Monograph:
Pursehouse, R.C. and Fleming, P.J. (2003) An Exploration of the Many-Objective Optimisation Process for a Class of Evolutionary Algorithms. Research Report. ACSE Research Report 840 . Department of Automatic Control and Systems Engineering

An Exploration of the Many-Objective Optimisation Process for a Class of Evolutionary Algorithms

R. C. Purshouse

P. J. Fleming

Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, S1 3JD
UK

Research Report No. 840
June 2003

An Exploration of the Many-Objective Optimisation Process for a Class of Evolutionary Algorithms

R. C. Purshouse and **P. J. Fleming**
Department of Automatic Control and Systems Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK.
E-mail: {r.purshouse, p.fleming}@sheffield.ac.uk

Abstract

This empirical inquiry explores the behaviour of a particular class of evolutionary algorithms as the number of conflicting objectives to be simultaneously optimised is increased. Population-based optimisers that perform selection according to Pareto dominance and density estimation are considered. The performances of abstracted algorithms, based on decompositions of the fundamental components of a modern optimiser, are considered across a wide range of mutation and recombination operating conditions. Configuration sweet-spots for these algorithms are identified and contrasted. The classical mutation settings are shown to be a robust choice, even when the total sweet-spot is seen to contract as the number of objectives is increased. Classical settings for recombination, by contrast, are shown to work well for small numbers of objectives but lead to very poor performance as the number of objectives is increased. Mutation performance is demonstrated to be largely invariant of population size across the standard range of values. The performance of recombination can be somewhat improved by using larger population sizes. Explanations are offered for the observed behaviour of the evolutionary optimisers.

1 Introduction

1.1 Background

Much of the research into evolutionary multi-objective optimisation algorithms (MOEAs) concentrates on optimisation tasks with two conflicting objectives. However, the real-world challenges to which these algorithms are applied often feature many more objectives (Coello *et al*, 2002). Hence, there is a clear need to extend evolutionary multi-objective optimisation (EMO) research into the realm of *many-objectives*[1].

Interactions often arise between objectives. These can be classified as *conflicting* or *harmonious*, and both interactions may co-exist between two objectives in the context of a single optimisation problem (Purshouse and Fleming, 2003). In the case of conflict, a solution modification that will improve performance in one objective is seen to cause deterioration in a second objective. In the case of harmony, the modification causes simultaneous improvement to both objectives. The conflict that exists in a many-objective task has been identified as a serious challenge for contemporary EMO researchers, and it is this relationship that is explored in this paper.

If the only assumption concerning decision-maker (DM) preferences is that a unidirectional *line of preference* (Edgeworth, 1932) exists for each objective then performance comparisons between solutions can be based on the notion of *Pareto dominance* (Coello *et al*, 2002). In these conditions, the optimal solution to an M-objective task, in which all objectives conflict, is an $(M\text{-}1)$-dimensional hypersurface. The number of samples required to represent the

[1] To the best of the authors' knowledge, this terminology was first introduced by Farina and Amato (2002).

surface at a fixed resolution is exponential in M. Even if such an *approximation set* (Zitzler *et al*, 2003) could be achieved, the quantity of information contained within the set may overwhelm the DM, who must ultimately select a single solution.

The inherent difficulties in solving many-objective problems have lead EMO researchers to incorporate preference-based schemes into their algorithms, as comprehensively reviewed by Coello *et al* (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 reason *per se* why 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 *et al* (2002b) to stimulate research into many-objective optimisation. In the first known study of its kind, this test suite was used by Khare *et al* (2003), to investigate the scalability of some contemporary MOEAs. The results suggested that *PESA* (Corne *et al*, 2000) could produce approximation sets with good proximity but with poor distribution as problem size increased, whilst *SPEA2* (Zitzler *et al*, 2001) and *NSGA-II* (Deb *et al*, 2002a) produced the converse result. However, since single configuration instances of each algorithm were used, and each algorithm is itself a complicated structure of basic EMO components, it is difficult to identify the exact mechanisms and rationale that underpin the observed behaviour.

In this paper, abstracted algorithms derived from the NSGA-II class are used to understand the changing behaviour of isolated EMO mechanisms, and combinations of such mechanisms, as the number of objectives is varied. Results are generated for a map of configuration settings for each algorithm. This permits analysis to be made in terms of the *exploration-exploitation* trade-offs in EMO (Bosman and Thierens, 2003) and for performance *sweet-spots* to be identified (Goldberg, 1998).

1.2 Document roadmap

The remainder of the paper is organised as follows. In Section 2, the class of evolutionary multi-objective optimisers explored in the inquiry is introduced. The NSGA-II algorithm established by Deb *et al* (2002a) is described, together with three related algorithms based on decompositions of the NSGA-II selection mechanisms. The single-parent (mutation) and multi-parent (recombination) variation operators that are combined with selection to form the complete evolutionary method are also introduced.

Preliminary analysis of the class of optimisation algorithms described in Section 2 is undertaken in Section 3. Analysis is documented for selection-for-variation and selection-for-survival mechanisms, together with insights into the behaviour of the variation operators under special selection conditions. This analysis is helpful in attempting to explain and understand the observed empirical results of Sections 5, 6, and 7.

The components of the empirical inquiry are detailed in Section 4. The scalable optimisation task used in simulations is described and discussed, together with performance indicators for measuring the quality, in terms of proximity and spread, of the generated approximation sets. An experimental framework, based on components from studies by Laumanns *et al* (2001) and Purshouse and Fleming (2002b), is introduced and contrasted to the methodology adopted in a recent study into many-objective optimisation by Khare *et al* (2003).

Experiments in Section 5 and Section 6 are conducted for a fixed population size of 100 individuals. The number of objectives in the optimisation task is varied between 3 and 12. Section 5 details the results obtained for the optimisation algorithms when combined with the mutation operator, whilst Section 6 describes results for the algorithms in combination with the recombination operator. In each section, a synopsis of the fundamental results and analysis is given, followed by more detailed observations and discussions. Section 7 documents the results for both mutation and recombination when the size of the optimisation task is fixed at 6 objectives and the population size is allowed to vary between 10 and 1000 individuals.

Section 8 concludes the inquiry with a summary of important findings and suggestions for essential future work into evolutionary many-objective optimisation.

2 Optimisation algorithms

2.1 Introduction

The evolutionary search mechanism, often described in terms of *exploration* and *exploitation*, employed by the optimisers considered in this inquiry can be summarised by Equation 2.1.

$$P[t+1] = s_s\left(v\left[s_v\left(P[t]\right)\right], P[t]\right) \tag{2.1}$$

where $P[t]$ is the population at iteration t,

s_v is the selection-for-variation operator,

v is the variation operator, and

s_s is the selection-for-survival operator.

The *selection-for-variation* operator, usually known simply as the *selection* operator, selects candidate solutions from the current population to form the *mating pool*. *Variation* operators are applied to the solutions in the mating pool to create a set of new candidate solutions. These new solutions then compete with the population of current solutions in the *selection-for-survival* stage to determine the composition of the subsequent population. The selection-for-survival operator is also known as the *population management* or *reinsertion* operator.

The different multi-objective evolutionary optimisers proposed in the literature are generally categorised by the manner in which selection is performed. One particular broad category of selection scheme can be represented by the *elitist non-dominated sorting genetic algorithm* (NSGA-II) developed by Deb *et al* (2002a). In the inquiry, NSGA-II is considered at an abstracted, or component, level to improve the clarity with which underlying algorithm mechanisms can be identified as being responsible for observed performance. Four algorithms based on *NSGA-type* components are investigated. These deconstructed algorithms are described in Section 2.2.

Abstracted versions of the NSGA-II algorithm have been chosen to represent a class of optimisers that use the *dominance* concept as a primary, relative discriminator between candidate solutions in the population. Some algorithms within the class also use population *density estimates* as a secondary discriminator in cases where the primary comparison fails to distinguish relative worth. The preservation of previously discovered solutions within the EA population can also feature in this class.

The NSGA family has been specifically chosen because it is heavily studied in the literature and is conceptually very simple. NSGA-II is often regarded as *parameter-free*, and is used in such a format in the inquiry, although this is not strictly accurate. This is explained further in the algorithm discussions below. The family has strong similarities with other algorithms of its class, such as the *multi-objective genetic algorithm* (MOGA) family introduced by Fonseca and Fleming (1993), the *strength Pareto evolutionary algorithm* (SPEA) family first described by Zitzler and Thiele (1999), and others. It is strongly argued that *generalised* evidence collected for the abstracted NSGA algorithms will be widely applicable across other algorithms that use similar selection mechanisms.

The different selection schemes considered in the inquiry are described in Section 2.2, and are related back to the fundamental search mechanism given in Equation 2.1. The variation operators used in the study are introduced in Section 2.3.

2.2 Selection mechanisms

2.2.1 `emo1`: non-dominated sorting

The fundamental step in the construction of an EMO optimiser is to translate a vector performance measure (where each element represents a particular objective) into a unary indicator of absolute or relative worth. All methods in the studied class use the notion of *comparison* across a population of solutions to perform this mapping. The original method for EAs, *non-dominated sorting*, was suggested by Goldberg (1989). In this layer-based procedure, the locally non-dominated solutions in the population are identified, are assigned best rank, and are temporarily removed. The new locally non-dominated solutions in the remaining population are then identified, assigned next-best rank, and are removed. This process continues until the population is empty, by which time all solutions have received a rank (a series of equivalence classes has been defined). A computationally efficient method of performing the operation was implemented in NSGA-II (Deb *et al*, 2002a).

