomposition alone

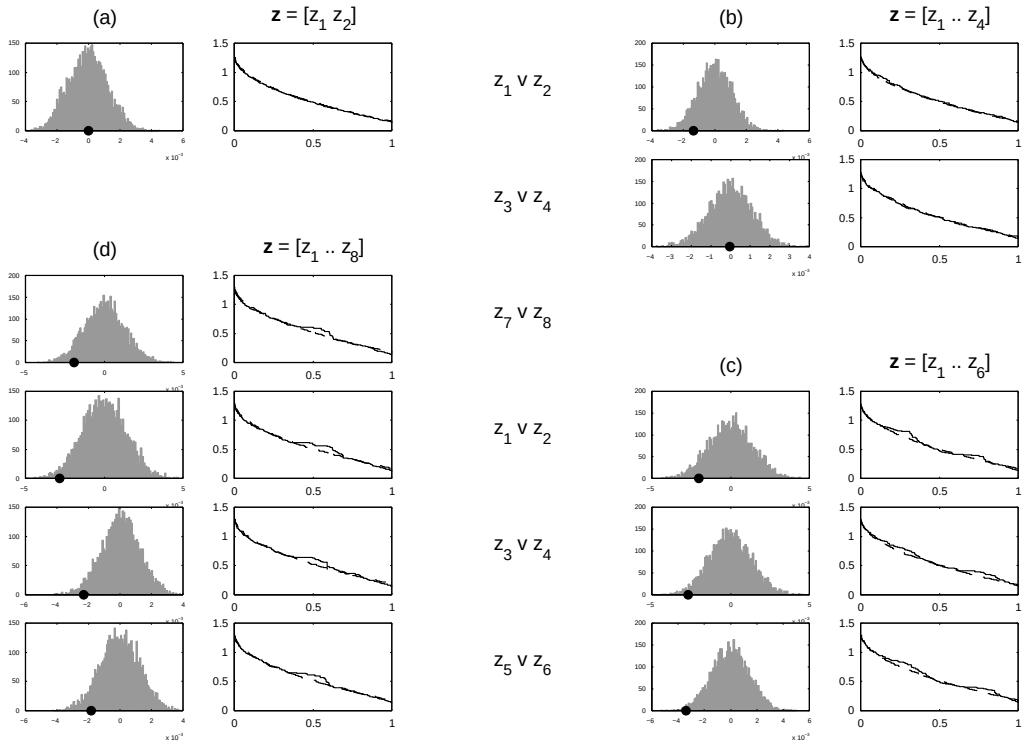


Figure 6.5: Decision-space decomposition versus objective-space decomposition

on which is less demanding to analyse. Note that if objective-space is decomposed then decision-space decomposition is also required at some point in order to synthesise a global solution. However, the converse is not the case. Decomposition may be *a priori*, progressive, or *a posteriori* with respect to the optimisation. A correct early decomposition in both spaces would be ideal but this may not be achievable due to insufficient knowledge of the interactions between objectives at this stage.

6.4 Exploiting Independence via Objective-Space Decomposition

6.4.1 Overview of the Methodology

A progressive decomposition of objectives, together with a retrospective decomposition of decision variables, is proposed in this section. This is appropriate for problem domains where the number of objectives is significantly fewer than the number of decision variables. An island-based sub-population approach is taken, in which the selection probability of an individual in a sub-population is determined on the subset of objectives assigned to that sub-population. The topology of this parallel model is permitted to vary over the course of the optimisation.

An overall schematic of the technique is given in Figure 6.6. The process begins with a global population model. The multi-objective performance of each candidate solution is then obtained. From the perspective of a single objective, the population provides a set of observations for that objective. Pair-wise statistical tests for independence are then performed for all possible pairs of objectives to determine between which objectives dependencies exist. Linkages are created for each dependent relationship. A sub-problem is then identified as a linked collection of objectives. This concept is illustrated for an ideal decomposition of *C-ZDT-1(2)* in Figure 6.7. Of all pair-wise dependency tests, significant dependencies have been identified between z_1 and z_2, and between z_3 and z_4.

The new topology of the population model follows from the decomposition. *Split* and *join* operations are implemented to allow objectives (and associated candidate solutions) to migrate between sub-problems as appropriate.

When each new sub-population has been formed, selection probabilities and the identi-

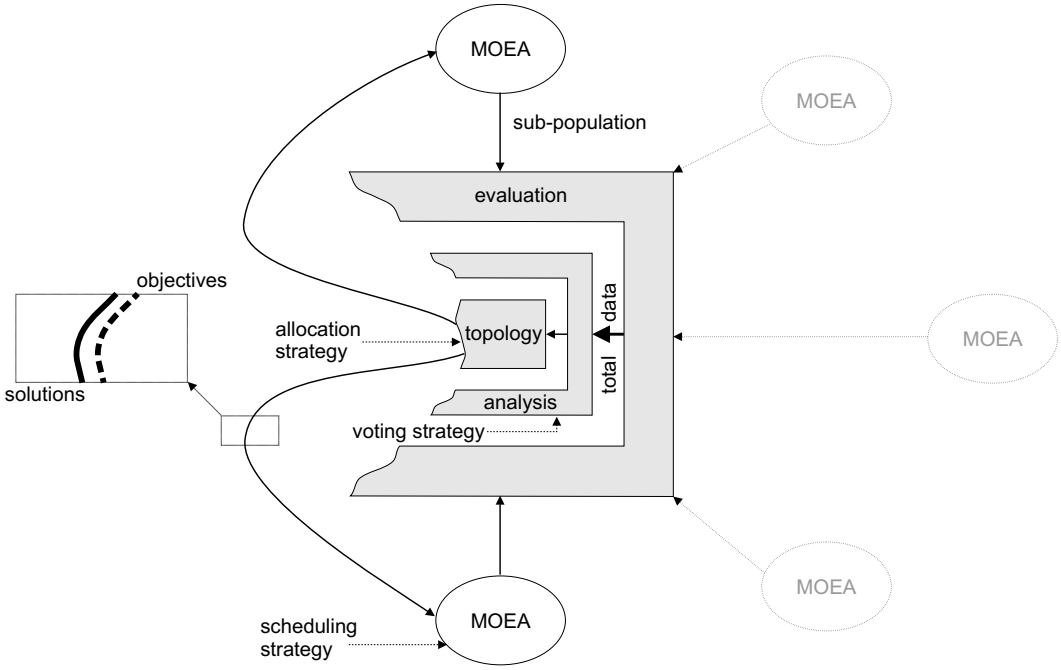


Figure 6.6: Schematic overview of the progressive decomposition methodology

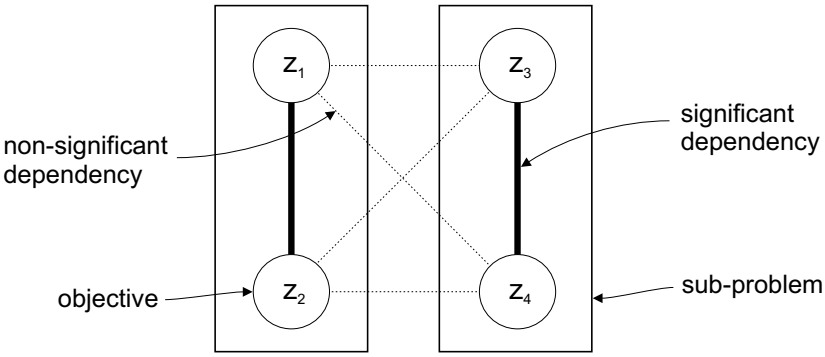


Figure 6.7: Identification of sub-problems via linkage

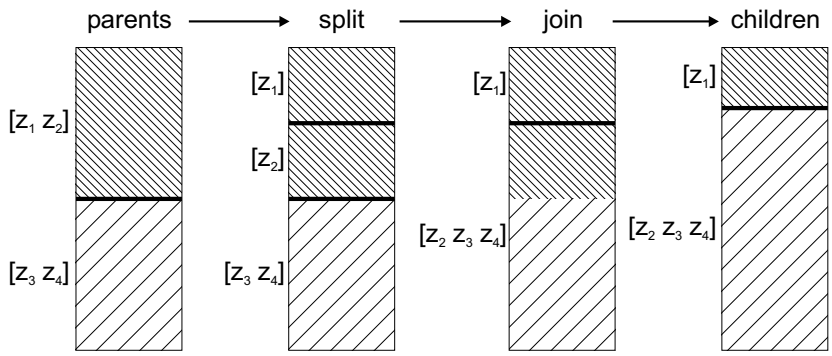


Figure 6.8: Example of the split and join operations

fication (and management) of elites are determined using the current subset of objectives. Performance across all other objectives is ignored. Selection and variation operators are then applied within the boundaries of the sub-population. The size of the resulting new sub-population is pre-determined by the population management process.

All new solutions are evaluated across the complete set of objectives. This new data is then used to determine an updated decomposition. In this study, the update is performed at every generation, although in general it could be performed according to any other schedule. The process then continues in the fashion described above.

6.4.2 Population Management

The population topology is dependent on the identified decomposition. This can change during the course of the optimisation, thus requiring some sub-problem resources to be re-allocated elsewhere. Operations are required to split some portion of the candidate solutions from a sub-population and subsequently join this portion on to another sub-population.

The size of the split is decided using an *allocation strategy*. Since the number of candidate solutions required to represent a trade-off surface grows exponentially with the number of associated conflicting objectives, Q, it would seem a reasonable heuristic to use an exponential allocation strategy such as 2^Q. As an example, consider a four-objective problem with an initial decomposition of $\{[z_1\ z_2], [z_3\ z_4]\}$. The suggested new decomposition is $\{[z_1], [z_2\ z_3\ z_4]\}$. The situation is depicted in Figure 6.8.

Prior to reallocation, each sub-population should have proportion $2^2/(2^2 + 2^2) = 1/2$

of the available resources. z_2 is now to be linked with z_3 and z_4, and must thus be split from its grouping with z_1. Both z_1 and z_2 will receive $2^1/(2^1 + 2^1) = 1/2$ of the resources in this sub-population. The actual candidate solutions to be assigned to each part of the split are determined randomly. The resources allocated to z_2 are then added to the $[z_3\ z_4]$ sub-population. Now z_1 has 1/4 of the resources, whilst $[z_2\ z_3\ z_4]$ has 3/4 of the resources. The selection and variation operations are then used to return to the required proportions of $2^1/(2^1 + 2^3) = 1/5$ for z_1 and $2^3/(2^1 + 2^3) = 4/5$ for $[z_2\ z_3\ z_4]$.

6.4.3 Tests for Independence

A sub-problem group is generated by collecting all objectives that are linked by observed pairwise dependencies. In this sub-section, the tests used to determine if a connection should be made are introduced. Several tests for variable independence based on sample data exist in the statistics literature (Hollander and Wolfe 1999). Two nonparametric procedures, the first based on the *Kendall K* statistic and the second on the *Blum-Kiefer-Rosenblatt D* statistic, are used in this work.

Both methods require special care for the handling of tied data. This is of concern in an evolutionary algorithm implementation since a particular solution may have more than a single copy in the current population. Large-sample approximations to each method have been implemented. This is possible because reasonably large population sizes have been used (100 individuals per independent bi-objective collection). All significance tests are two-tailed at the 1%-level, the null hypothesis being that the objectives are independent.

Kendall *K*

A distribution-free bivariate test for independence can be made using the Kendall sample correlation statistic, K. This statistic measures the level of *concordance* (as one variable increases/decreases, the other increases/decreases) against the level of *discordance* (as one variable increases/decreases, the other decreases/increases). This is somewhat analogous to the concepts of *harmony* and *conflict* in multi-objective optimisation (see Chapter 4). The standardised statistic can then be tested for significance using the normal distribution $N(0, 1)$. Ties are handled using a modified paired sign statistic. A modified null variance is also used in the standardisation procedure. For further details, refer to Hollander and Wolfe (1999).

The main concern with this method is that if $K = 0$ this does not necessarily imply that the two objectives are independent (although the converse is true). This restricts the applicability of the method beyond bi-objective dependencies, where relationships may not be monotonic.

Blum-Kiefer-Rosenblatt D

As an alternative to the above test based on the sample correlation coefficient, Blum, Kiefer, and Rosenblatt's large-sample approximation to the *Hoeffding D* statistic has also been considered in this study. This test is able to detect a much broader class of alternatives to independence. For full details, refer to Hollander and Wolfe (1999).