Non-dominated sorting is a common feature across the NSGA family and is used in all algorithms in the inquiry. Other dominance-based procedures have been suggested, such as Fonseca and Fleming's (1993) *Pareto-based ranking* and Zitzler and Thiele's (1999) *strength*-based approach. Pareto-based ranking assigns the rank of a solution to be the number of solutions in the population by which it is dominated. This is a more fine-grained approach than non-dominated sorting. The strength system uses the number of solutions that a particular solution dominates in addition to the Pareto-based rank to determine a fitness value for a solution. The granularity of this approach is greater still than Pareto-based ranking. There is little evidence available to argue that any one of these techniques is superior to any other under any optimisation conditions. This is because most EMO performance studies, with the notable exception of the inquiry by Laumanns *et al* (2001), consider *complete* algorithms rather than concentrate on specific components. Studies by Purshouse and Fleming (2002a) and Bosman and Thirens (2003) found no significant difference in performance between non-dominated sorting and Pareto-based ranking when solving tasks from the so-called *ZDT* test suite (Ziztler *et al*, 2000).

The selection-for-variation mechanism for `emo1` is a binary tournament. From two solutions selected at random from the population (with replacement), the solution that is accepted into the mating pool (*the winner*) is the one with the lowest non-dominated rank in the partial ordering defined by the above sorting procedure. In the case of a tie, selection is performed at random. In the selection-for-survival stage of `emo1`, the new population is composed entirely of the post-variation (child) solutions.

2.2.2 emo2: non-dominated sorting and crowding comparison

Some optimisers use population density estimates to discriminate between solutions of the same dominance-based equivalence class in order to promote diversity of solutions. Diversity is commonly required in objective-space, although sometimes the explicit promotion of diversity in decision-space is also appropriate. Deb *et al* (2002a) proposed a *crowding distance* estimator, which computes the average side length of the hypercube formed using nearest neighbour (NN) values in each objective as the vertices. This operator is described as parameter-less, but is actually a specific case of kth NN with $k=1$.

Boundary conditions for the estimator are handled slightly differently in the inquiry than to the original specification of Deb *et al* (2002a). All boundary distances are set to the maximum distance found rather than infinity to ensure that all estimates are equal for the equilibrium condition of a perfectly distributed (in the sense defined by the general form of the estimator) Pareto front.

Various density estimators have been proposed in the context of EMO, all with the aim of providing sufficient information to bias selection probabilities in favour of solutions from less-dense areas of objective-space. It is argued that the behaviour of emo2 *when compared to* emo1 will be generally indicative of the effect of introducing any type of explicit diversity-promoting mechanism. However, it should be noted that the relative effectiveness of the crowding distance estimator when compared to other estimators has been questioned by some researchers (Laumanns *et al*, 2001; Deb *et al*, 2003).

A two-step tournament process, formally defined by Deb *et al* (2002a) as the *crowded-comparison* operator, is used at the selection-for-variation stage of emo2. As for emo1, the winner is the solution with the lowest non-dominated rank. In the case of tied rank, the solution with the lowest density estimate is chosen. If the density estimates are also tied, the winner is selected at random.

This selection-for-variation mechanism is very similar in spirit to the fine-grained ranking procedure described by Purshouse and Fleming (2002b) and the modified strength measure in Zitzler *et al*'s (2001) SPEA2. These latter schemes also permit the use of proportional selection since they map the complete partial-ordering to actual fitness values.

At the selection-for-survival stage, the current population is replaced in its entirety by the child population.

2.2.3 emo3: elitist non-dominated sorting

Elitism refers to the policy of preserving previously-found good solutions in the population of the EA from one generation to the next. Selection-for-survival techniques of this nature can offer improved (faster) convergence to near-optimal regions and are also used to develop asymptotic convergence proofs for MOEAs (Rudolph, 2001; Laumanns *et al*, 2002).

In emo3 the so-called $(\mu+\lambda)$ strategy is implemented for selection-for-survival, in which the new EA population is taken as the best half of the elements from the combined parent (μ) and offspring (λ) populations. *Best* is defined according to the partial ordering obtained from a non-dominated sorting of the combined ($\mu+\lambda$) population. If the cut-off point for the new population occurs within an equivalence class then solutions are selected at random from this class.

Given settings for $|\mu|$ and $|\lambda|$, the ($\mu+\lambda$) method is sometimes then regarded as parameter-free, but the *elitism intensity* (a form of selective pressure, defined as the probability of selecting a member of the new population from μ) can in fact be varied to control the oft-cited *exploration – exploitation* (EE) trade-off (Laumanns *et al*, 2001). One particular method of

varying the elitism intensity in the context of NSGA-II is known as *controlled elitism* (Deb and Goel, 2001). In this scheme, the number of solutions that can be selected to occupy each dominance-based equivalence class in the new population is constrained according to some pre-specified plan. Controlled elitism has been shown to overcome some of the convergence problems encountered by the standard NSGA-II on multimodal task landscapes, in cases where variation operators have insufficient exploratory power. In this study, in order to limit the complexity of the forthcoming analysis, it is assumed that the EE trade-off can be adequately controlled with fixed elitism intensity. Also, multimodal task landscapes are not considered.

For `emo3`, the selection-for-variation method is identical to that described above for `emo1`.

2.2.4 `emo4`: elitist non-dominated sorting and crowding comparison – NSGA-II

This algorithm combines non-dominated sorting, secondary density-dependent selection-for-variation via the crowding distance estimator, and an enhanced form of $(\mu+\lambda)$ elitism to form the complete NSGA-II algorithm described by Deb *et al* (2002a). The enhanced $(\mu+\lambda)$ selection-for-survival mechanism discriminates based on new density estimates from the combined $(\mu+\lambda)$ population in cases where the cut-off point for the new population lies within an equivalence class obtained from non-dominated sorting. If the cut-off point still resides within the finer hierarchy, then selection-for-survival is made at random from this sub-set of solutions.

2.3 Variation operators

2.3.1 Introduction to variation

Variation refers to the process of making new solutions from current solutions. When combined with selection, it forms a compelling Darwinian search mechanism that is at the very heart of the EA heuristic.

In general, within a single variation operation, m current solutions (known as *parents*) are used to produce n new solutions (known as *children*). $m = n$ for most variation operators, including those considered in the inquiry. An operator is usually classified based on whether it is *single-parent* ($m = 1$) or *multi-parent* ($m > 1$).[2] Single-parent variations are also known as *mutations*, whilst multi-parent variations are called *recombination* or *crossover* operators. Other distinctions between the natures of recombination and mutation can be made, being often representation-dependent. A discussion of these is beyond the scope of this inquiry.

In the context of the multi-objective performance of a child compared to its parents, a variation will have one of the following mutually exclusive outcomes:

- The child dominates at least one of its parents.
- The child is non-dominated with respect to all of its parents.
- The child is equal to at least one of its parents and does not dominate any of its parents.
- The child is dominated by at least one of its parents and is non-dominated with respect to all of its other parents.

---

[2] The special case of $m = 0$ could be regarded as an operator that creates a new solution from a distribution rather than a set of exemplars. This is the common method for initialising the EA population.

In a single-parent setting, from the perspective of evolutionary progress, these outcomes can be interpreted as success, incomparability, equality, and failure respectively. The multi-parent situation is more complex (there are 4^m distinct outcomes in terms of comparison across all m parents) and is clouded further by the possible relationships between parents. In the two-parent case, the above classification of outcomes is sufficient but cannot be strictly interpreted in the same way as the one-parent situation.

2.3.2 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 paper (Deb *et al*, 2002b; Khare *et al*, 2003). Variable-wise mutation is performed according to a probability distribution function centred over the parent value. The operator is defined by Equations 2.2 and 2.3.

$$c_i = \begin{cases} p_i + (p_i - l_i)\delta_i & \text{if } r_i < 0.5 \\ p_i + (u_i - p_i)\delta_i & \text{otherwise} \end{cases} \tag{2.2}$$

$$\delta_i = \begin{cases} (2r_i)^{1/(\eta_m+1)} - 1 & \text{if } r_i < 0.5 \\ 1 - \left[2(1-r_i)\right]^{1/(\eta_m+1)} & \text{otherwise} \end{cases} \tag{2.3}$$

where p_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.

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 $O(1/\eta_m)$. 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 2.4.

$$p(\text{mutate}) = 1 - (1 - p_m)^n \tag{2.4}$$

2.3.3 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 *et al*, 2002b; Khare *et al*, 2003), and is defined in Equations 2.5, 2.6, and 2.7.

$$c_{1,i} = 0.5\left[(1+\beta_i)p_{1,i} + (1-\beta_i)p_{2,i}\right] \tag{2.5}$$

$$c_{2,i} = 0.5\left[(1-\beta_i)p_{1,i} + (1+\beta_i)p_{2,i}\right] \tag{2.6}$$

$$\beta_i = \begin{cases} (2r_i)^{1/(\eta_c+1)} & \text{if } r_i < 0.5 \\ \left[1/(1-r_i)\right]^{1/(\eta_c+1)} & \text{otherwise} \end{cases} \tag{2.7}$$

where $p_{1,i}$ and $p_{2,i}$ 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.