6.4.4 Decision-Space Decomposition: An Aside

Discussion

In the above methodology, and throughout the forthcoming empirical analysis of this method in Section 6.5, different sub-populations evolve solutions to different collections of objectives. Decision-space decomposition is not attempted. Thus, at the end of the optimisation process, complete candidate solutions exist for each objective collection. It is now unclear which decision variables relate to which objective collection. In order to finalise the global solution, solutions from each trade-off surface must be synthesised via partitioning of the decision variables.

An *a posteriori* decomposition, as described below, is simple to implement but has two clear disadvantages: (i) some reduction in EA efficiency will be incurred because the operators search over inactive areas of the chromosome (operators that are independent of chromosome length should be used), and (ii) further analysis is required to obtain the global solution.

Method

A candidate solution should be selected at random from the overall final population. Each variable is then perturbed in turn and the effect on the objectives should be observed. The variable should be associated with whichever objectives are affected. Then, when the decision-maker selects a solution from the trade-off featuring a particular group of objectives, the

corresponding decision variables are selected from the sub-population corresponding to this grouping.

This method requires as many extra candidate solution evaluations as there are decision variables in the problem. For the 4-collection concatenated ZDT-1 test function, this is 120 evaluations or 30% of a single generation of the baseline algorithm.

Two special cases must be addressed:

- If the perturbation of a decision variable affects objectives in more than one objective subset, this indicates an invalid decomposition of objective-space. Information of this kind could be used progressively to increase the robustness of the decomposition.
- It is possible that no disturbance of objectives is seen when the decision variable is perturbed. Here, the alternatives are to consider another candidate solution or to consider more complicated variable interactions. This may also be an indication that the variable is globally redundant.

6.5 Preliminary Results

Proof-of-principle results for the adaptive divide-and-conquer strategy devised in Section 6.4 are presented herein for the concatenated ZDT-1 test function (Definition 6.1) with $\psi = [1, \ldots, 4]$. A summary of the chosen strategy is given in Table 6.2. Both the Kendall K method and the Blum-Kiefer-Rosenblatt D method have been considered.

Table 6.2: Divide-and-Conquer settings

EMO component	Strategy	
Independence test	(either)	Blum-Kiefer-Rosenblatt D
	(or)	Kendall K
Resource allocation	2^Q	
Schedule	Revise the decomposition every generation	

6.5.1 Blum-Kiefer-Rosenblatt D Results

The performance of this strategy when compared to the baseline case of no decomposition is shown in Figure 6.9. Both the hypervolume metric results and the attainment surfaces indicate that the divide-and-conquer strategy produces trade-off surfaces of higher quality

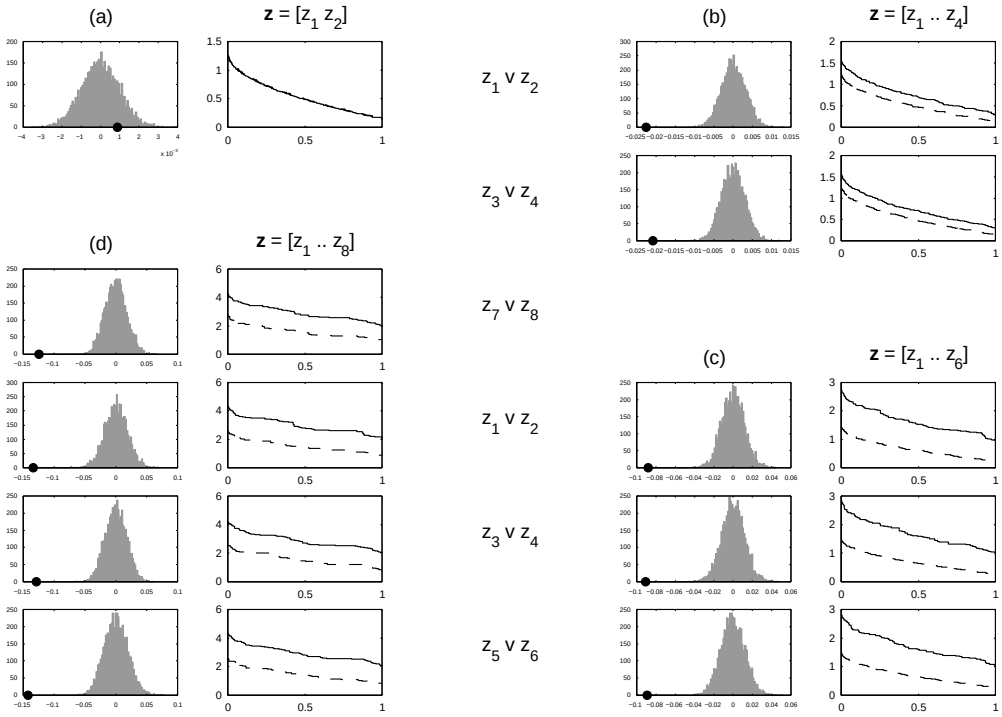


Figure 6.9: No decomposition versus Blum-Kiefer-Rosenblatt D divide-and-conquer

in cases where independence exists. However, the attainment surfaces also show that the absolute performance of the method degrades as more independent collections are included.

The degradation can be partially explained by considering the percentage of correct decompositions made by the algorithm at each generation (measured over the 35 runs) shown in Figure 6.10. As the number of independent collections increases, the proportion of correctly identified decompositions decreases rapidly. Note that this does not necessarily mean that the algorithm is making invalid decompositions or no decomposition: other valid decompositions exist, for example $\{[z_1\ z_2], [z_3\ z_4\ z_5\ z_6]\}$ for $\psi = 3$, but these are not globally optimal (also the number of possible decompositions increases exponentially with ψ.). Indeed, on no occasion did the test produce an invalid decomposition (identified independence when dependency exists). This is evident from plots of the decomposition history over the course of the optimisation. A typical history is depicted in Figure 6.11. Each objective is labelled on the vertical axis, whilst the horizontal axis depicts the current generation of the evolution. At a particular generation, objectives that have been identified as an independent collection are

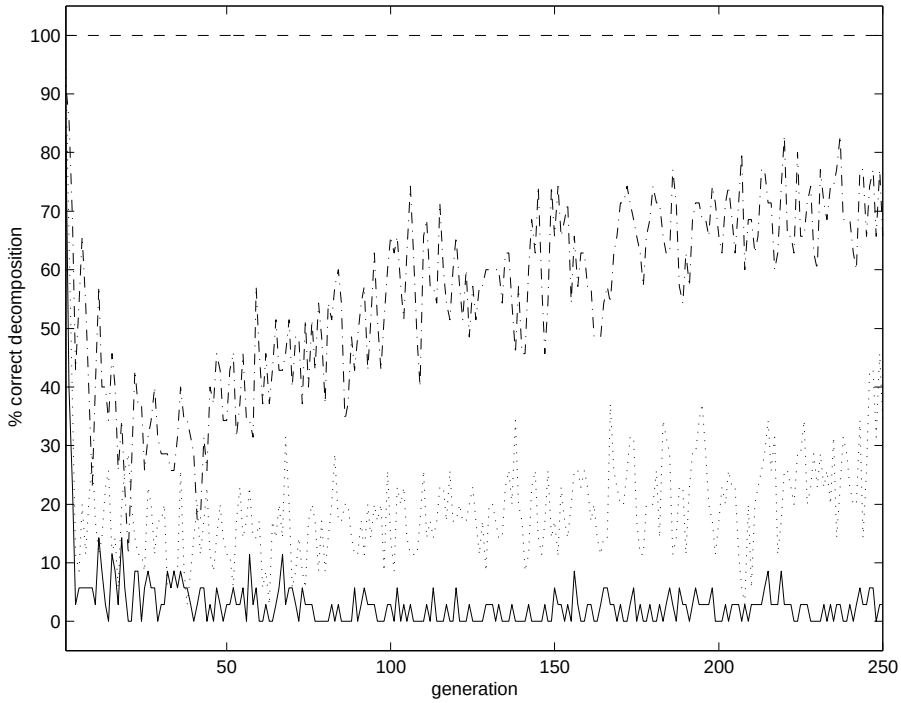


Figure 6.10: Blum-Kiefer-Rosenblatt D: Correct decompositions as a percentage of total runs over the course of the optimisation.
$- - - -$ $[z_1\ z_2]$, $- \cdot - \cdot$ $[z_1, \ldots, z_4]$, $\cdots\cdot$ $[z_1, \ldots, z_6]$, —— $[z_1, \ldots, z_8]$

associated with a unique colour. Thus, as shown in Figure 6.11, at the initial generation z_1 and z_2 have been identified as a cluster (white), as have $[z_3\ z_4]$ (black), $[z_5\ z_6]$ (light grey), and $[z_7\ z_8]$ (dark grey). At generation 200 all the objectives have been grouped together, as indicated by the complete whiteness at this point in the graph. Note that there is no association between the colours across the generations.

6.5.2 Kendall K Results

The performance of the divide-and-conquer algorithm with the Kendall K test for independence is compared to Blum-Kiefer-Rosenblatt D in Figure 6.12. The former test appears to offer superior performance as the number of independent collections increases. No significant performance difference can be found for $\psi = 2$, but such a difference can be seen for two of the surfaces for $\psi = 3$, and every surface for $\psi = 4$.

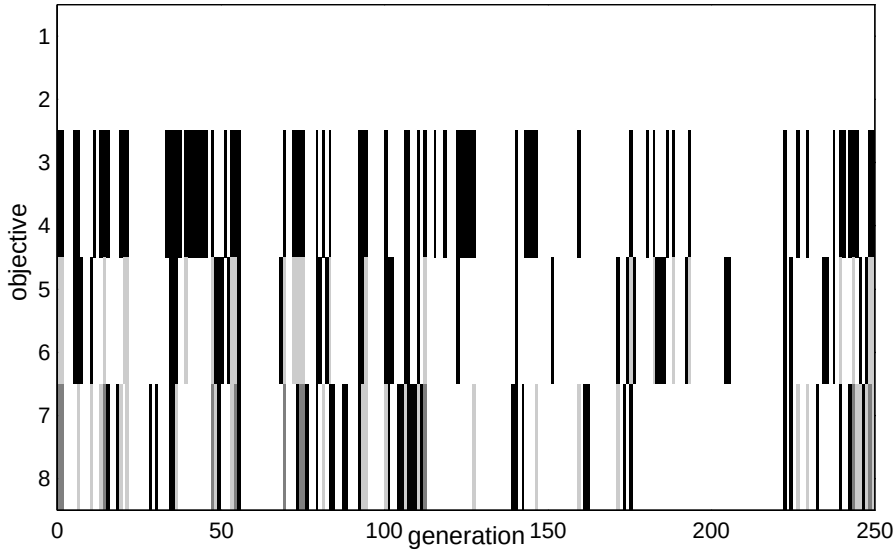


Figure 6.11: Blum-Kiefer-Rosenblatt D: Typical decomposition history for a single replication. Each identified objective cluster is represented by a colour (white, light grey, dark grey, black)

This difference in performance can be explained by the plot of correct decompositions shown in Figure 6.13. Whilst the proportion of correct decompositions degrades as ψ increases, this degradation is not as severe as for Blum-Kiefer-Rosenblatt D (previously shown in Figure 6.10). Also, from the typical decomposition history depicted in Figure 6.14, the valid decompositions tend to be of higher resolution than those developed by the alternative method (shown in Figure 6.11).

6.5.3 Discussion

The obtained results show that the adaptive divide-and-conquer strategy offers substantially better performance than the global approach in terms of the quality of trade-off surfaces generated.

Of the two independence tests considered, Kendall K would appear more capable of finding good decompositions on the benchmark problem considered, especially as the number of independent collections increases. However, Kendall K may experience difficulties when the dimension of the trade-off surface increases, since it may incorrectly identify independence due to the variation in the nature of bi-objective relationships over the surface (the relationship

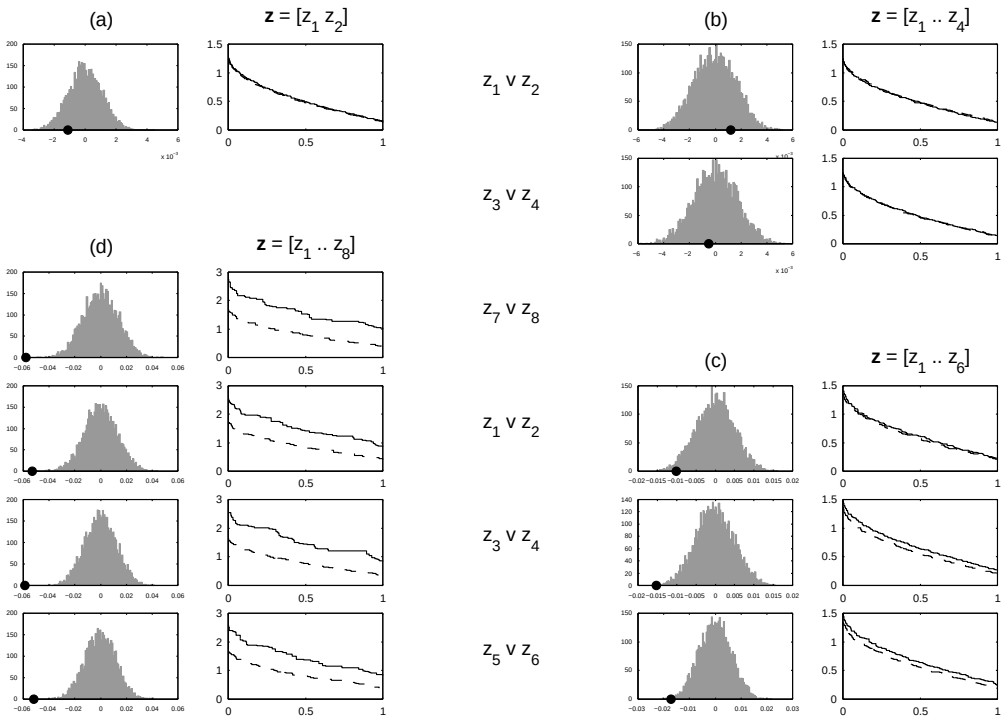


Figure 6.12: Blum-Kiefer-Rosenblatt D divide-and-conquer versus Kendall K divide-and-conquer

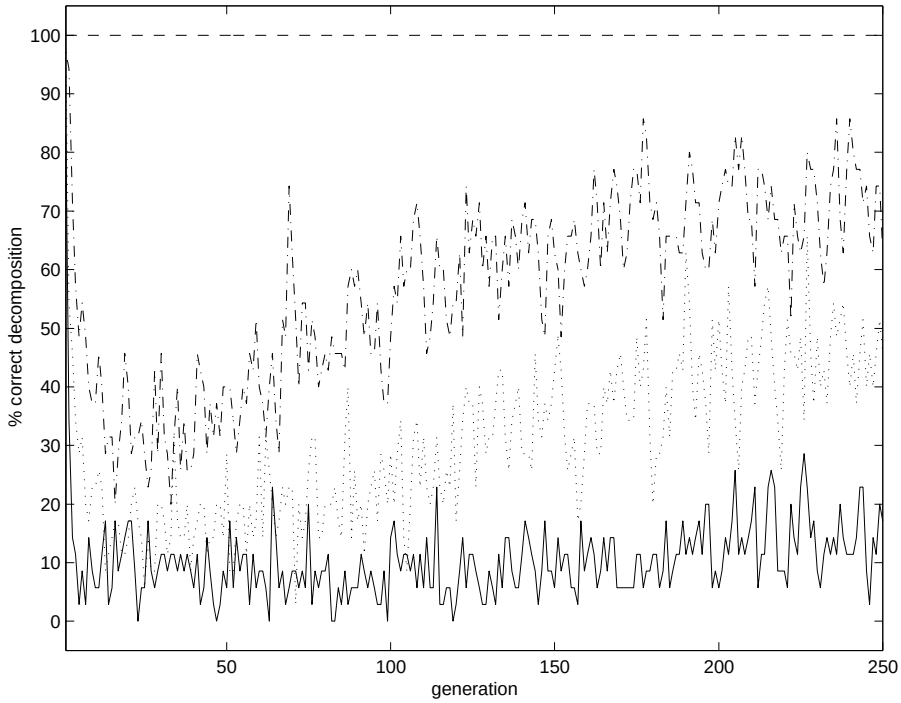


Figure 6.13: Kendall K: Correct decompositions as a percentage of total runs over the course of the optimisation.

$- - - -$ $[z_1\ z_2]$, $- \cdot - \cdot$ $[z_1, \ldots, z_4]$, $\cdots\cdot$ $[z_1, \ldots, z_6]$, —— $[z_1, \ldots, z_8]$

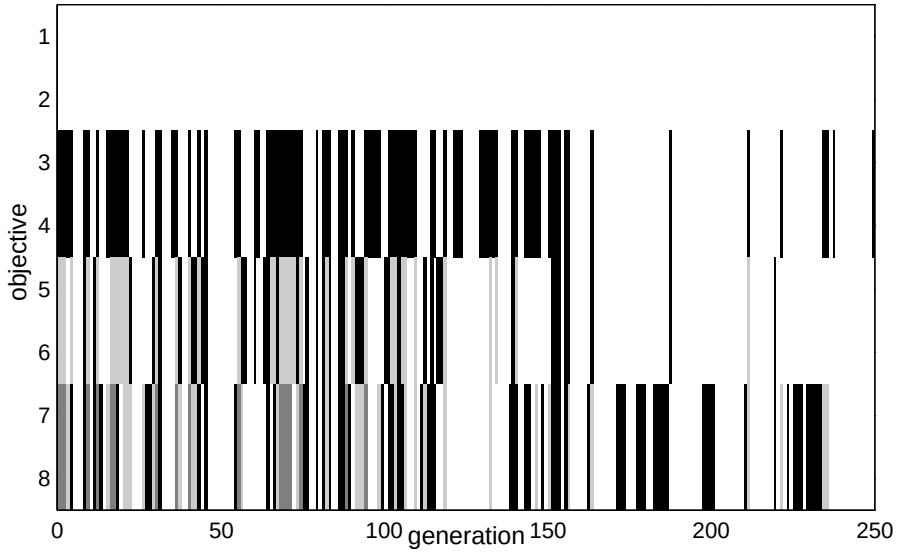


Figure 6.14: Kendall K: Typical decomposition history for a single replication. Each identified objective cluster is represented by a colour (white, light grey, dark grey, black)

is not always conflicting, as it is for a bi-objective problem). By contrast, Blum-Kiefer-Rosenblatt D offers a more robust search in these conditions, but is more conservative.

There is a clear need for the procedure to be robust (invalid decompositions should be avoided, although the progressive nature of the process may somewhat mitigate the damage from these), but conservatism should be minimised in order to increase the effectiveness of the methodology. Under these circumstances, it may be prudent to adopt a voting strategy, in which a decision is made based on the results from several tests for independence.

The adaptive divide-and-conquer strategy carries some overhead in terms of the test for independence and the sub-population management activity, which may be controlled using a scheduling strategy. This must be balanced against the improvements in the quality of the trade-off surfaces identified and the reduction in the complexity of the MOEA ranking and density estimation procedures.

6.6 Summary

This study has shown that, if feasible, a divide-and-conquer strategy can substantially improve MOEA performance. The decomposition may be made in either objective-space or

decision-space, with a joint decomposition proving the most effective. Objective-space decomposition is particularly appealing because it reduces the complexity of the trade-off surfaces to be presented to the decision-maker. Furthermore, no loss of trade-off surface shape was observed for the ideal objective-space decomposition as it was for the sole decomposition of decision variables.

An adaptive objective-space decomposition methodology has been presented and proof-of-principle results on the concatenated ZDT-1 problem have been shown to be very encouraging. It should be noted that the approach is not confined to objective-space: independence tests and identified linkages could equally well have been applied to decision variable data. In this case, the sub-populations would evolve different decision variables, whilst the evaluation would be global. In such a decision-space decomposition, careful consideration would be required of the methodology by which a global evaluation could be based on a subset of decision variables. The simultaneous decomposition of both objective-space and decision-space could also be attempted using the techniques in this chapter.

The main limitation of the methodology is the number of pair-wise comparisons that have to be conducted for high-dimensional spaces. In the concatenated ZDT-1 problem, analysis of the decision variables would be very compute-intensive. Thus, further techniques for the progressive decomposition of high-dimensional spaces are required to enhance the framework.

Chapter 7

Conclusions

7.1 Evolutionary Multi-Objective Optimisation

In the general case, a candidate solution to a problem will be assessed against multiple performance criteria. If conflict arises between these different objectives then no single solution can be considered optimal in the classical sense of exhibiting best performance across all the objectives. Instead, optimality encompasses a family of alternative solutions whose corresponding objective vectors define the trade-off surface for the problem. Thus a multi-objective optimiser, in the absence of decision-maker preferences between the objectives, is required to find a good representation of the trade-off surface to present to the DM.

The appeal of evolutionary algorithms as multi-objective optimisers rests primarily on their population-based nature. The required family of alternative solutions can be evolved in parallel, with exchange of information between the multiple search points providing the potential for superior efficiency over a single-solution strategy requiring multiple restarts. The use of a population also permits solutions to be evaluated via peer group comparisons based on Pareto dominance. Thus, the forced cohesion of non-commensurable objectives can be avoided. Evolutionary computing also has general benefits as an optimisation methodology. Primarily, it permits considerable flexibility with regard to solution representation and performance assessment. The multi-point and stochastic properties of the EA also provide an element of robustness in difficult problem environments (with features such as local minima and noise).

Since the emergence of EMO in the mid-1980s, many different EAs have been proposed

for multi-objective optimisation. Of the mechanisms within such algorithms, the selection processes are those most commonly found to relate directly to the multi-objective nature of the search. Much EMO research effort is devoted to the generation of new algorithms and subsequent comparisons with existing algorithms across a variety of benchmark problems. The results frequently show that the new algorithm has outperformed the existing algorithms (under the restricted conditions imposed) but only very infrequently provide information on the underlying causes of this outcome. The identification of such causes is interesting to the EMO theoretician and also very important to the analyst solving a real-world problem.

In order to exploit the flexibility of the EA methodology as a problem-solving tool, an EA should be tailored to each individual problem based on both innovation and experience gained from other problems (Michalewicz and Fogel 2000). If this argument is accepted, then the resulting movement away from the technique of so-called *black-box* problem solving requires that algorithms be considered on a component basis as well as on a composite scale. In order to choose suitable components and parameter settings to form an optimiser that will provide good solutions to the problem — such that the algorithm is *competent* according to the terminology of Goldberg (2002) — knowledge is required of how these components work in various contexts. This knowledge is not currently available because the type of research framework required to provide it has yet to be adopted by the EMO community on a sufficient scale.

7.2 Performance Assessment

This thesis has sought to develop an experimental performance assessment methodology that will provide useful information to an analyst developing an EMO algorithm at a component level. MOEA performance assessment is a challenging task. In addition to the standard EA difficulty that many design factors must be considered simultaneously (such as choice of selection method and associated selective pressure, variation operators and associated parameters, and so forth), the result for a single run of an MOEA produces a set of vectors rather than the single scalar value that would result from a single-objective EA. Comparisons between the results of MOEAs are thus rather more involved than for a standard EA. Also, visualisation of results becomes problematic for tasks with more than two objectives.

The approach developed in this thesis has stressed the need for baselining. A fundamental

engineering heuristic is to develop a solution that is as simple as possible (whilst meeting the required specifications). Thus, the baseline MOEA considered used Pareto-based selection-for-variation in isolation. No additional selective biases were applied. The effects of further selective mechanisms were then considered with respect to this baseline. Thus, it is possible to identify at a fundamental level whether the extra complexity in the optimiser produces benefits in terms of search performance. Through this process of isolation of components and higher-order component structures it is then possible to begin to identify which mechanisms are crucial for good performance and in what contexts.

In addition to baselining, the methodology emphasises three other important factors: the use of appropriate test problems, the choice of appropriate indicators to evaluate algorithm performance, and a rigorous statistical analysis of the results.

- **Test problems.** MOEA performance is evaluated using benchmark problems. These may be functions developed specifically for testing purposes or may be real-world problems. Any performance conclusions inferred from the results for a particular algorithm can only be directly applied to the local class of problems to which the exemplar problem belongs. Wolpert and Macready's (1997) *no free lunch* (NFL) theorem requires that care is taken when attempting to extend or generalise the performance conclusions since, on average across the set of all problems, all optimisation algorithms exhibit equal performance. However, the insights gained into the *reasons* for the observed performance behaviour can, and should, receive consideration when tailoring an algorithm to a problem from any given class.