SBX generates child values from a probability distribution, with standard deviation derived from the distance between parent values and a distribution parameter η_c. Distance is the primary factor in determining the magnitude of the distribution, with larger distances leading to larger standard deviations. 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. Variation may be *introverted* (the child values are within the range defined by the parents) or *extroverted* (the child values are outside the range defined by the parents). Child values 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 recombination scheme is chosen, in which each individual decision variable is selected for variation independently of any other (given the fact 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 2.8.

$$p(\text{recombine}) = p_c\left[1-(1-p_{ic})^n\right] \tag{2.8}$$

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 2.4.

3 Preliminary analysis

3.1 Introduction

Preliminary analysis of the selection and variation mechanisms described in Equation 2.1 will prove helpful when attempting to explain the forthcoming experimental results of Sections 5, 6, and 7. In Section 3.2.1, the conditions when various aspects of the selection-for-variation (s_v) mechanisms of the emo algorithms are likely to be active (or deactive) are described, and the expected effect of these mechanisms on overall search performance is discussed. Similar analysis is undertaken in Section 3.2.2 for the selection-for-survival mechanism (s_s). Section 3.3 studies the behaviour of a population under the variation operators (v) in the absence of selective bias. This analysis provides useful information because some settings for variation configurations can substantially reduce the influence of selection mechanisms over the population evolution.

3.2 Selection mechanisms

3.2.1 Selection-for-variation

The binary tournament selection process in the emo algorithms makes comparisons of the following forms between solutions:

A. primary dominance-based comparison (all emo algorithms),
B. secondary density-based comparison (emo2 and emo4).

Considering stage A alone, if the dominance-based equivalence classes are small in relation to the population size then s_s will direct the application of variation operators towards high quality (from the perspective of the current population) regions of the search space. Actual progress then depends on the ability of v, given this genetic material, to find superior solutions. If the equivalence classes are large in relation to the population size (generally meaning that a single equivalence class has a large membership in the population, which in practice implies that most of the solutions are locally non-dominated) then s_s will tend to select solutions on a more uniform basis from $P[t]$. This does not necessarily mean that search progress will be attenuated, since this still depends on the ability of v to find improved solutions under these conditions. The rate of improvement (and hence convergence towards the globally optimal surface) depends on (i) the probability of improvement and (ii) the expected magnitude of improvement. Note that, in the absence of diversity-promoting mechanisms, random sampling errors are likely to progressively focus the population on to a small region of objective-space. This phenomenon is known as *genetic drift*.

EMO algorithms also seek to obtain a good distribution of solutions to present to the decision-maker. Stage B of the selection-for-variation operator is one method designed to take account of this aim. Stage B only has a real impact on the results of s_s when many population members share the same dominance-based equivalence class. Assuming that the locally non-dominated solutions comprise the largest class, then the probability of stage B being the key arbiter in a selection decision is approximately equal to the proportion of non-dominated solutions in the population. When B is active, s_s is biased in favour of solutions with low density estimates. For the crowding distance estimator illustrated in Figure 3.1, noting that density is the inverse of distance, these solutions are likely to be

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

The volume of objective-space increases exponentially as the number of conflicting objectives is increased linearly. For a finite population in these higher-dimensional spaces there is 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. Thus, s_s can bias in favour of solutions that are locally non-dominated but are actually poor in relation to others from the perspective of the off-line proximity indicator. This could ultimately affect the ability of an MOEA to discover an approximation set that is of value to the decision-maker, since the set may be far from Pareto-optimal and lacking in compromise solutions.

If the variation operators have difficulties in producing children that dominate badly performing parents then MOEA performance can be adversely affected. This problem of *dominance resistant solutions* was first identified by Ikeda *et al* (2001), in the context of real number representations and particular real parameter variation operators, for a specific class of multi-objective functions. Deb *et al* (2002b) have also encountered this behaviour for the scalable, real parameter, *constraint surface* tasks. The authors also identified that, in this

instance, the problem of dominance resistance increased with the dimension of objective-space.

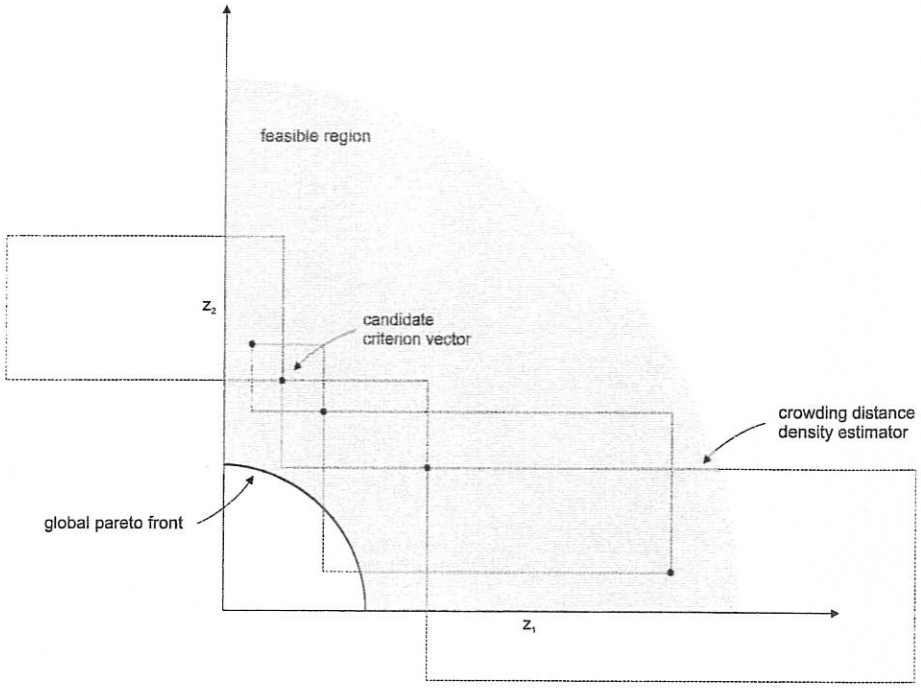

Figure 3.1: Crowding distance density estimator

3.2.2 Selection-for-survival

In the non-elitist algorithms (`emo1` and `emo2`) the children produced by the variation operators are certain to be included in the new population. Members of the current population are entirely supplanted by these new solutions. Thus, the only way for the genetic material of any member of the current population to survive into the next iteration is via selection-for-variation. In the elitist algorithms (`emo3` and `emo4`) by contrast, members of the current population have a non-zero probability of being included at the next iteration.

In the elitist schemes, reinsertion into the new population is based on a deterministic hierarchy formed from comparisons between candidate solutions:

A. primary dominance-based hierarchy (`emo3`),
B. secondary density-based hierarchy (`emo4`).

Considering stage A alone, this approach will conserve all locally non-dominated solutions discovered during the search until the number of such solutions exceeds the population size. In this latter case, non-dominated solutions will be rejected on a random basis. It is thus conceivable that this approach can suffer from oscillatory behaviour, in which previously rejected solutions are rediscovered, due to the absence of an asymptotic convergence property. This issue has been identified and addressed by Everson *et al* (2002), Knowles and Corne (2003b), and Laumanns *et al* (2002).

Elitism is known to increase the convergence rate of evolutionary algorithms, with a corresponding increased risk of premature convergence in multimodal cost environments if mutation search power is insufficient. Increased convergence can also quickly limit the variability within the genetic material available to the search operators. For multi-parent operators that make use of the differences between parents to adapt the search magnitude, this can cause EA search stagnation.

With the implementation of stage A alone, the algorithm would be predicted to suffer from genetic drift. Since EMO algorithms aim for good diversity in optimal areas of objective-space, stage B was proposed by Deb *et al* (2002a) to bias selection-for-survival towards solutions with low density estimates. Stage B will only be properly operational when many population members share the same dominance-based equivalence class in the hierarchy. In the context of the multi-objective search, this usually corresponds to a large number of locally non-dominated solutions in the combined P[t] and child solution pools. Stage B will preserve solutions with low density estimates. As demonstrated previously for selection-for-variation, these will be boundary solutions, other remote solutions, and neighbours of remote solutions. As the dimension of objective-space increases, the opportunity for a solution to be remote and non-dominated, and yet poor in terms of relative location to the optimal hypersurface increases. s_s inclusive of stage B can also suffer from oscillatory behaviour, as described earlier for stage A.

Note that stage A of s_s does not have any in-built copy rejection mechanism. Thus, it is possible for a particular solution to be present in the post-s_s population several times. However, since the crowded distance density estimator is an example of first-nearest-neighbour, solutions with copies will have the maximum possible density estimate (corresponding to a crowding distance of zero). Non-dominated solutions of this nature will be biased against in stage B when the number of non-dominated solutions in the combined pool exceeds the population size. Indeed, if the number of non-dominated solutions without copies exceeds or equals the population size then the solutions with copies are certain not to be preserved. If there is sufficient space in the new population for non-dominated solutions with zero crowding distance then these will be subject to purely random inclusion.

3.3 Variation mechanisms

Variation operator analysis is a challenging task for anything but the most simple of test problems. This is because of the generally complicated relationship between decision-space and objective-space (complicated further in the presence of many objectives), and the population-based nature of the search. However, one key variation operator result that affects s_v can immediately be suggested: if the variation operator has the conceptual maximum exploratory power, meaning that there is an equal probability of generating a child at each point in the search space (regardless of parent location), then the effect of s_v is completely eliminated. The best solutions may be chosen but none of their genetic information is passed on to their children. The total variation is equivalent to simply generating the child solution at random from a uniform distribution (equivalent to the classical method of initialising a population). Hence, the behaviour of the population over the optimisation process would be neutral (subject to random noise) in the absence of biased selection-for-survival mechanisms (elitism).