- **Performance indicators.** The thesis has generally advocated the use of unary indicators to measure various properties of an obtained approximation set. Thus, in each study, a specific indicator has been chosen to measure proximity and another indicator has been used to assess the quality of the distribution. In this approach, as proved by Zitzler *et al.* (2003), it is not possible to conclude that one algorithm is actually superior to another from the fundamental perspective of Pareto dominance using a finite combination of such indicators. Thus, care is taken in the thesis to focus on individual aspects of performance. Care must also be taken to ensure that the unary indicators correctly measure the desired approximation set attributes. In particular, confusion over the definition of a good distribution may lead to incorrect inferences over results.

Two possible areas of concern in Chapter 3 are that the Δ metric attempts to combine both extent and uniformity into a single value and that the globally optimal distributions for the trade-off surfaces of ZDT-3 and ZDT-5 are not uniform in a Euclidean sense (as implicitly assumed by the metric). In Chapter 5, these concerns were eliminated by concentrating on a specific aspect of distribution and using a problem with a continuous trade-off surface.

- **Statistical analysis.** EMO algorithms are stochastic systems. Thus, multiple replications and statistical confidence tests are required in order to determine the significance of any observed performance differences between algorithms. The use of median indicator values and randomisation testing advocated in the thesis provides strong grounds for inference of algorithm performance (providing that the indicators are chosen appropriately as mentioned above). Visualisation is used in the thesis to reinforce the statistical results and to assist understanding. It is at no point used to make firm judgements on performance, but *is* used in an exploratory capacity in Chapter 5.

The above framework was used in the thesis to identify that (i) diversity promotion schemes can indeed improve the distribution of the obtained approximation set, but that (ii) the genetic drift experienced for an algorithm that did not include such a scheme was not too severe in an absolute sense, and that (iii) an elitist selection-for-survival method based on SPEA2 could offer significant benefits in terms of both proximity and diversity. These conclusions are for the most part limited to the class of real-parameter function optimisation problems, but do cover various landscape characteristics such as convexity, non-convexity, discontinuity, multimodality, and non-uniformity.

7.3 Evolutionary Many-Objective Optimisation

A mismatch has been identified between the number of objectives considered in theoretical EMO research and the number of objectives engaged by real-world applications. In general, the theoretical community restricts itself to bi-objective problems whilst it is not unusual for application developers to attempt to simultaneously optimise as many as 20 objectives. Further understanding of MOEAs in order to support the work of such developers has been identified as a key issue in the thesis. The terminology *many-objective* ($\mathbb{M}$) has been adopted

to describe problems in which there are more than two objectives to be optimised, after Farina and Amato (2002).

A platform for research into many-objective optimisation has been established through consideration of the relationships between objectives. Two objectives may exhibit a dependency relationship, and this may be either conflicting or harmonious. In the former case, as performance is one objective is improved, performance in the other deteriorates. In the latter case, improvement in one objective is rewarded with simultaneous improvement in the other. Alternatively, performance in the two objectives may be entirely independent. In this case, a modification that is responsible for change in one objective produces no change in the other.

Relative to the general high level of research activity in the EMO field, the amount of energy directed toward problems with more than two or three objectives is very small. However, some related work has been identified in the thesis, in both the EMO and MCDM communities. When attempting to optimise many conflicting objectives, EMO researchers have noted that the proportion of non-dominated solutions in a population can become very high (even when large population sizes are used). This can cause problems with selection methods that discriminate on the basis of Pareto dominance and requires the DM to consider a very large number of alternative solution options. Thus, researchers have resorted to the use of further preference data to discriminate between otherwise equivalent solutions. Dimensionality reduction techniques have also been used, especially to provide enhanced visualisation of the relationships between objectives. Harmonious objectives have received some consideration in the MCDM community, wherein much of the discussion has concerned whether or not such objectives should be removed prior to optimisation. This decision is probably best left to the analyst and decision-maker for the particular problem. Prior to the work documented in Chapter 6 of this thesis, no research had been published concerning independence in the context of EMO.

Two studies of many-objective optimisation have been documented in the thesis. In Chapter 5, an exploratory analysis was conducted of the effect of many conflicting objectives on the behaviour of a popular class of MOEAs. In Chapter 6, the benefits of identifying and exploiting independence between objectives were considered. The main conclusions that were drawn from these studies are discussed below.

- **Conflict.** This inquiry considered the additional effect of diversity enhancement mech-

anisms over pure Pareto-based discrimination in selection processes. An exploratory framework was deployed and the number of conflicting objectives was varied between 3 and 12. The study indicated that the conclusions that can be drawn from simulations on a small number of objectives cannot be generalised to problems with a higher number of objectives. In particular, active diversity promotion was found to help prevent genetic drift but, if a high level of dominance resistance is encountered by the variation operators, then the final approximation set can prove to be of very low quality. The performance of mutation-based schemes was found to be less variable than the recombination-based equivalents as the number of objectives changed. Even for high numbers of objectives, some algorithm configurations could still provide good quality results. However, under a fixed population size, the resolution of the approximation set becomes increasingly low (this aspect of quality is DM-dependent). Therefore, a fixed approximation set size may not be appropriate: rather, the DM may wish to specify a desired resolution for each objective. Alternatively, the resolution can be improved by restricting the extent of the trade-off surface that must be represented via the use of DM goals and priorities.

- **Independence.** This inquiry considered an example benchmark problem that could be decomposed into sub-problems (each comprising two conflicting objectives), for which there were no dependencies between different sub-problems. The study found that exploitation of this separability led to improved results (measured using the hypervolume indicator). *A priori* divide-and-conquer approaches of either objective-space or decision-space enhanced the quality of the approximation sets generated. The best results were achieved by enforcing a simultaneous decomposition of spaces. A scheme for dynamically identifying, and subsequently exploiting, a suitable decomposition during the optimisation process was proposed and was shown to work well for the concatenated bi-objective problem considered. The scheme, in its current basic implementation, carries significant overhead in terms of the required tests for independence. This overhead must be balanced against the reduced complexity obtained through the partitioning of objective-space.

7.4 Future Perspectives

7.4.1 Many-Objective Optimisation

The exploratory analysis of conflict in many-objective optimisation produced some noteworthy findings of potential interest to the EMO research community. However, only a single real-parameter function optimisation problem was used during simulations. Thus, despite the realisation that the factors that are believed to underpin the observed behaviour could occur outside of this problem, the results cannot be generalised further at this stage. Similar studies are therefore required for other classes of many-objective problem, such as the *multi-objective quadratic assignment problem* class developed by Knowles and Corne (2003a) which encompasses both the *travelling salesman problem* and the *graph partitioning problem*. Analysis is also required of conflict scalability in real-world scenarios. Other classes of algorithm also require a detailed many-objective analysis. For example, the ϵ-dominance selection-for-survival mechanism developed by Laumanns, Thiele, Deb and Zitzler (2002) is predicted to become an increasingly popular component of EMO algorithms and is thus a good candidate process for a study of this kind. The effect of alternative diversity promotion mechanisms in the context of fixed population sizes should also be considered.

The problem of dominance resistance has been identified as a key concern when optimising many conflicting objectives. Parallels exist here with the problem of *lethals* identified in single-objective multimodal function optimisation by Deb and Goldberg (1989). These researchers discovered that the recombination of spatially dissimilar solutions tended to produce children that performed relatively badly in the problem domain. The authors were able to improve the efficiency of the search by only allowing recombination to occur between parents located within the same local neighbourhood. The neighbourhood size was calculated using the same methods as those used for fitness sharing. In many-objective optimisation, lethals can perhaps be regarded as locally non-dominated remote solutions with a highly substandard component in one or more objectives. Thus, the incorporation of some form of *mating restriction* may prove fruitful in a many-objective context. Indeed, Fonseca and Fleming (1993) suggested such a methodology in the original MOGA.

The study of independence in many-objective optimisation only considered dependency relationships between two objectives in each independent set. The complexity of the depen-

dency relationships could be extended further by, for example, using concatenated problems from the scalable DTLZ test suite. The decomposition of real-world problems should also be considered. Furthermore, the requirement for complete separability could be relaxed to permit weak dependencies between objectives. This may require revision of the adaptive divide-and-conquer strategy employed, possibly importing concepts from coevolution and other EA-based decomposition methodologies. Even if these advancements were not made, the computational complexity of the current method may need to be reconsidered in order to manage any high dimensionality in the objective- and decision-space under consideration.

7.4.2 General EMO Issues

Many avenues of EMO research remain under-explored. Salient issues in five key research areas — theoretical convergence, hypervolume applications, diversity promotion, incorporation of DM preferences, and real-world considerations — are discussed below.

Convergence Analysis

The convergence properties of MOEAs have only recently received consideration in the EMO community. These properties are important because they provide the analyst and decision-maker with improved confidence in the ability of an MOEA to solve the task at hand.

The limit behaviour of MOEAs has been the most heavily studied form of convergence. In this case, an MOEA should provably converge to the 'optimal solution' given an unlimited number of candidate solution evaluations. If the optimal solution in EMO is defined as the set of all Pareto optimal solutions then this property can be achieved by (i) preserving all locally non-dominated solutions during selection-for-survival and (ii) including a variation operator that can generate any solution in the search space with non-zero probability regardless of the input genetic material. This latter requirement can be fulfilled through, for example, the use of a standard mutation operator or the injection of new random solutions at each iteration of the optimiser.

If a finite upper bound to the population size is specified, then it is not in general possible to represent every Pareto optimal solution within the population. The upper bound may be an arbitrary, directly defined, specific integer limit that exists independently from the selection-for-survival processes. Alternatively, it may be an indirectly defined value that is

intrinsic to the selection-for-survival mechanism (rather than a quantity that must simply be respected by the process, as in the former approach). In either case, the notion of optimality must be revised to account for an optimal distribution of solutions. Note that an acceptable definition of such optimality has yet to be fully agreed by the EMO community.

Rudolph and Agapie (2000) developed a scheme in which new locally non-dominated solutions could not be retained during selection-for-survival if this would cause the specific integer limit on population size to be breached. This method ensures convergence in the limit to Pareto optimal solutions but does not preserve a good distribution. Of all the schemes devised for an arbitrary finite limit that *do* actively attempt to maintain a good distribution, none guarantee convergence in the limit (except under special conditions) despite much research effort in this area (Knowles and Corne 2003b).

Unfortunately, this shortcoming is more than a mere theoretical irritation. A significant body of empirical evidence has been accumulated to show that MOEAs can suffer the problem of partial deterioration (where current locally non-dominated solutions in the population are dominated by solutions that have previously been discarded) as a specific result of the lack of a convergence property (Everson *et al.* 2002, Laumanns, Thiele, Deb and Zitzler 2002).

However, if the upper bound on population size is specified implicitly (as a fundamental component of selection-for-survival interacting with objective-space) then convergence in the limit can be achieved for both proximity and (one particular definition of) distribution. This extremely important result was achieved by Laumanns, Thiele, Deb and Zitzler (2002) using a hyperbox selection-for-survival method based on ϵ-dominance.

The upper bound on the number of solution evaluations required for convergence is an important property of optimisation algorithms. For stochastic methods, such as EAs, knowledge of the expected order of magnitude of the evaluations required is desirable. Very little information of this nature is available for MOEAs and this is one of the main sources of criticism of the methods. Whilst noting that this criticism may be somewhat unfair since — unlike with more classical approaches — MOEAs are generally used in problem environments that are difficult to analyse mathematically, information in this area is still valuable. The first known analysis of expected time to convergence has recently been carried out by Laumanns, Thiele, Zitzler, Welzl and Deb (2002) for a simple class of bi-objective problems. Much more research is required in this field, although it should be noted that MOEAs are

generally regarded as approximation methods: thus, they are usually expected to produce *acceptable* solutions rather than globally optimal solutions (to a cost function that is itself generally an approximation to reality).

Potential of the Hypervolume Measure

The concept of hypervolume was introduced by Zitzler (1999) in the context of performance assessment. The hypervolume, or *Lebesgue integral*, of an approximation set is the volume of objective-space that is dominated by the set. Its main application is as a unary performance indicator and it was used in such a form in Chapter 6.

Recent work by Zitzler *et al.* (2003) has shown that the hypervolume metric is the most informative (from a comparative Pareto dominance perspective) of the unary operators currently available. The minor limitations of the metric are that it is sensitive to the choice of a reference parameter and requires objectives to be multiplied together (perhaps requiring transformations of some types of objectives) (Knowles and Corne 2002).

Until recently, the major drawback of using the hypervolume metric was that the prevalent algorithm employed to calculate the measure had a time complexity that was exponential in the number of objectives (Knowles 2002). This has constrained the application of the hypervolume metric to problems with a small number of objectives. However, Fleischer (2003) has now developed an efficient polynomial-time algorithm that should enable widespread uptake of the measure.

Fleischer (2003) also formally proved that the necessary and sufficient condition for all solutions in the corresponding approximation set to be Pareto optimal is the maximisation of hypervolume. Thus, the hypervolume measure can be used for selection purposes to advance the search. This type of approach has already been suggested by Knowles and Corne (2003b) for application at the selection-for-survival stage. In this method, new solutions are selected depending on the contribution they make to the hypervolume. With the advances made in methods to compute the measure, the appeal of approaches of this nature is strengthened. Thus, more innovations based on the hypervolume metric are predicted.

Diversity Promotion

The mechanisms used to promote diversity in an approximation set have arguably proved the most problematic for EMO researchers. In addition to the difficulties such methods introduce when attempting to obtain an algorithm with good theoretical convergence properties, their effectiveness is often very sensitive to parameter settings. The majority of density estimators used in EMO, such as fitness sharing and nearest-neighbour variants, operate in Euclidean space and require the aggregation of the objectives. This is undesirable, since one of the main advantages cited for favouring EMO over classical approaches is that the use of dominance-based comparisons prevents the requirement for any such cohesion. Diversity enhancement mechanisms that do not require such an aggregation include the ϵ-dominance approach of Laumanns, Thiele, Deb and Zitzler (2002) and methods that make use of distributed topologies such as that described by Rowe *et al.* (1996). Further research in this area is required.

Incorporation of DM Preferences

Decision-maker preference data can be incorporated into EMO algorithms to provide a search that is more efficient and produces results that are of greater relevance to the DM than those achieved through pure Pareto optimality. Despite these advantages, little research effort has been devoted to the incorporation of such methods within MOEAs. This is likely to be because most EMO research has only considered bi-objective problems in which the benefits of preference articulation are less apparent. However several excellent preference schemes do exist for EMO, of which Fonseca and Fleming's (1998b) goal and priority method and Todd and Sen's (1999) neural network scheme are particularly notable. The former provides a rigorously defined unification of many OR schemes in the literature using a modified dominance relation, whilst the latter attempts to automate DM preferences via machine learning.

Despite these, and similar, methods the sophistication of DM preference exploitation in EMO remains low. Essential research is required, perhaps borrowing techniques from the wider MCDM field, into (i) the facilitation of multiple DM entities organised in a hierarchy and (ii) methods for handling uncertain preference data.

Real-World Considerations

The immediate concerns of the analyst developing a solution to a real-world problem can be summarised as the *speed* (in real-time) within which the optimiser can produce results and the *confidence* that the analyst has in those results. A brief discussion of these issues completes the future perspectives offered in the thesis.

- **Speed.** Whilst the computational complexity of some MOEA operations may be larger than desired, the main resource-related factor that is likely to trouble the analyst is the amount of (real) time required to perform a candidate solution evaluation. Whilst it is tempting to speculate that these concerns will be remedied by the ever increasing power-to-cost ratio of computer hardware, more innovative and immediate solutions can be sought elsewhere. Note also that techniques such as parallel processing have for many years been cited, correctly, as being of benefit to EAs (for example, the *farmer-worker* approach to conducting evaluations) but their use continues to remain the exception rather than the rule. Advances in metamodelling are predicted, both in the type of approximate models used and the way in which they are integrated within the holistic optimisation process. The enhanced efficiency benefits of *memetic* algorithms (metaheuristics combined with local search methods) are also likely to be evident. Techniques and heuristics extracted from competent EA theory will also have a more prominent role in improving MOEA efficiency.

- **Confidence.** The key to building confidence in the ability of an MOEA to produce good solutions to a particular problem is to obtain knowledge of how MOEA components perform in similar problem environments. This knowledge may be theoretical upper bounds on the number of evaluations required for convergence or, and perhaps more likely, it may be previous experience gained from empirical studies on similar problems. The incorporation of elements from Goldberg's (2002) competent EA methodology should also help to improve the confidence of both the analyst and decision-maker in the performance of a prospective EMO application.

Appendix A

ZDT Benchmark Suite

Introduction

The bi-objective problems in the suite proposed by Zitzler *et al.* (2000) are of the following form:

$$\left.\begin{array}{lrcl} \text{min.} & z_1(\boldsymbol{x}), & & \\ \text{min.} & z_2(\boldsymbol{x}) & = & g(\boldsymbol{x})h\left(z_1(\boldsymbol{x}), g(\boldsymbol{x})\right), \\ \text{w.r.t.} & \boldsymbol{x} & = & [x_1, \ldots, x_n]\,. \end{array}\right\} \tag{A.1}$$

The functions $z_1(\boldsymbol{x})$, $g(\boldsymbol{x})$, and $h\left(z_1(\boldsymbol{x}), g(\boldsymbol{x})\right)$, together with ancillary functions and the decision vector $\boldsymbol{x}$, are defined for each particular problem below:

ZDT-1

The equations for ZDT-1 are shown in Equation A.2, where $n = 30$ is the number of decision variables, $\boldsymbol{x} \in [0\ 1]^n$.

$$\left.\begin{array}{lcl} z_1(\boldsymbol{x}) & = & x_1 \\ g(\boldsymbol{x}) & = & 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i \\ h(z_1, g) & = & 1 - \sqrt{\frac{z_1}{g}} \end{array}\right\} \tag{A.2}$$

ZDT-2

The equations for ZDT-2 are shown in Equation A.3, where $n = 30$ is the number of decision variables, $\boldsymbol{x} \in [0\ 1]^n$.

$$\left.\begin{array}{lcl} z_1(\boldsymbol{x}) & = & x_1 \\ g(\boldsymbol{x}) & = & 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i \\ h(z_1, g) & = & 1 - \left(\frac{z_1}{g}\right) \end{array}\right\} \tag{A.3}$$

ZDT-3

The equations for ZDT-3 are shown in Equation A.4, where $n = 30$ is the number of decision variables, $\boldsymbol{x} \in [0\ 1]^n$.

$$\left.\begin{array}{lcl} z_1(\boldsymbol{x}) & = & x_1 \\ g(\boldsymbol{x}) & = & 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i \\ h(z_1, g) & = & 1 - \sqrt{\frac{z_1}{g}} - \left(\frac{z_1}{g}\right)\sin(10\pi z_1) \end{array}\right\} \tag{A.4}$$

ZDT-4

The equations for ZDT-4 are shown in Equation A.5, where $n = 10$ is the number of decision variables. $x_1 \in [0\ 1]$ and $x_{2,\ldots,n} \in [-5\ 5]$.

$$\left.\begin{array}{lcl} z_1(\boldsymbol{x}) & = & x_1 \\ g(\boldsymbol{x}) & = & 1 + 10(n-1) + \sum_{i=2}^{n}\left(x_i^2 - 10\cos(4\pi x_i)\right) \\ h(z_1, g) & = & 1 - \sqrt{\frac{z_1}{g}} \end{array}\right\} \tag{A.5}$$

ZDT-5

The equations for ZDT-5 are shown in Equation A.6, where $n = 11$ is the number of decision variables. $x_1 \in \{0,1\}^{30}$ and $x_{2,\ldots,n} \in \{0,1\}^5$.

$$\left.\begin{array}{lcl} z_1(\boldsymbol{x}) & = & 1 + u(x_1) \\ u(x_1) & = & \sum_{k=1}^{5} (x_i(k) \wedge 1) \\ g(\boldsymbol{x}) & = & \sum_{i=2}^{n} v\left(u(x_i)\right) \\ v\left(u(x_i)\right) & = & \begin{cases} 2 + u(x_i) & \text{if} \quad u(x_i) < 5 \\ 1 & \text{if} \quad u(x_i) = 5 \end{cases} \\ h(z_1, g) & = & \frac{1}{z_1} \end{array}\right\} \tag{A.6}$$

ZDT-6

The equations for ZDT-6 are shown in Equation A.7, where $n = 10$ is the number of decision variables, $\boldsymbol{x} \in [0\ 1]^n$.

$$\left.\begin{array}{lcl} z_1(\boldsymbol{x}) & = & 1 - \mathrm{e}^{-4x_1} \sin^6(6\pi x_1) \\ g(\boldsymbol{x}) & = & 1 + 9\left(\left(\sum_{i=2}^{n} x_i\right)/9\right)^{\frac{1}{4}} \\ h(z_1, g) & = & 1 - \left(\frac{z_1}{g}\right) \end{array}\right\} \tag{A.7}$$

References

Abbass, H. A., Sarker, R. and Newton, C.: 2001, PDE: A Pareto-frontier differential evolution approach for multi-objective optimization problems, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2001 Congress on Evolutionary Computation (CEC 2001)*, Vol. 2, IEEE Service Center, Piscataway, New Jersey, pp. 971–978.

Agrell, P.: 1997, On redundancy in multi criteria decision-making, *European Journal of Operational Research* **98**, 571–586.

Anderson, M. B. and Lawrence, W. R.: 1996, Launch conditions and aerodynamic data extraction by an elitist Pareto genetic algorithm, *AIAA Atmospheric Flight Mechanics Conference*, American Institute of Aeronautics and Astronautics, Reston, Virginia. AIAA Paper 96-3361.

Aouni, B. and Kettani, O.: 2001, Goal programming model: A glorious history and a promising future, *European Journal of Operational Research* **133**(2), 225–231.

Bäck, T.: 1996, *Evolutionary Algorithms in Theory and Practice*, Oxford University Press, New York.

Baker, J. E.: 1987, Reducing bias and inefficiency in the selection algorithm, *in* J. Grefenstette (ed.), *Proceedings of the Second International Conference on Genetic Algorithms*, Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 14–21.

Baluja, S.: 1994, Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning, *Technical Report CMU-CS-94-163*, Carnegie Mellon University, Pittsburgh.

Booker, L. B.: 1982, *Intelligent Behaviour as an Adaptation to the Task Environment*, PhD thesis, University of Michigan, Ann Arbor. Cited by Deb and Goldberg (1989).

Bosman, P. A. N. and Thierens, D.: 1999, An algorithmic framework for density estimation based evolutionary algorithms, *Technical Report UU-CS-1999-46*, Utrecht University, Utrecht.

Bosman, P. A. N. and Thierens, D.: 2000, Mixed IDEAS, *Technical Report UU-CS-2000-45*, Utrecht University, Utrecht.

Bosman, P. A. N. and Thierens, D.: 2003, The balance between proximity and diversity in multiobjective evolutionary algorithms, *IEEE Transactions on Evolutionary Computation* **7**(2), 174–188.

Bouysson, D.: 1990, Building criteria: a prerequisite for MCDA, *in* C. B. e Costa (ed.), *Readings in Multiple Criteria Decision Aid*, Springer-Verlag, Berlin, pp. 58–80.

Branke, J., Kaußler, T. and Schmeck, H.: 2001, Guidance in evolutionary multi-objective optimization, *Advances in Engineering Software* **32**, 499–507.

Brindle, A.: 1981, *Genetic Algorithms for Function Optimization*, PhD thesis, University of Alberta, Edmonton. Cited by Goldberg and Deb (1991).

Bull, L.: 1999, On model-based evolutionary computing, *Soft Computing* **3**(2), 183–190.

Carlsson, C. and Fullér, R.: 1995, Multiple criteria decision making: The case for interdependence, *Computers and Operations Research* **119**, 251–260.

Censor, Y.: 1977, Pareto optimality in multiobjective problems, *Applied Mathematics and Optimization* **4**, 41–59.

Chernoff, H.: 1973, The use of faces to represent points in k-dimensional space graphically, *Journal of the American Statistical Association* **68**, 361–368.

Chipperfield, A. J. and Fleming, P. J.: 1995, Parallel genetic algorithms, *in* A. Zomaya (ed.), *Parallel and Distributed Computing Handbook*, McGraw-Hill, New York, pp. 1034–1059.

Chipperfield, A. J., Fleming, P. J., Pohlheim, H. and Fonseca, C. M.: 1994, Genetic algorithm toolbox user's guide, *ACSE Research Report 512*, University of Sheffield.

Cieniawski, S. E.: 1993, *An investigation of the ability of genetic algorithms to generate the trade-off curve of a multi-objective groundwater monitoring problem*, Master's thesis, University of Illinois at Urbana-Champaign. Cited by Fonseca and Fleming (1995b).

Cieniawski, S. E., Eheart, J. W. and Ranjithan, S.: 1995, Using genetic algorithms to solve a multiobjective groundwater monitoring problem, *Water Resources Research* **31**(2), 399–409.

Cleveland, W. S.: 1993, *Visualizing Data*, Hobart Press, Summit, New Jersey.