The behaviour of the population under the mutation and recombination operators described in Section 2.3, with settings chosen to provide high exploratory power, is studied in the remainder of this section.

3.3.1 Polynomial mutation

For the mutation operator configuration of [p_m = 1.0, η_m = 0.0] every element of every candidate solution in the selection pool is certain to be mutated, with a mutation distribution that is piece-wise uniform about the parent value. This is the maximum possible expected perturbation for this operator. The distribution is piece-wise uniform rather than globally uniform because there exists an equal probability of mutating in either direction along the feasible range of the decision variable regardless of parent position. By contrast, for a properly uniform distribution, the probability of movement in each direction *would* be dependent on the parent position. However, once a direction has been chosen, the distribution in that sub-region *is* locally uniform for $\eta_m = 0$.

The special piece-wise uniformity of the polynomial mutation operator changes the distribution of the population in the absence of any selective bias. Since the direction of mutation is always decided with probability 0.5, this internal bias causes a uniform random sample of decision values to be perturbed towards the boundaries of the feasible decision-space with greater probability than that resulting from a globally uniform distribution. This is simulated in Figure 3.2 for a single decision variable with feasible range [0 1]. A population of 100,000 is used (the large population reduces the noise on the histogram, but the results are equally applicable to standard EMO population sizes). The movement of the population distribution translates into movement through the cost landscape that is independent of fitness-based selection.

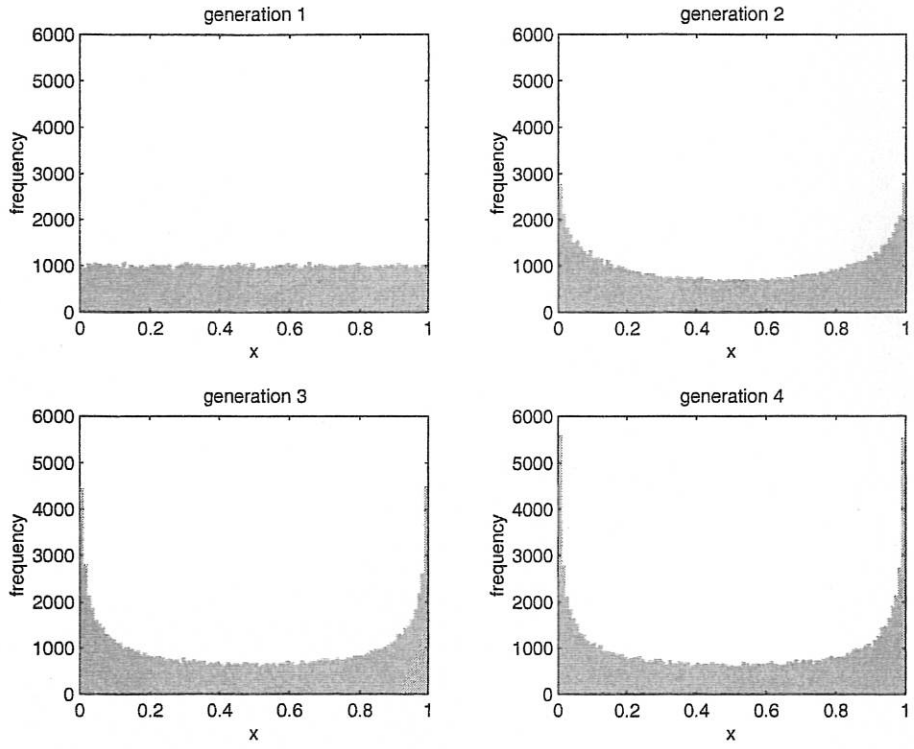

Figure 3.2: Population distribution change via polynomial mutation [$p_m = 0.5$, $\eta_m = 2.5$] for $x \in [0\ 1]$ in the absence of fitness-biased selection.

3.3.2 SBX recombination

It is instructive to consider the effects of repeated application of the SBX variation operator to a population in the absence of fitness-based selection schemes. If parents are selected randomly from the population, and the variance on the SBX distribution is large (corresponding to low values of η_c), then the population distribution evolves as shown for a single decision variable, $x \in [0\ 1]$, in Figure 3.3. A large distributional bias to the decision-variable boundary values is established. The observed distribution will ultimately be generated for any initial population distribution, except for the case where the entire population is generated at a single point.

The observed bias towards remoter values occurs because SBX is likely to generate children outside of the feasible bounds of the decision variable, requiring the child values to be cropped to the nearest feasible equivalent. If the SBX variance is low, then this behaviour is not observed. Instead, no long-term adjustments are seen to the initial population distribution.

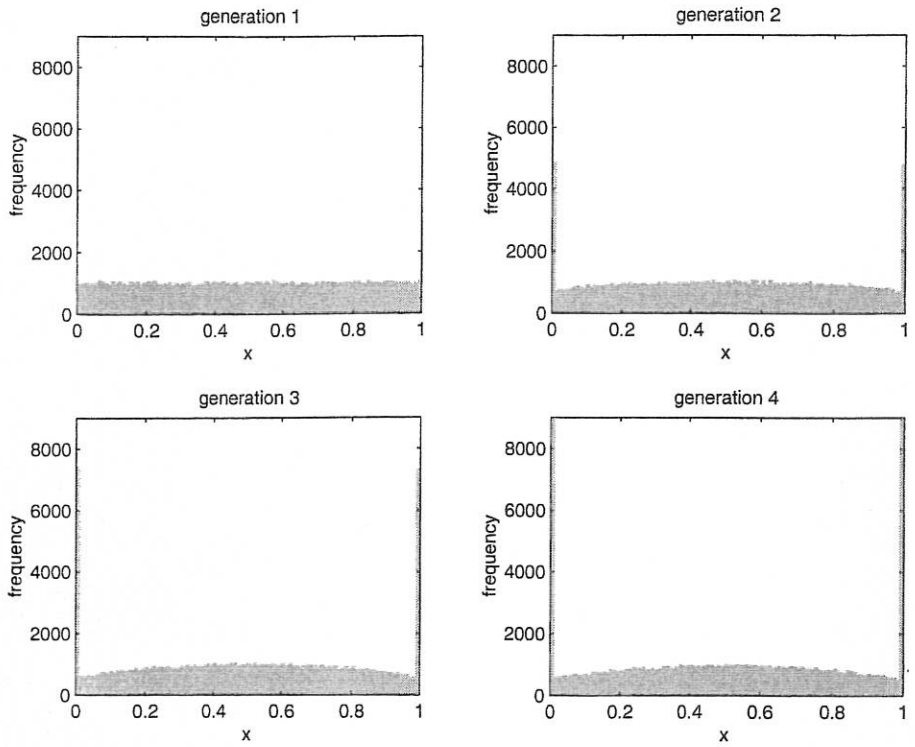

Figure 3.3: Population distribution change via SBX [$p_{ic} = 0.5$, $\eta_c = 0.25$] for $x \in [0\ 1]$ in the absence of fitness-based selection.

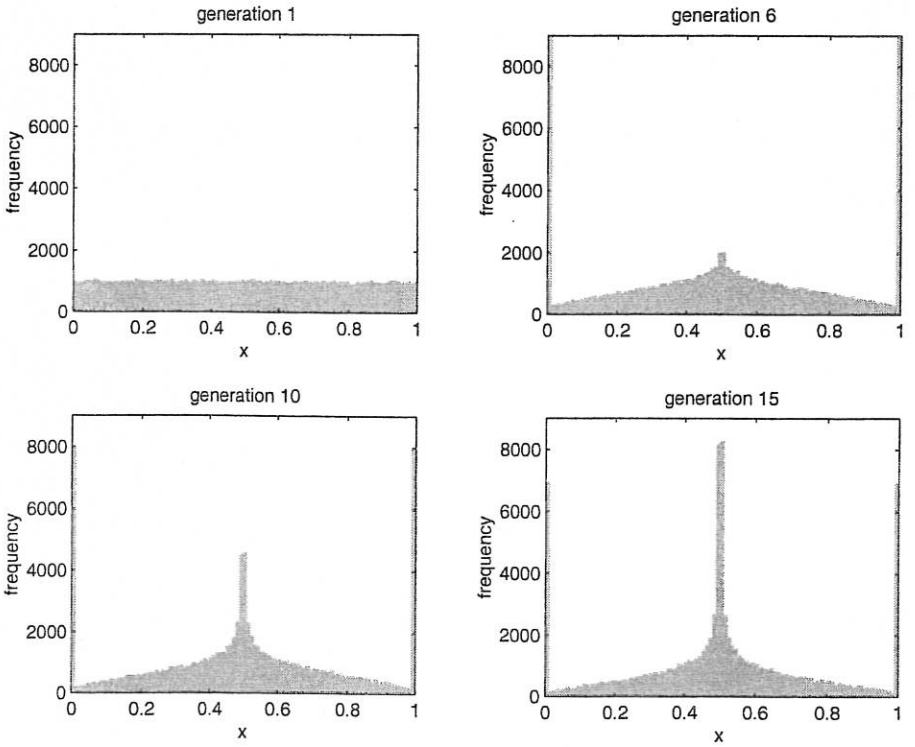

Figure 3.4: Population distribution change via SBX [$p_{ic} = 1.0$, $\eta_c = 0.0$] for $x \in [0\ 1]$ under repetitive selection.

If SBX is repeatedly applied to the same pair of evolving solutions in the population, then the long-term behaviour is to converge to the mid-point between the original parent values. This requires successive introspective variations or (and possibly combined with) a single large introspection. The self-adaptive nature of the search magnitude then makes it increasingly less likely that distant children will be produced. The symmetric form of the SBX distribution ensures that convergence will be to the mid-point.

If the scale of the distribution (relative to the decision variable range) is likely to cause cropping of infeasible child values to the feasible bounds then the symmetry property is lost. For a decision variable with feasible range $[l\ u]$, if the search power is strong enough to cause one child value to be set to u and the other child to be set to l, then the new centre becomes $(u + l)/2$ and convergence is ultimately achieved on this value. The first few applications of SBX to a population under these conditions, with $l = 0$ and $u = 1$, are shown in Figure 3.4 for $[p_{ic} = 1.0, \eta_c = 0.0]$.

4 Inquiry design

4.1 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 4.2. The performance indicators chosen to reduce an approximation set to a more tractable number of summary statistics are presented in Section 4.3. The experimental framework that has been developed to facilitate understanding of many-objective optimiser behaviour is introduced in Section 4.4, and contrasted to a recent study in this research field. In Section 4.5, an introduction is made to the empirical results detailed in the remainder of the paper.

4.2 Scalable optimisation task

This inquiry considers a real-parameter function optimisation task known as *DTLZ2*, which is defined in Equation 4.1. The task is taken from a highly tractable set of problems developed by Deb *et al* (2002b) 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.

$$
\begin{aligned}
&\text{minimise} \quad z_1(\mathbf{x}) = \left[1 + g(\mathbf{x}_M)\right]\cos(x_1\pi/2)\cos(x_2\pi/2)\ldots\cos(x_{M-2}\pi/2)\cos(x_{M-1}\pi/2),\\
&\text{minimise} \quad z_2(\mathbf{x}) = \left[1 + g(\mathbf{x}_M)\right]\cos(x_1\pi/2)\cos(x_2\pi/2)\ldots\cos(x_{M-2}\pi/2)\sin(x_{M-1}\pi/2),\\
&\text{minimise} \quad z_3(\mathbf{x}) = \left[1 + g(\mathbf{x}_M)\right]\cos(x_1\pi/2)\cos(x_2\pi/2)\ldots\sin(x_{M-2}\pi/2),\\
&\vdots \qquad\qquad \vdots\\
&\text{minimise} \quad z_{M-1}(\mathbf{x}) = \left[1 + g(\mathbf{x}_M)\right]\cos(x_1\pi/2)\sin(x_2\pi/2),\\
&\text{minimise} \quad z_M(\mathbf{x}) = \left[1 + g(\mathbf{x}_M)\right]\sin(x_1\pi/2),\\
&\text{where} \quad g(\mathbf{x}_M) = \sum_{x_i \in \mathbf{x}_M}(x_i - 0.5)^2, \text{ with } \mathbf{x}_M = [x_M, \ldots, x_n],\\
&\text{and} \quad 0 \le x_i \le 1, \text{ for } i = 1,2,\ldots,n, \text{ with } n = M + k - 1.
\end{aligned}
\tag{4.1}
$$

M is the number of objectives, n=M+k-1 is the number of decision variables, and k is a difficulty parameter (k = 10 in this study). DTLZ2 is comprised of decision variables of two

distinct functional types: those that control convergence towards 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 k-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 to the global Pareto front of DTLZ2 using an EMO optimiser (Deb *et al*, 2002b). 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 and 6, 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. The effect of more complex real-parameter cost features remains a matter for future research. There is also a clear need to investigate other optimisation tasks of a different underlying theme, such as the multi-objective quadratic assignment problem class recently developed by Knowles and Corne (2003a).

The proportion of a population that is locally non-dominated is thought to be one of the key factors in determining EMO algorithm success (Deb, 2001). Empirical values, averaged across 1000 random samples sets from DTLZ2, for various population sizes and numbers of conflicting objectives are shown in Figure 4.1.

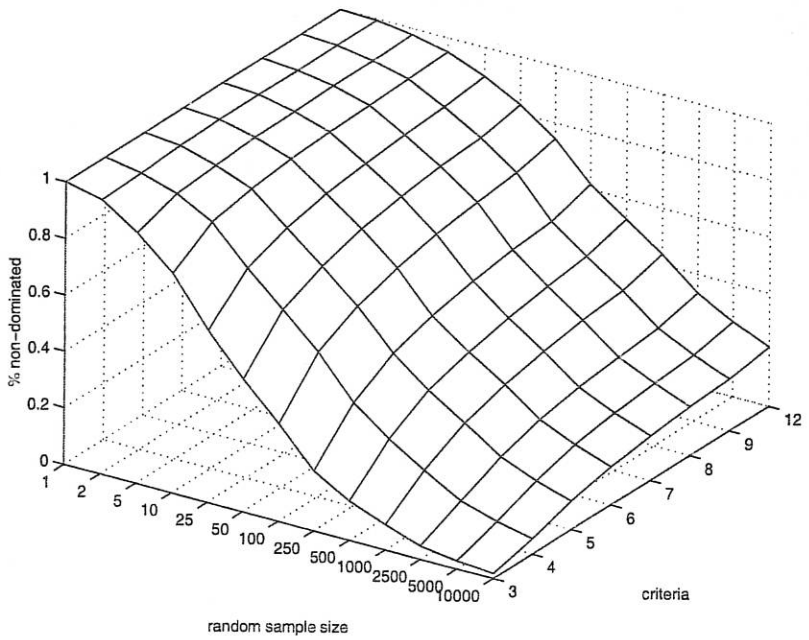

Figure 4.1: Estimated non-dominated proportions of random samples of DTLZ2(M) solutions.

4.3 Performance indicators

In the context of this inquiry, *performance* relates to the quality of the trade-off surface discovered by an optimiser, given a finite number of candidate solution evaluations. 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, and that are as near uniformly distributed as the true surface permits.

Various performance indicators have been proposed to measure these aspects of quality (Deb, 2001). The indicators used in the inquiry are described in this section. 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). Zitzler *et al* (2003) have shown that no finite combination of unary measures can indicate whether one approximation set is superior to the other (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.

4.3.1 Proximity indicator

The proximity indicator measures a median level of proximity of the approximation set, $\mathbf{Z}_A$, to the global surface. In terms of attainment across the objectives, a objective vector for DTLZ2 will respect Equation 4.2. 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 4.3. This is essentially the same as Veldhuizen's (1999) *generational distance* metric, for the case of a continuous globally optimal reference set, $\mathbf{Z}_*$.

$$\left[\sum_{i=1}^{M}(z_i)^2\right]^{1/2} \geq 1 \tag{4.2}$$

$$I_P = \text{median}_{\mathbf{z}_A \in \mathbf{Z}_A}\left\{\left[\sum_{i=1}^{M}\left(\mathbf{z}_{A,i}\right)^2\right]^{1/2} - 1\right\} \tag{4.3}$$

4.3.2 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. *Good* means that the identified non-dominated vectors should span the complete surface, with appropriate distances between each. Given limited preference information and a continuous surface, the vectors should be equal distances apart.

To achieve high quantization 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 *et al*'s (2002a) Δ metric. Unfortunately, it can become unclear which aspect of the distribution is responsible for the observed indicator value. For example, errors on the spread of the distribution can potentially mask difficulties with uniformity. To manage the complexity of the inquiry, only the spread 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 4.4.

$$D = \left[\sum_{m=1}^{M} \left(\max_{i=1}^{|\mathbf{z}_A|} \mathbf{z}_{A,i,m} - \min_{i=1}^{|\mathbf{z}_A|} \mathbf{z}_{A,i,m} \right)^2 \right]^{1/2} \tag{4.4}$$

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 4.5. 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.

$$I_s = D \Bigg/ \left[\sum_{m=1}^{M} \left(\max_{i=1}^{|\mathbf{z}_*|} \mathbf{z}_{*,i,m} - \min_{i=1}^{|\mathbf{z}_*|} \mathbf{z}_{*,i,m} \right)^2 \right]^{1/2} \tag{4.5}$$

4.4 Investigative framework

Very little of the documented EMO research has considered the issue of scalability. Indeed, with the advent of a scalable test suite, results of the first such study have only recently been published (Khare *et al*, 2003). A description and critique of the methodology used in that work is described below. This is followed by a reasoned account of the framework pursued in this inquiry.

4.4.1 Published framework: Khare *et al* (2003)

Khare *et al* (2003) investigated the scalability of three well-known *functionally complete* EA optimisers: Corne *et al*'s (2000) *Pareto-envelope based selection algorithm* (PESA), Zitzler *et al*'s (2001) SPEA2, and a controlled elitism version of emo4 (Deb and Goel, 2001). The 2, 3, 4, 6, and 8-objective instances of tasks DTLZ2, DTLZ3, DTLZ4, and DTLZ6 were undertaken. These extra tests feature multimodality and reduced dimensionality behaviour[3]. Population size was increased with the number of objectives.

The main conclusions of the study were that the algorithms of the class considered in this inquiry – SPEA2 and NSGA-II – increasingly encountered difficulties converging to the global surface as the number of objectives was enlarged. However, the distribution of the identified front remained satisfactory. The hypergrid-based PESA, which is arguably not directly linked to the emo class of algorithms, displayed the opposite behaviour.