Coello, C. A. C.: 1999, A comprehensive survey of evolutionary-based multiobjective optimization techniques, *Knowledge and Information Systems. An International Journal* **1**(3), 269–308.

Coello, C. A. C.: 2000, Handling preferences in evolutionary multiobjective optimization: a survey, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2000 Congress on Evolutionary Computation (CEC 2000)*, Vol. 1, IEEE Service Center, Piscataway, New Jersey, pp. 30–37.

Coello, C. A. C.: 2002, Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: A survey of the state of the art, *Computer Methods in Applied Mechanics and Engineering* **191**(11–12), 1245–1287.

Coello, C. A. C. and Toscano Pulido, G.: 2001, A micro-genetic algorithm for multiobjective optimization, *in* E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello and D. Corne (eds), *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, Springer-Verlag, Berlin, pp. 126–140.

Coello, C. A. C., Veldhuizen, D. A. V. and Lamont, G. B.: 2002, *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer Academic Publishers, New York.

Cohoon, J., Martin, W. and Richards, D.: 1991, A multi-population genetic algorithm for solving the k-partition problem on hyper-cubes, *in* R. K. Belew and L. B. Booker (eds), *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann Publishers, San Mateo, California.

Collette, Y., Siarry, P. and Wong, H.-I.: 2000, Multidimensional data representation aimed at the interpretation of results from multiobjective optimization, *in* J. D. Knowles (ed.), *PPSN/SAB Workshop on Multiobjective Problem Solving from Nature (MPSN).*

Corne, D. W., Jerram, N. R., Knowles, J. D. and Oates, M. J.: 2001, PESA-II: Region-based selection in evolutionary multiobjective optimization, *in* L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon and E. Burke (eds), *Proceedings of the 2001 Genetic and Evolutionary Computation Conference (GECCO 2001)*, Morgan Kaufmann, pp. 283–290.

Corne, D. W., Knowles, J. D. and Oates, M. J.: 2000, The Pareto envelope-based selection algorithm for multiobjective optimization, *in* M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo and H.-P. Schwefel (eds), *Proceedings of the Parallel Problem Solving from Nature VI Conference*, Springer-Verlag, Berlin, pp. 839–848.

Cvetković, D. and Parmee, I. C.: 2002, Preferences and their application in evolutionary multiobjective optimisation, *IEEE Transactions on Evolutionary Computation* **6**(1), 42–57.

Dakev, N. V., Whidborne, J. F., Chipperfield, A. J. and Fleming, P. J.: 1997, H_∞ design of an EMS control system for a maglev vehicle using evolutionary algorithms, *Proceedings of the Institution of Mechanical Engineers Part I: Journal of Systems and Control Engineering* **311**(4), 345–355.

Darwin, C.: 1859, *The Origin of Species*, John Murray, London.

De Jong, K. A.: 1975, *An Analysis of the Behaviour of a Class of Genetic Adaptive Systems*, PhD thesis, University of Michigan, Ann Arbor, Michigan.

Deb, K.: 1999a, Evolutionary algorithms for multi-criterion optimization in engineering design, *in* K. Miettinen, M. M. Mäkelä, P. Neittaanmäki and J. Periaux (eds), *Evolutionary Algorithms in Engineering and Computer Science*, John Wiley & Sons, Chichester, UK, chapter 8, pp. 135–161.

Deb, K.: 1999b, Multi-objective evolutionary algorithms: Introducing bias among Pareto-optimal solutions, *KanGAL Report 99002*, Indian Institute of Technology Kanpur, Kanpur.

Deb, K.: 2001a, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, Chichester, UK.

Deb, K.: 2001b, Nonlinear goal programming using multi-objective genetic algorithms, *Journal of the Operational Research Society* **52**(3), 291–302.

Deb, K. and Agrawal, R. B.: 1995, Simulated binary crossover for continuous search space, *Complex Systems* **9**(2), 115–148.

Deb, K. and Goel, T.: 2001, Controlled elitist non-dominated sorting genetic algorithms for better convergence, *in* E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello and D. Corne (eds), *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, Springer-Verlag, pp. 67–81.

Deb, K. and Goldberg, D. E.: 1989, An investigation of niche and species formation in genetic function optimization, *in* J. D. Schaffer (ed.), *Proceedings of the Third International Conference on Genetic Algorithms*, Morgan Kaufmann Publishers, San Mateo, California, pp. 42–50.

Deb, K. and Goyal, M.: 1996, A combined genetic adaptive search (GeneAS) for engineering design, *Computer Science and Informatics* **26**(4), 30–45.

Deb, K. and Jain, S.: 2002, Running performance metrics for evolutionary multi-objective optimization, *KanGAL Report 200200*, Indian Institute of Technology Kanpur, Kanpur.

Deb, K., Mohan, M. and Mishra, S.: 2003, Towards a quick computation of well-spread Pareto-optimal solutions, *in* C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb and L. Thiele (eds), *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, Springer-Verlag, Berlin, pp. 222–236.

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T.: 2002, A fast and elitist multiobjective genetic algorithm: NSGA–II, *IEEE Transactions on Evolutionary Computation* **6**(2), 182–197.

Deb, K., Thiele, L., Laumanns, M. and Zitzler, E.: 2002, Scalable multi-objective optimization test problems, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2002*

Congress on Evolutionary Computation (CEC 2002), Vol. 1, IEEE Service Center, Piscataway, New Jersey, pp. 825–830.

Deb, K., Zope, P. and Jain, A.: 2003, Distributed computing of Pareto-optimal solutions with evolutionary algorithms, *in* C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb and L. Thiele (eds), *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, Springer, Berlin, pp. 534–549.

Demartines, P. and Hérault, H.: 1997, Curvilinear component analysis: a self-organising neural network for non-linear mapping of data sets, *IEEE Transactions on Neural Networks* **8**, 148–154.

Drechsler, N., Drechsler, R. and Becker, B.: 2001, Multi-objective optimisation based on relation *favour*, *in* E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello and D. Corne (eds), *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, Springer-Verlag, pp. 154–166.

Edgeworth, F. Y.: 1932, *Mathematical Psychics. An Essay on the Application of Mathematics to the Moral Sciences*, reprint edn, The London School of Economics and Political Science, London.

Eshelman, L. J. and Schaffer, J. D.: 1993, Crossover's niche, *in* S. Forrest (ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms*, Morgan Kaufmann Publishers, San Mateo, California, pp. 9–14.

Everson, R. M., Fieldsend, J. E. and Singh, S.: 2002, Full elite sets for multi-objective optimisation, *in* I. C. Parmee (ed.), *Proceedings of the Fifth International Conference on Adaptive Computing Design and Manufacture (ACDM 2002)*, Springer-Verlag, Berlin, pp. 343–354.

Farhang-Mehr, A. and Azarm, S.: 2002, Diversity assessment of Pareto optimal solution sets: An entropy approach, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2002 Congress on Evolutionary Computation (CEC 2002)*, Vol. 1, IEEE Service Center, Piscataway, New Jersey, pp. 825–830.

Farina, M. and Amato, P.: 2002, On the optimal solution definition for many-criteria optimization problems, *in* J. Keller and O. Nasraoui (eds), *Proceedings of the 2002*

NAFIPS-FLINT International Conference, IEEE Service Center, Piscataway, New Jersey, pp. 233–238.

Fisher, R.: 1930, *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford.

Fleischer, M.: 2003, The measure of Pareto optima. applications to multi-objective metaheuristics, *in* C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb and L. Thiele (eds), *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, Springer, Berlin, pp. 519–533.

Fleming, P. J. and Pashkevich, A. P.: 1985, Computer aided control system design using a multiobjective optimization approach, *Proceedings of the IEE Control '85 Conference*, pp. 174–179.

Fleming, P. J. and Purshouse, R. C.: 2001a, Evolutionary computing: Opportunities for control, *in* H. Verbruiggen, C. W. Chan and R. A. Vingerhoeds (eds), *Proceedings of the IFAC Conference on New Technologies for Computer Control (NTCC 2001)*, Elsevier Science, Oxford, UK, pp. 21–32.

Fleming, P. J. and Purshouse, R. C.: 2001b, Genetic algorithms in control systems engineering, *in* P. Albertos, J. Bokor, A. Crespo, P. Fleming, A. Haddad, R. Isermann, J. Park and M. Zaremba (eds), *IFAC Professional Briefs*, International Federation of Automatic Control.

Fleming, P. J. and Purshouse, R. C.: 2002, Evolutionary algorithms in control systems engineering: A survey, *Control Engineering Practice* **10**, 1223–1241.

Fonseca, C. M. and Fleming, P. J.: 1993, Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization, *in* S. Forrest (ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms*, Morgan Kauffman Publishers, San Mateo, California, pp. 416–423.

Fonseca, C. M. and Fleming, P. J.: 1995a, Multiobjective genetic algorithms made easy: Selection, sharing, and mating restriction, *in* A. M. S. Zalzala (ed.), *Proceedings of the First International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications (GALESIA 95)*, Institution of Electrical Engineers, Stevenage, UK, pp. 42–52.

Fonseca, C. M. and Fleming, P. J.: 1995b, An overview of evolutionary algorithms in multiobjective optimization, *Evolutionary Computation* **3**(1), 1–16.

Fonseca, C. M. and Fleming, P. J.: 1996, On the performance assessment and comparison of stochastic multiobjective optimizers, *in* H.-M. Voigt, W. Ebeling, I. Rechenberg and H.-P. Schwefel (eds), *Proceedings of the Parallel Problem Solving from Nature IV Conference*, Springer-Verlag, Berlin, pp. 584–593.

Fonseca, C. M. and Fleming, P. J.: 1997, Multiobjective optimization, *in* T. Bäck, D. B. Fogel and Z. Michalewicz (eds), *Handbook of Evolutionary Computation*, Vol. 1, Institute of Physics Publishing and Oxford University Press, New York, pp. C4.5:1–C4.5:9.

Fonseca, C. M. and Fleming, P. J.: 1998a, Multiobjective optimization and multiple constraint handling with evolutionary algorithms — Part I: A unified formulation, *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans* **28**(1), 26–37.

Fonseca, C. M. and Fleming, P. J.: 1998b, Multiobjective optimization and multiple constraint handling with evolutionary algorithms — Part II: An application example, *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans* **28**(1), 38–47.

Fonseca, C. M., Fleming, P. J., Zitzler, E., Deb, K. and Thiele, L. (eds): 2003, *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, Vol. 2632 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin.

Fourman, M. P.: 1985, Compaction of symbolic layout using genetic algorithms, *in* J. Grefenstette (ed.), *Proceedings of the First International Conference on Genetic Algorithms*, Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 141–153.

Gal, T. and Hanne, T.: 1999, Consequences of dropping nonessential objectives for the application of MCDM methods, *European Journal of Operational Research* **119**, 373–378.

Goldberg, D. E.: 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company, Reading, Massachusetts.

Goldberg, D. E.: 1998, The race, the hurdle, and the sweet spot: Lessons from genetic algorithms for the automation of design innovation and creativity, *IlliGAL Report 98007*, University of Illinois at Urbana-Champaign, Urbana, Illinois.

Goldberg, D. E.: 2002, *The Design of Innovation: Lessons From and For Competent Genetic Algorithms*, number 7 in *Genetic Algorithms and Evolutionary Computation*, Kluwer Academic Publishers, New York.

Goldberg, D. E. and Deb, K.: 1991, A comparative analysis of selection schemes used in genetic algorithms, *in* G. J. E. Rawlins (ed.), *Foundations of Genetic Algorithms*, Morgan Kaufmann, San Mateo, California, pp. 69–93.

Goldberg, D. E. and Richardson, J.: 1987, Genetic algorithms with sharing for multimodal function optimization, *in* J. Grefenstette (ed.), *Proceedings of the Second International Conference on Genetic Algorithms*, Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 41–49.

Gordon, V., Whitley, D. and Böhm, A.: 1992, Dataflow parallelism in genetic algorithms, *in* R. Männer and B. Manderick (eds), *Proceedings of the Parallel Problem Solving from Nature II Conference*, Springer-Verlag, Berlin, pp. 533–542.

Gray, G. J., Murray-Smith, D. J., Li, Y., Sharman, K. C. and Weinbrunner, T.: 1998, Nonlinear model structure identification using genetic programming, *Control Engineering Practice* **6**(11), 1341–1352.