Khare *et al* (2003) have shown that conclusions drawn from tasks with small numbers of objectives cannot be generalised to higher numbers of objectives. This is an important result. However, because complete algorithms were considered in the study, the specific mechanisms and reasons behind the observed behaviours remains unclear. An approach based on abstracted algorithms, with consideration of the dynamic processes under which they operate, may prove more appropriate in this context and is described below.

---

[3] Deb *et al* (2002b) claim that for the *curve problem* (of which DTLZ6 is an example) only one independent variable describes the global front, regardless of the number of objectives considered. This behaviour is obtained by creating a dependency between the convergence-variables and the distribution-variables. This invalidates the assumption of independence upon which the equation describing the global front can be obtained, and this result must thus be treated with care.

4.4.2 Proposed framework

The changing behaviour of evolutionary multi-objective optimisers in environments of increasing conflict remains largely unclear. Multi-objective operations, such as dominance checks and density estimates, are exponential in the number of objectives and are thus time-consuming. Advanced techniques are required for visualising relationships between more than three objectives. Increased dimensionality, and the corresponding reduction in tractability, poses a significant challenge to the analyst. Hence, the selected framework is aimed more towards *exploratory data analysis* than rigorous statistical methodology. With this in mind, each run of each algorithm configuration is executed once only. Whilst repeated runs are still preferable, this is infeasible within the context of this inquiry because of the vast processing and data storage requirements. The aim is to expose broad relationships within and between algorithms as scale demands vary. There will be no attempt to draw statistically significant performance comparisons between one optimiser and another.

It is a well-known fact that the configuration of the variation operators can have a substantial effect on the performance of evolutionary algorithms. Thus, it is unsurprising that different variation operator configurations can lead to very different performance for the emo class of algorithms, and that behaviour for a single configuration can vary considerably with the number of conflicting objectives. Rather than attempting to tune the variation operators to obtain best performance, a broad set of configurations are considered in the inquiry with the aim of uncovering regional behaviour for a range of exploration-exploitation choices.

Variation operators have two key controllable parameters: the *probability* of applying variation and the expected *magnitude* of such a variation, as described in Section 2.3. The inquiry uses a sample set of probabilities and magnitudes, with pseudo-logarithmic scales that have been heuristically chosen to show relativity within and between different maps. Algorithm runs are conducted for all pair-wise permutations of magnitude and probability samples, with all other parameters held constant, to yield a set of optimiser *responses*. Each response is the population evolution of an EA over 1000 generations. Summary statistics for performance indicators and other observable variables can be calculated and displayed in a *response map*. This is a two-dimensional grid with operator magnitude and probability as horizontal and vertical axes respectively. Note that the spatial relationships that arise through the use of multiple configuration instances offer improved confidence in the one-replication results.

The response map provides the basis for developing an understanding of optimiser performance. Laumanns *et al* (2001) took this approach and also fitted a *response surface* (Myers and Montgomery, 2002) to the raw samples to obtain a good description of the data. For the purposes of this inquiry, the raw data itself proves sufficient to expose the underlying spatial relationships.

As described in Section 2.3, two types of variation operator are considered for all algorithms in the study. The first is a one-parent mutation with controllable parameters p_m and η_m, whilst the second is a two-parent recombination with parameters p_{ic} and η_c. Thus maps are generated for each algorithm, for each variation type, and for each objective-value instance of DTLZ2. Population size is held constant (at 100) in these studies. Deb (2001) argues that a main method for coping with increased conflict is to increase the population size. However, this may not be practical in the context of many real-world applications (where evaluation of a single candidate solution can be a time-consuming process). Furthermore, Deb's results indicate that the benefit of increased population sizes decreases rapidly with the number of conflicting objectives to be considered. However, a population scaling study is undertaken as part of the inquiry for a fixed 6-objective instance of DTLZ2.

4.5 Introduction to Results

Experiments were conducted to generate response maps for the emo1, emo2, emo3, and emo4 algorithms on the DTLZ2 optimisation task. Maps were generated for the 3, 4, 5, 6, 7, 9, and 12-objective instances of the function. In the analysis of the mutation and recombination operators that follows, only the results for the 3, 6, and 12 objective instances are presented. These prove sufficient for demonstrating behaviours and identifiable potential causes both within and between algorithm configurations. The next two sub-sections explain what types of results are considered and how the results are presented. In Section 5 results and analysis for the mutation operator are discussed, followed by those for recombination (both with and without exchange) in Section 6. A population-sizing study in Section 7 completes the inquiry.

4.5.1 Types of response map

Performance measures

Proximity is measured by subtracting the initial proximity indicator value, computed as described in Section 4.3.1, from the median of the values obtained from the final 100 iterations of the optimisation process. Thus, a value of zero indicates no progress from the initial population, a negative value indicates convergence towards the global trade-off surface, and a positive value indicates divergence away from the true surface. Increasing magnitudes indicate increasing levels of convergence or divergence. Improved proximity indicator values of at least 0.5 are identified by highlighted map squares.

Spread is measured as the median value of the spread indicator, described in Section 4.3.2, 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. A spread measure in the range [0.75 1.25] is identified by a highlighted grid square.

Other measurements

Optimiser behaviour, in terms of the above performance indicators, can be explained by observing other measurable process variables. The proportion of the population that is locally non-dominated, the proportion of children that dominate their parent(s), and the Euclidean distance in objective-space between parent and child are all considered in the analysis. In all these cases, the observed value was taken as the median across the first ten iterations of the optimiser. Early measurements were taken because, whilst these process variables may be partially responsible for the observed behaviour, they are also affected by the behaviour due to the iterative, feedback-like, nature of the optimisation process. Early measurements limit the impact of feedback on the results.

4.5.2 Viewing the response maps

Sections 5, 6, and 7 include many figures that show algorithm response maps. A set of example response maps for the proximity indicator, I_P, are shown in Figure 4.2. The maps in the first row of the figure relate to the non-elitist algorithms (emo1 and emo2) whereas the maps in the second row relate to the elitist algorithms (emo3 and emo4). Column-wise, functionality is distinguished by the inclusion, or otherwise, of the crowding diversity-promotion mechanism. In the first column, the responses relate to algorithms without such a mechanism (emo1 and emo3), whereas the second column corresponds to diversity-promoting algorithms (emo2 and emo4).

Performance for each probability-magnitude setting is indicated by a grey-scale square at the appropriate location. Lighter shades of grey indicate better proximity, as shown by the colour-

bars of indicator values to the right of the maps. Thus, for the example emo4 shown, a region of good proximity is evident for η_c values in the range [0 5] together with variation probabilities in the order of [0.01 0.1]. Conversely, algorithm divergence is exhibited for intermediate η_c, around [1 10], and high p_{ic} [0.5 1.0].

Some grid squares in the response maps are highlighted by means of a solid boundary: [p_m = 0.005, η_m = 5] being one such example in Figure 4.2a. Variation operator configurations identified in this way are those that lie within a certain threshold for whichever measure of the optimisation process is considered for that map. In Figure 4.2, the highlighted grid points correspond to configurations that attain a proximity measure of -0.5 or under.

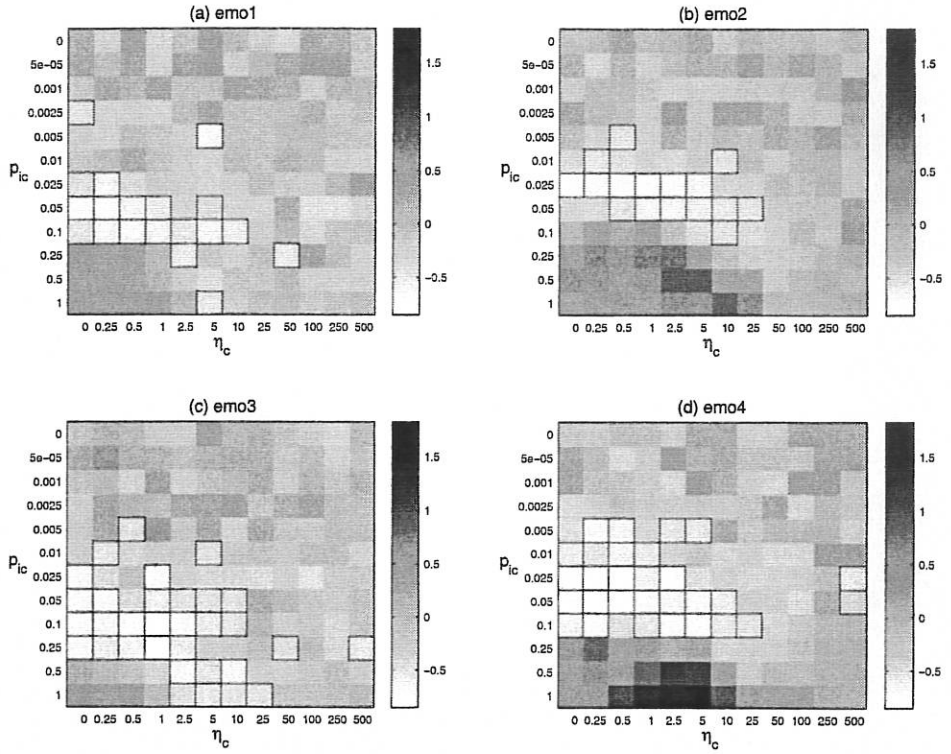

Figure 4.2: Example emo response maps

5 Results and Analysis: Mutation

5.1 Mutation: Synopsis

The fundamental results to be drawn from the mutation inquiry are summarised in this section. For the interested reader, more details and explanations are provided in the observation and discussion sections that follow.