Griffin, I. A., Schroder, P., Chipperfield, A. J. and Fleming, P. J.: 2000, Multi-objective optimization approach to the ALSTOM gasifier problem, *Proceedings of the Institution of Mechanical Engineers Part I: Journal of Systems and Control* **214**, 453–468.

Grunert da Fonseca, V., Fonseca, C. M. and Hall, A. O.: 2001, Inferential performance assessment of stochastic optimisers and the attainment function, *in* E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello and D. Corne (eds), *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, Springer-Verlag, pp. 213–225.

Gunawan, S., Farhang-Mehr, A. and Azarm, S.: 2003, Multi-level multi-objective genetic algorithm using entropy to preserve diversity, *in* C. M. Fonseca, P. J. Fleming, E. Zitzler,

K. Deb and L. Thiele (eds), *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, Springer, Berlin, pp. 148–161.

Haleja, P. and Lin, C.-Y.: 1992, Genetic search strategies in multicriterion optimal design, *Structural Optimization* **4**, 99–107.

Helbig, S. and Pateva, D.: 1994, On several concepts for ϵ-efficiency, *OR Spektrum* **16**(3), 179–186.

Herrera, F., Lozano, M. and Verdegay, J. L.: 1998, Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis, *Artificial Intelligence Review* pp. 265–319.

Hiroyasu, T., Miki, M. and Watanabe, S.: 2000, The new model of parallel genetic algorithm in multi-objective optimization problems — divided range multi-objective genetic algorithm, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2000 Congress on Evolutionary Computation (CEC 2000)*, Vol. 1, IEEE Service Center, Piscataway, New Jersey, pp. 333–340.

Holland, J. H.: 1975, *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, Michigan.

Hollander, M. and Wolfe, D. A.: 1999, *Nonparametric Statistical Methods*, 2nd edn, Wiley, New York.

Horn, J.: 1997, *The Nature of Niching: Genetic Algorithms and the Evolution of Optimal, Cooperative Populations*, PhD thesis, University of Illinois at Urbana-Champaign, Urbana, Illinois.

Horn, J. and Nafpliotis, N.: 1993, Multiobjective optimization using the niched Pareto genetic algorithm, *IlliGAL Report 93005*, University of Illinois at Urbana-Champaign, Urbana, Illinois.

Hughes, E. J.: 2001, Evolutionary multi-objective ranking with uncertainty and noise, *in* E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello and D. Corne (eds), *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, Springer-Verlag, pp. 329–343.

Ikeda, K., Kita, H. and Kobayashi, S.: 2001, Failure of Pareto-based MOEAs: Does non-dominated really mean near to optimal?, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2001 Congress on Evolutionary Computation (CEC 2001)*, Vol. 2, IEEE Service Center, Piscataway, New Jersey, pp. 957–962.

Inselberg, A.: 1985, The plane with parallel coordinates, *The Visual Computer* **1**, 69–91.

Ishibuchi, H. and Murata, T.: 1996, Multi-objective genetic local search algorithm, *in* T. Fukuda and T. Furuhashi (eds), *Proceedings of the 1996 International Conference on Evolutionary Computation*, IEEE Service Center, Piscataway, New Jersey, pp. 119–124.

Jin, Y., Okabe, T. and Sendhoff, B.: 2001a, Adapting weighted aggregation for multiobjective evolution strategies, *in* E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello and D. Corne (eds), *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, Springer-Verlag, Berlin, pp. 96–110.

Jin, Y., Okabe, T. and Sendhoff, B.: 2001b, Dynamic weighted aggregation for evolutionary multi-objective optimization: Why does it work and how?, *in* L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon and E. Burke (eds), *Proceedings of the 2001 Genetic and Evolutionary Computation Conference (GECCO 2001)*, Morgan Kaufmann Publishers, San Francisco, California, pp. 1042–1049.

Joliffe, I. T.: 1986, *Principal Component Analysis*, Springer, New York.

Kendall, M. G.: 1938, A new measure of rank correlation, *Biometrika* **30**, 81–93.

Khan, N., Goldberg, D. E. and Pelikan, M.: 2002, Multi-objective Bayesian optimization algorithm, *in* W. B. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke and N. Jonoska (eds), *Proceedings of the 2002 Genetic and Evolutionary Computation Conference (GECCO 2002)*, Morgan Kaufmann Publishers, San Francisco, California, p. 684.

Khare, V., Yao, X. and Deb, K.: 2003, Performance scaling of multi-objective evolutionary algorithms, *in* C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb and L. Thiele (eds),

Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003), Springer, Berlin, pp. 376–390.

Knowles, J. D.: 2002, *Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization*, PhD thesis, The University of Reading, Reading, UK.

Knowles, J. D. and Corne, D. W.: 1999, The Pareto archived evolution strategy: A new baseline algorithm for multiobjective optimisation, IEEE Service Center, Piscataway, New Jersey, pp. 98–105.

Knowles, J. D. and Corne, D. W.: 2000, Approximating the nondominated front using the Pareto archived evolution strategy, *Evolutionary Computation* **8**(2), 149–172.

Knowles, J. D. and Corne, D. W.: 2002, On metrics for comparing nondominated sets, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2002 Congress on Evolutionary Computation (CEC 2002)*, Vol. 1, IEEE Service Center, Piscataway, New Jersey, pp. 711–716.

Knowles, J. D. and Corne, D. W.: 2003a, Instance generators and test suites for the multiobjective quadratic assignment problem, *in* C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb and L. Thiele (eds), *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, Springer, Berlin, pp. 295–310.

Knowles, J. D. and Corne, D. W.: 2003b, Properties of an adaptive archiving algorithm for storing nondominated vectors, *IEEE Transactions on Evolutionary Computation* **7**(2), 100–116.

Kohonen, T.: 1995, *Self-Organizing Maps*, Springer, Berlin.

Koza, J. R.: 1992, *Genetic Programming — on the Programming of Computers by Means of Natural Selection*, The MIT Press, Cambridge, Massachusetts.

Kursawe, F.: 1991, A variant of evolution strategies for vector optimization, *in* H.-P. Schwefel and R. Männer (eds), *Proceedings of the Parallel Problem Solving from Nature I Conference*, Springer-Verlag, Berlin, pp. 193–197.

Laumanns, M. and Ocenasek, J.: 2002, Bayesian optimization algorithms for multi-objective optimization, *in* J. J. Merelo Guervós, P. Adamidis, H.-G. Beyer, J.-L. Fernández-Villacañas and H.-P. Schwefel (eds), *Proceedings of the Parallel Problem Solving from Nature VII Conference*, Springer-Verlag, Berlin, pp. 298–307.

Laumanns, M., Thiele, L., Deb, K. and Zitzler, E.: 2002, Combining convergence and diversity in evolutionary multi-objective optimization, *Evolutionary Computation* **10**(3), 263–282.

Laumanns, M., Thiele, L., Zitzler, E., Welzl, E. and Deb, K.: 2002, Running time analysis of multi-objective evolutionary algorithms on a simple discrete optimization problem, *in* J. J. Merelo Guervós, P. Adamidis, H.-G. Beyer, J.-L. Fernández-Villacañas and H.-P. Schwefel (eds), *Proceedings of the Parallel Problem Solving from Nature VII Conference*, Springer-Verlag, Berlin, pp. 44–53.

Laumanns, M., Zitzler, E. and Thiele, L.: 2000, A unified model for multi-objective evolutionary algorithms with elitism, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2000 Congress on Evolutionary Computation (CEC 2000)*, Vol. 1, IEEE Service Center, Piscataway, New Jersey, pp. 46–53.

Laumanns, M., Zitzler, E. and Thiele, L.: 2001, On the effects of archiving, elitism, and density based selection in evolutionary multi-objective optimization, *in* E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello and D. Corne (eds), *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, Springer-Verlag, Berlin, pp. 181–196.

Lee, J. and Kuo, J.-Y.: 1998, Fuzzy decision making through trade-off analysis between criteria, *Journal of Information Sciences* **107**, 107–126.

Lobo, F. G. and Goldberg, D. E.: 2001, The parameter-less genetic algorithm in practice, *IlliGAL Report 2001022*, University of Illinois at Urbana-Champaign, Urbana, Illinois.

Lohn, J. D., Kraus, W. F. and Haith, G. L.: 2002, Comparing a coevolutionary genetic algorithm for multiobjective optimization, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2002 Congress on Evolutionary Computation (CEC 2002)*, Vol. 2, IEEE Service Center, Piscataway, New Jersey, pp. 1157–1162.

Lu, H. and Yen, G.: 2002, Rank-density based multiobjective genetic algorithm, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2002 Congress on Evolutionary Computation (CEC 2002)*, Vol. 1, IEEE Service Center, Piscataway, New Jersey, pp. 944–949.

Manly, B. F. J.: 1991, *Randomization and Monte Carlo Methods in Biology*, Chapman and Hall, London.

Massebeuf, S., Fonteix, C., Kiss, L. N., Marc, I., Pla, F. and Zaras, K.: 1999, Multicriteria optimization and decision engineering of an extrusion process aided by a diploid genetic algorithm, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 1999 Congress on Evolutionary Computation (CEC 99)*, IEEE Service Center, Piscataway, New Jersey, pp. 14–21.

Michalewicz, Z.: 1992, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Berlin.

Michalewicz, Z. and Fogel, D. B.: 2000, *How to Solve It: Modern Heuristics*, Springer, Berlin.

Morse, J. N.: 1980, Reducing the size of the nondominated set: Pruning by clustering, *Computers and Operations Research* **7**(1–2), 55–66.

Murata, T., Ishibuchi, H. and Tanaka, H.: 1996, Multi-objective genetic algorithm and its application to flowshop scheduling, *Computers and Industrial Engineering* **30**(4), 957–968.

Myers, R. H. and Montgomery, D. C.: 2002, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd edn, Wiley, Chichester.

Noble, J. and Watson, R. A.: 2001, Pareto coevolution: Using performance against coevolved opponents in a game as dimensions for Pareto selection, *in* L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon and E. Burke (eds), *Proceedings of the 2001 Genetic and Evolutionary Computation Conference (GECCO 2001)*, Morgan Kaufmann Publishers, San Francisco, California, pp. 493–500.

Obayashi, S. and Sasaki, D.: 2003, Visualization and data mining of Pareto solutions using self-organizing map, *in* C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb and

L. Thiele (eds), *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, Springer, Berlin, pp. 796–809.

Oei, C. K., Goldberg, D. E. and Chang, S.-J.: 1991, Tournament selection, niching, and the preservation of diversity, *IlliGAL Report 91011*, University of Illinois at Urbana-Champaign, Urbana, Illinois.

Okabe, T., Jin, Y. and Sendhoff, B.: 2002, On the dynamics of evolutionary multi-objective optimization, *in* W. B. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke and N. Jonoska (eds), *Proceedings of the 2002 Genetic and Evolutionary Computation Conference (GECCO 2002)*, Morgan Kaufmann Publishers, San Francisco, California, pp. 247–255.

Okuda, T., Hiroyasu, T., Miki, M. and Watanabe, S.: 2002, DCMOGA: Distributed cooperation model of multi-objective genetic algorithm, *in* J. D. Knowles (ed.), *Proceedings of the PPSN/SAB Workshop on Multiobjective Problem Solving from Nature II (MPSN-II)*.

Papadimitriou, C. H. and Yannakakis, M.: 2000, On the approximability of trade-offs and optimal access of Web sources, *Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS 2000)*, IEEE Computer Society, Los Alamitos, California, pp. 86–92.

Parks, G. T. and Miller, I.: 1998, Selective breeding in a multiobjective genetic algorithm, *in* A. E. Eiben, M. Schoenauer and H.-P. Schwefel (eds), *Proceedings of the Parallel Problem Solving From Nature V Conference*, Springer-Verlag, Berlin, pp. 250–259.

Parmee, I. C., Cvetković, D., Watson, A. H. and Bonham, C. R.: 2000, Multiobjective satisfaction within an interactive evolutionary design environment, *Evolutionary Computation* **8**(2), 197–222.

Pelikan, M. and Goldberg, D. E.: 2000, Genetic algorithms, clustering, and the breaking of symmetry, *in* M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo and H.-P. Schwefel (eds), *Proceedings of the Parallel Problem Solving from Nature VI Conference*, Springer-Verlag, Berlin, pp. 385–394.

Pelikan, M., Goldberg, D. E. and Cantú-Paz, E.: 1998, Linkage problem, distribution estimation, and Bayesian networks, *IlliGAL Report 98013*, University of Illinois at Urbana-Champaign, Urbana, Illinois.

Pelikan, M., Goldberg, D. E. and Lobo, F.: 1999, A survey of optimization by building and using probabilistic models, *IlliGAL Report 99018*, University of Illinois at Urbana-Champaign, Urbana, Illinois.

Pelikan, M. and Mühlenbein, H.: 1999, The bivariate marginal distribution algorithm, *in* R. Roy (ed.), *Advances in Soft Computing — Engineering Design and Manufacturing*, Springer-Verlag, London, pp. 521–535.

Potter, M. A. and De Jong, K. A.: 2000, Cooperative coevolution: An architecture for evolving coadapted subcomponents, *Evolutionary Computation* **8**(1), 1–29.

Purshouse, R. C. and Fleming, P. J.: 2001, The multi-objective genetic algorithm applied to benchmark problems — an analysis, *ACSE Research Report 796*, University of Sheffield, Sheffield, UK.

Purshouse, R. C. and Fleming, P. J.: 2002, Why use elitism and sharing in a multi-objective genetic algorithm?, *in* W. B. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke and N. Jonoska (eds), *Proceedings of the 2002 Genetic and Evolutionary Computation Conference (GECCO 2002)*, Morgan Kaufmann Publishers, San Francisco, California, pp. 520–527.

Purshouse, R. C. and Fleming, P. J.: 2003a, An adaptive divide-and-conquer methodology for evolutionary multi-criterion optimisation, *in* C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb and L. Thiele (eds), *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, Springer, Berlin, pp. 133–147.

Purshouse, R. C. and Fleming, P. J.: 2003b, Conflict, harmony, and independence: Relationships in evolutionary multi-criterion optimisation, *in* C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb and L. Thiele (eds), *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, Springer, Berlin, pp. 16–30.

Purshouse, R. C. and Fleming, P. J.: 2003c, Evolutionary many-objective optimisation: An exploratory analysis, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2003 Congress on Evolutionary Computation (CEC 2003)*, IEEE Service Center, Piscataway, New Jersey, pp. 2066–2073.

Purshouse, R. C. and Fleming, P. J.: 2003d, An exploration of the many-objective optimisation process for a class of evolutionary algorithms, *ACSE Research Report 840*, University of Sheffield, Sheffield, UK.

Ray, T., Kang, T. and Chye, S. K.: 2001, Multiobjective design optimization by an evolutionary algorithm, *Engineering Optimization* **33**(3), 399–424.

Rechenberg, I.: 1973, *Evolutionßtrategie: Optimierung technischer systeme nach prinzipien der biologischen evolution*, Frommann-Holzboog, Stuttgart.

Reed, P., Minsker, B. S. and Goldberg, D. E.: 2000, Designing a competent simple genetic algorithm for search and optimization, *Water Resources Research* **36**(12), 3757–3761.

Rekiek, B., De Lit, P., Pellichero, F., L'Eglise, T., Falkenauer, E. and Delchambre, A.: 2000, Dealing with user's preferences in hybrid assembly lines design, *in* Z. Binder (ed.), *Proceedings of the Second Conference on Management and Control of Production and Logistics (MCPL 2000)*, Elsevier Science, Oxford, UK.

Ritzel, B.: 1992, Genetic algorithms in multiple objective optimization. Unpublished analytical class project for GE 493, University of Illinois at Urbana-Champaign. Cited by Horn and Nafpliotis (1993).

Rosenman, M. A. and Gero, J. S.: 1985, Reducing the Pareto optimal set in multicriteria optimization, *Engineering Optimization* **8**, 189–206.

Rowe, J., Vinsen, K. and Marvin, N.: 1996, Parallel GAs for multiobjective functions, *in* J. T. Alander (ed.), *Proceedings of the Second Nordic Workshop on Genetic Algorithms and Their Applications (2NWGA)*, University of Vaasa, Vaasa, Finland, pp. 61–70.

Rudolph, G. and Agapie, A.: 2000, Convergence properties of some multi-objective evolutionary algorithms, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 2000*

Conference on Evolutionary Computation (CEC 2000), Vol. 2, IEEE Service Center, Piscataway, New Jersey, pp. 1010–1016.

Sarker, R., Liang, K. and Newton, C.: 2002, A new evolutionary algorithm for multiobjective optimization, *European Journal of Operational Research* **140**(1), 12–23.

Schaffer, J. D.: 1985, Multiple objective optimization with vector evaluated genetic algorithms, *in* J. Grefenstette (ed.), *Proceedings of the First International Conference on Genetic Algorithms*, Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 93–100.

Schott, J. R.: 1995, *Fault tolerant design using single and multicriteria genetic algorithm optimization*, Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Schroder, P.: 1998, *Multivariable Control of Active Magnetic Bearings*, PhD thesis, University of Sheffield, Sheffield, UK.

Scott, D. W.: 1992, *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley, New York.

Shaw, K. J. and Fleming, P. J.: 1997, Including real-life preferences in genetic algorithms to improve optimisation of production schedules, *in* A. M. S. Zalzala (ed.), *Proceedings of the Second International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications (GALESIA 97)*, Institution of Electrical Engineers, Stevenage, UK, pp. 239–244.

Shaw, K. J. and Fleming, P. J.: 2000, Genetic algorithms for scheduling: Incorporation of user preferences, *Transactions of the Institute of Measurement and Control* **22**(2), 195–210.

Silverman, B. W.: 1986, *Density Estimation for Statistics and Data Analysis*, Vol. 26 of *Monographs on Applied Statistics and Probability*, Chapman and Hall, London.

Smith, R., Dike, B. and Stegmann, S.: 1995, Fitness inheritance in genetic algorithms, *in* K. M. George (ed.), *Proceedings of the 1995 ACM Symposium on Applied Computing*, ACM Press, New York, pp. 345–350.

Srinivas, N. and Deb, K.: 1994, Multiobjective optimization using nondominated sorting in genetic algorithms, *Evolutionary Computation* **2**(3), 221–248.

Steuer, R. E.: 2001, An overview in graphs of multiple objective programming, *in* E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello and D. Corne (eds), *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimisation (EMO 2001)*, Springer-Verlag, Berlin, pp. 41–51.

Sun, M., Stam, A. and Steuer, R. E.: 2000, Interactive multiple objective programming using Tchebycheff programs and artificial neural networks, *Computers and Operations Research* **27**, 601–620.

Tamaki, H., Mori, M., Araki, M., Mishima, Y. and Ogai, H.: 1994, Multi-criteria optimization by genetic algorithms: A case of scheduling in hot rolling process, *in* M. Fushimi and K. Tone (eds), *Proceedings of the Third Conference of the Association of Asian-Pacific Operational Research Societies within IFORS (APORS 94)*, World Scientific Publishing, pp. 374–381.

Tan, K. C., Khor, E. F., Lee, T. H. and Sathikannan, R.: 2003, An evolutionary algorithm with advanced goal and priority specification for multi-objective optimization, *Journal of Artificial Intelligence Research* **18**, 183–215.

Tan, K. C., Lee, T. H. and Khor, E. F.: 1999, Evolutionary algorithms with goal and priority information for multi-objective optimization, *in* IEEE Neural Networks Council (ed.), *Proceedings of the 1999 Congress on Evolutionary Computation (CEC 99)*, IEEE Service Center, Piscataway, New Jersey, pp. 106–113.

Teich, J.: 2001, Pareto-front exploration with uncertain objectives, *in* E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello and D. Corne (eds), *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, Springer-Verlag, Berlin, pp. 314–328.

Thierens, D. and Bosman, P. A. N.: 2001a, Multi-objective mixture-based iterated density estimation evolutionary algorithms, *in* L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon and E. Burke (eds), *Proceedings of the 2001 Genetic and Evolutionary Computation Conference (GECCO 2001)*, Morgan Kaufmann Publishers, San Francisco, California, pp. 663–670.

Thierens, D. and Bosman, P. A. N.: 2001b, Multi-objective optimization with iterated density estimation evolutionary algorithms using mixture models, *Proceedings of the Third International Symposium on Adaptive Systems — Evolutionary Computation and Probabilistic Graphical Models*, Institute of Cybernetics, Mathematics and Physics, Havana, Cuba, pp. 129–136.

Thomas, M. W.: 1998, *A Pareto Frontier for Full Stern Submarines Via Genetic Algorithm*, PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Thompson, H. A., Chipperfield, A. J., Fleming, P. J. and Legge, C.: 1999, Distributed aero-engine control systems architecture selection using multi-objective optimisation, *Control Engineering Practice* **7**(5), 655–664.

Tiwari, A. and Roy, R.: 2002, Variable response interaction and multi-objective optimisation, *in* W. B. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke and N. Jonoska (eds), *Proceedings of the 2002 Genetic and Evolutionary Computation Conference (GECCO 2002)*, Morgan Kaufmann Publishers, San Francisco, California, pp. 602–609.

Todd, D. S. and Sen, P.: 1997, A multiple criteria genetic algorithm for containership loading, *in* T. Bäck (ed.), *Proceedings of the Seventh International Conference on Genetic Algorithms*, Morgan Kaufmann Publishers, San Mateo, California, pp. 674–681.

Todd, D. S. and Sen, P.: 1999, Directed multiple objective search of design spaces using genetic algorithms and neural networks, *in* W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela and R. E. Smith (eds), *Proceedings of the 1999 Genetic and Evolutionary Computation Conference (GECCO 99)*, Vol. 2, Morgan Kaufmann Publishers, San Francisco, California, pp. 1738–1743.

Valenzuela, C. L. and Jones, A. J.: 1994, Evolutionary divide-and-conquer (I): A novel genetic approach to the TSP, *Evolutionary Computation* **1**(4), 313–333.

Veldhuizen, D. A. V.: 1999, *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*, PhD thesis, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Veldhuizen, D. A. V. and Lamont, G. B.: 2000, Multiobjective evolutionary algorithms: Analyzing the state-of-the-art, *Evolutionary Computation* **8**(2), 125–147.

Veldhuizen, D. A. V., Zydallis, J. B. and Lamont, G. B.: 2003, Considerations in engineering parallel multiobjective evolutionary algorithms, *IEEE Transactions on Evolutionary Computation* **7**(2), 144–173.

Watson, R. A.: 2002, *Compositional Evolution: Interdisciplinary Investigations in Evolvability, Modularity, and Symbiosis*, PhD thesis, Brandeis University, Waltham, Massachusetts.

Wegman, E. J.: 1990, Hyperdimensional data analysis using parallel coordinates, *Journal of the American Statistical Association* **85**, 664–675.

Whitley, D.: 2001, An overview of evolutionary algorithms: Practical issues and common pitfalls, *Information and Software Technology* **43**, 817–831.

Wienke, P. B., Lucasius, C. and Kateman, G.: 1992, Multicriteria target optimization of analytical procedures using a genetic algorithm, *Analytica Chimica Acta* **265**(2), 211–225.

Wolpert, D. and Macready, W.: 1997, No free lunch theorems for optimization, *IEEE Transactions on Evolutionary Computation* **1**(1), 67–82.

Zeleny, M.: 1973, Compromise programming, *in* J. L. Cochrane and M. Zeleny (eds), *Multiple Criteria Decision Making*, University of South Carolina Press, Columbia, South Carolina, pp. 262–301.

Zitzler, E.: 1999, *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*, PhD thesis, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

Zitzler, E.: 2002, A tutorial on evolutionary multiobjective optimization, Invited Talk, 2002 Workshop on Multi-Objective Meta-Heuristics. 4–5 November 2002, Paris.

Zitzler, E., Deb, K. and Thiele, L.: 2000, Comparison of multiobjective evolutionary algorithms: Empirical results, *Evolutionary Computation* **8**(2), 173–195.

Zitzler, E., Deb, K., Thiele, L., Coello, C. A. C. and Corne, D. (eds): 2001, *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, Vol. 1993 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin.

Zitzler, E., Laumanns, M. and Thiele, L.: 2001, SPEA2: Improving the strength Pareto evolutionary algorithm, *TIK Report 103*, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

Zitzler, E. and Thiele, L.: 1998, Multiobjective optimization using evolutionary algorithms — a comparative study, *in* A. E. Eiben (ed.), *Proceedings of the Parallel Problem Solving from Nature V Conference*, Springer-Verlag, Berlin, pp. 292–301.

Zitzler, E. and Thiele, L.: 1999, Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach, *IEEE Transactions on Evolutionary Computation* **3**(4), 257–271.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M. and Grunert da Fonseca, V.: 2003, Performance assessment of multiobjective optimizers: An analysis and review, *IEEE Transactions on Evolutionary Computation* **7**(2), 117–132.