- A suitable trade-off between exploration and exploitation is required for good many-objective performance, and this can be specified via mutation operator configuration settings.

- The mutation operator has a large proximity sweet-spot but good spread values are more difficult to obtain. The sweet-spots contract as the number of conflicting objectives to be optimised increases.

- For large numbers of objectives, it is difficult to obtain an approximation set with both good proximity and good spread since the combined sweet-spot is very small. Classical settings in the literature appear suitable when combined with NSGA-II (emo4) selection mechanisms.

- Whilst NSGA-II can provide the best performance, inappropriate variation operator settings can lead to poor performance when large numbers of objectives are considered. The abstracted algorithms offer more robust performance in these situations, but the best results for these do not match the best results for NSGA-II.

- If convergence towards the true front is difficult to achieve because of low mutation success rates, then *active* diversity-promoting mechanisms can lead to an approximation set with many poor-proximity solutions. The level of activity is indirectly controllable via mutation operator configuration settings.

5.2 Observations[4]

5.2.1 6 objectives

The proximity maps obtained for the six-objective instance of DTLZ2 are shown in Figure 5.1. A band of good proximity is seen to stretch from configurations of low mutation probabilities coupled with large expected mutation perturbations, such as [$p_m = 0.005$, $\eta_m = 0.25$], to configurations of high probabilities coupled with small perturbations, for example [$p_m = 0.5$, $\eta_m = 250$]. This region is identifiable across all four algorithms but is noticeably thinner for the crowding algorithms emo2 and emo4. The elitist algorithms seem to offer better proximity in areas of intermediate p_m and high $1/\eta_m$, where the band thickens in Figures 5.1c and 5.1d.

Poor proximity is evident across the four algorithms for regions of high mutation probability combined with large mutation perturbation. Divergence appears particularly strong for the crowding algorithms as shown in Figures 5.1b and 5.1d. Proximity is also quite poor for very low mutation probabilities and very small perturbations, especially when these two settings occur together.

DTLZ2(6) response maps for the spread indicator are given in Figure 5.2. Good spread is found to lie in a band close to the lower-left boundary of the good proximity region identified in Figure 5.1. Spread is too small in regions when p_m is low and η_m is high. This means that the optimiser has produced approximation sets that reside in a small sub-region of objective-space. As p_m and $1/\eta_m$ are increased, spread values increase in magnitude until they become larger than optimal. Spread values become very large for [$p_m = 1.0$, $\eta_m = 0.0$]. In regions such as this, the members of the approximation set are spread widely throughout objective-space. Many of these locations will be far from the true trade-off surface.

The non-crowding algorithms, emo1 and emo3, have very similar spread response maps, as shown in Figures 5.2a and 5.2c. The main difference between the non-crowding maps and the crowding versions of Figures 5.2b and 5.2d is that spread values are generally larger on the latter maps, except in the region with exemplar [$p_m = 0.0025$, $\eta_m = 250$]. Spread values become particularly large in the lower-left map area for the elitist, crowding emo4. Note that the classical choice of variation operator settings for this problem, namely [$p_m = |\mathbf{x}|^{-1} = 0.07$ to 1 s.f., $\eta_m = 20$] (Deb *et al*, 2002b), is located in a region that appears to produce

[4] For details on the types of response map presented in this section, and for details of how to view the maps, refer to Section 4.5.

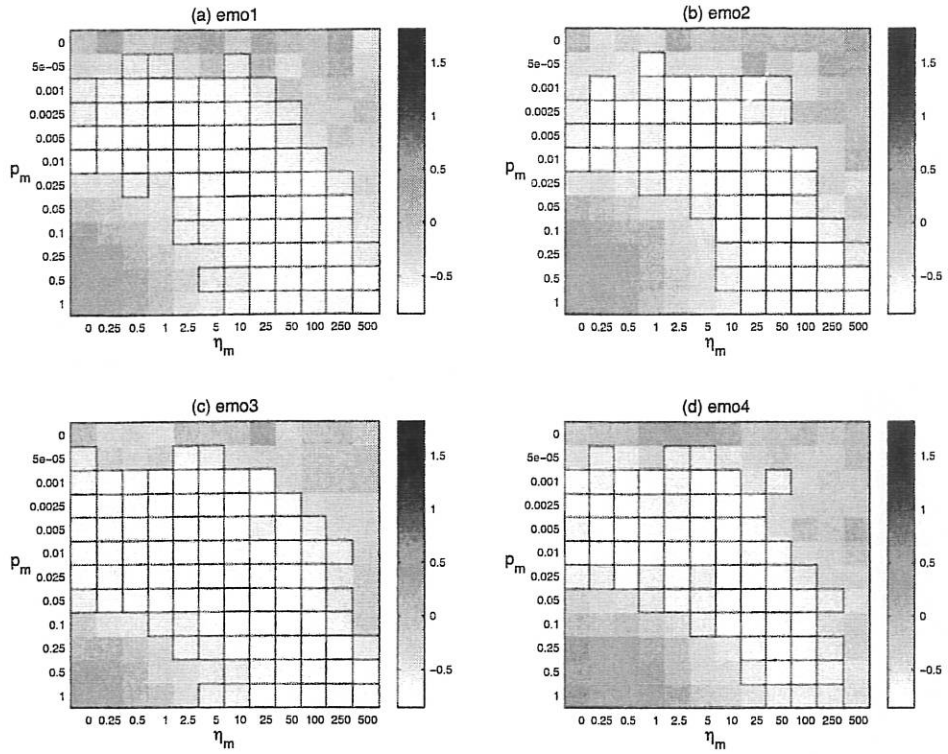

Figure 5.1: Mutation proximity maps for DTLZ2(6).

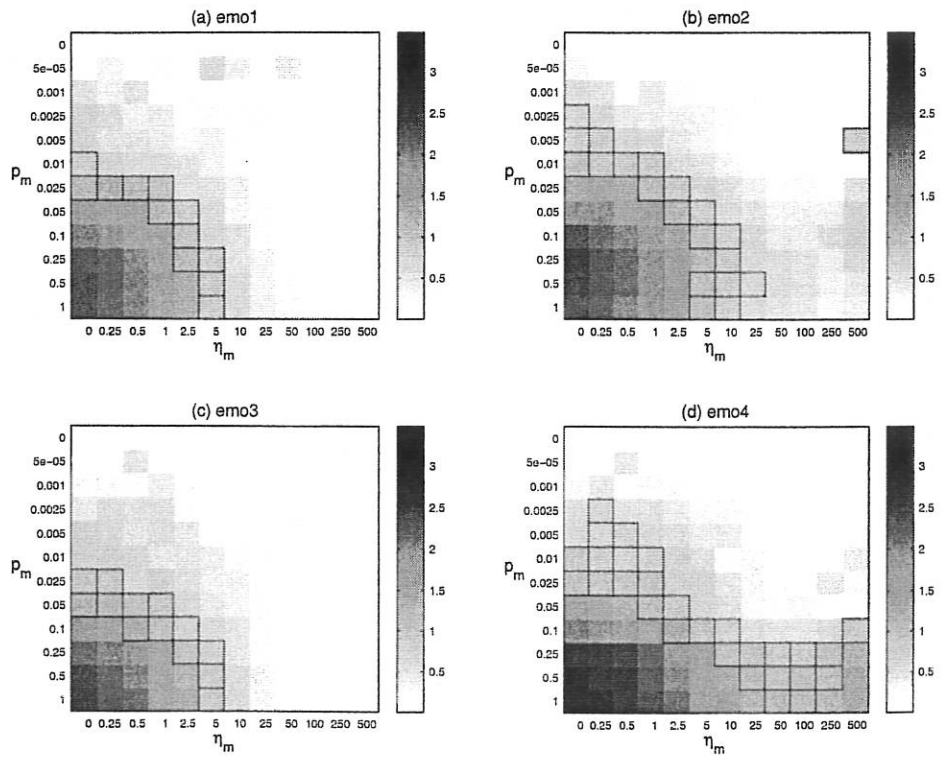

Figure 5.2: Mutation spread maps for DTLZ2(6).

approximation sets with both good proximity and good spread for the NSGA-II algorithm (Figure 5.1d and Figure 5.2d).

5.2.2 Varying the number of conflicting objectives

Figure 5.3 shows proximity maps for the 3-objective instance of DTLZ2. Maps for the 12-objective instance are shown in Figure 5.4. Considering these maps together with Figure 5.1, it can be seen that the highlighted region of good proximity remains largely invariant as the number of conflicting objectives, M, changes for the baseline emo1. The added benefit of elitism for improved proximity is evident across all instances (compare Figure 5.1a to 5.1c, 5.3a to 5.3c, and 5.4a to 5.4c) but the magnitude of that benefit appears to decrease with increasing M. The band of good proximity becomes increasingly thin for the non-elitist crowding emo2 as the number of conflicting objectives increases, as shown by Figures 5.3b, 5.1b, and 5.4b. The addition of elitism becomes increasingly incapable of compensating for these losses and can make matters worse, especially in regions of high mutation probability and small mutation perturbation (Figures 5.3d, 5.1d, and 5.4d).

Interesting variation in behaviour is observed in the map region of high mutation probabilities coupled with large mutation perturbations. In the non-crowding algorithms, emo1 and emo3, few changes are evident as M increases. However, levels of divergence are seen to increase with M for the diversity-promoting emo2 and emo4. The incorporation of elitism would appear to lead to further deterioration in this region (compare Figure 5.4d to Figure 5.4b) for 12 objectives, but makes little difference for 6 objectives (Figure 5.1b and 5.1d) and provides slightly improved proximity for the 3-objective task, as shown in Figure 5.3b and 5.3d.

The bands of good spread identified in Figure 5.2 are seen to broaden for smaller numbers of conflicting objectives, as demonstrated by the results for the 3-objective instance of DTLZ2 in Figure 5.5. In particular, the large values of spread in the lower-left corner of the map for emo4 are eliminated. As M increases, the bands of proximity become slightly thinner but do not disappear. Response maps for the 12-objective instance are shown in Figure 5.6. Performance in the region to the lower-left of the band is seen to become increasingly uniform except for the case of emo4 in Figure 5.6d where spread continues to increase.

As M increases, the number of occasions on the configuration grid where good proximity is matched by good spread becomes increasingly few, with the elitist algorithms appearing to offer the greatest opportunities. The classical configuration of $[p_m = |\mathbf{x}|^{-1}, \eta_m = 20]$ can be seen to be one of the more likely locations for this occurrence for emo4. The evidence in Figures 5.4 and 5.6 suggests that this will not be the case for any of the decomposed versions of NSGA-II.

5.3 Discussion

Explanations for the observed behaviour described in Section 5.2 can be sought by considering the processes (outlined in Section 3) that drive an EMO algorithm towards a good approximation of the global trade-off surface, and by monitoring the variables that impact on these processes. EMO search can be described as a trade-off between exploration and exploitation (Bosman and Thierens, 2003). In this section, particular focus is applied to three areas of configuration-space that relate to three different EE trade-off alternatives.

5.3.1 Highly exploratory configurations

The bottom-left map region is an area of high exploration and low exploitation. For all four emo algorithms the proximity cost is seen to increase beyond the initial measured value. The large standard deviation on the mutation distribution (corresponding to a near piece-wise uniform mutation) leads to a bias towards boundary values for decision variables independently from selection, as detailed in Section 3.3.1. Since proximity to the global trade-off surface is determined for DTLZ2 by a quadratic function $g(\mathbf{x})$ with minimum at $\mathbf{x} = 0.5$

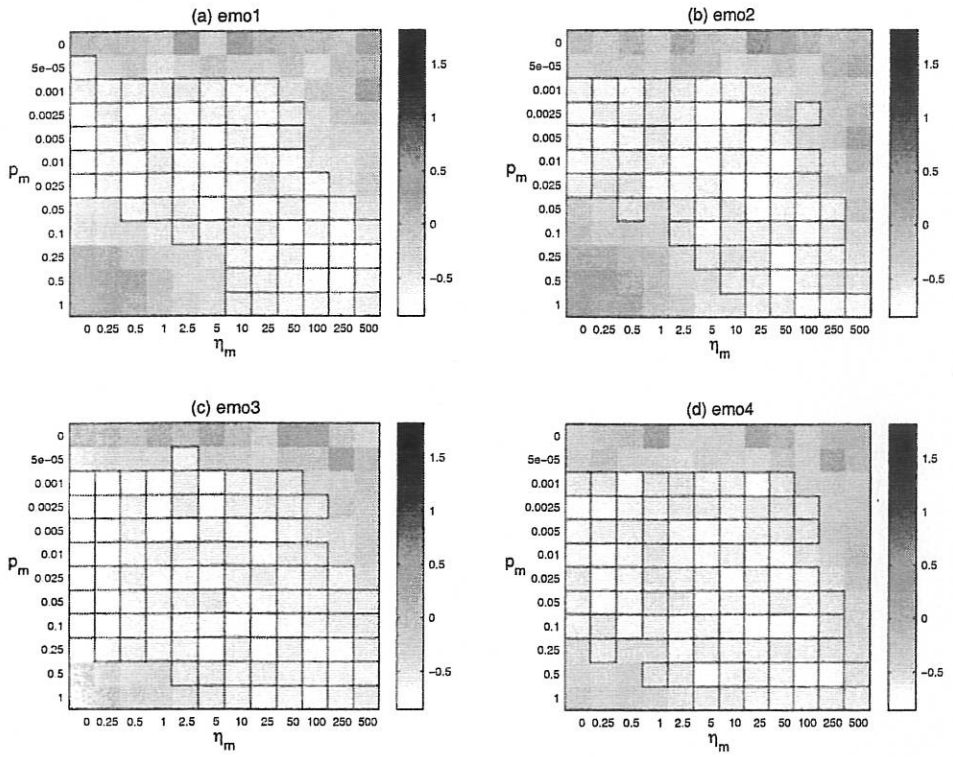

Figure 5.3: Mutation proximity maps for DTLZ2(3).

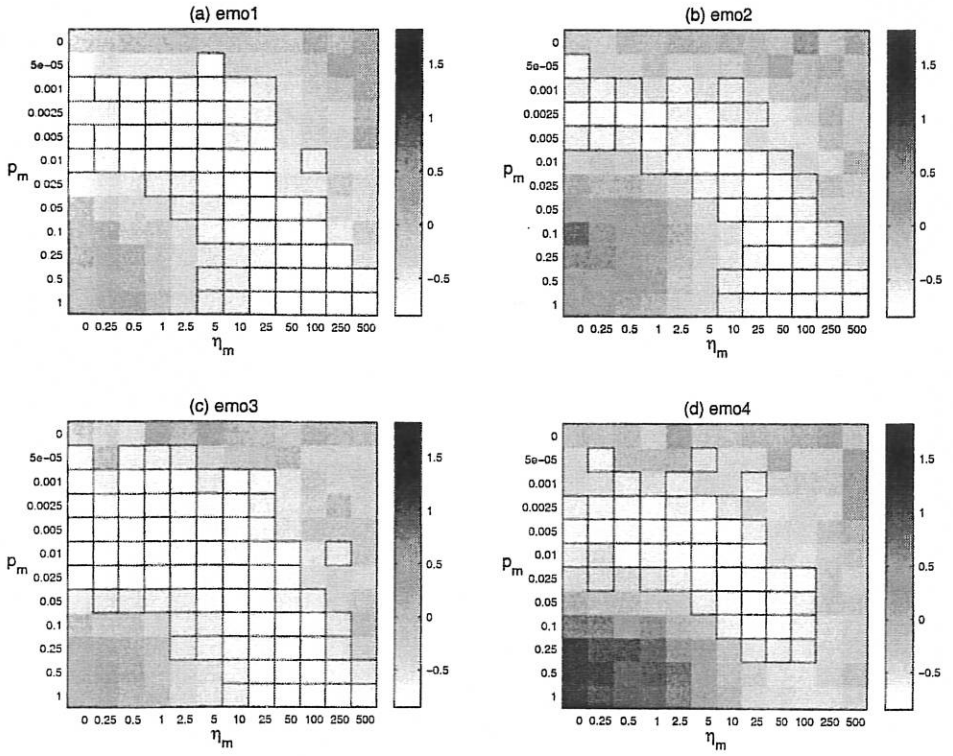

Figure 5.4: Mutation proximity maps for DTLZ2(12).

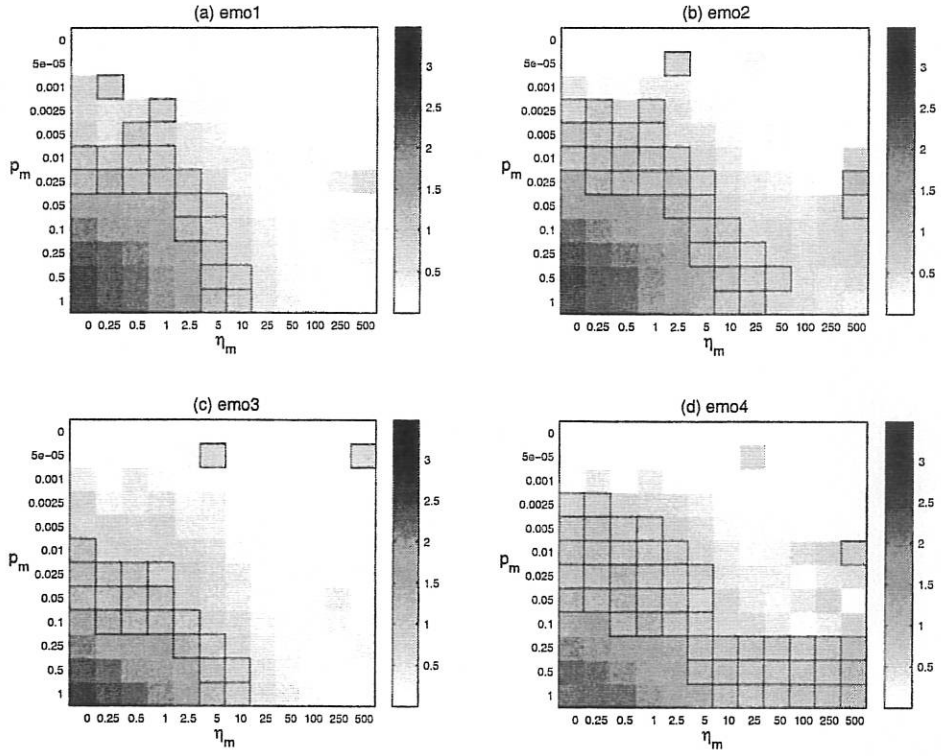

Figure 5.5: Mutation spread maps for DTLZ2(3).

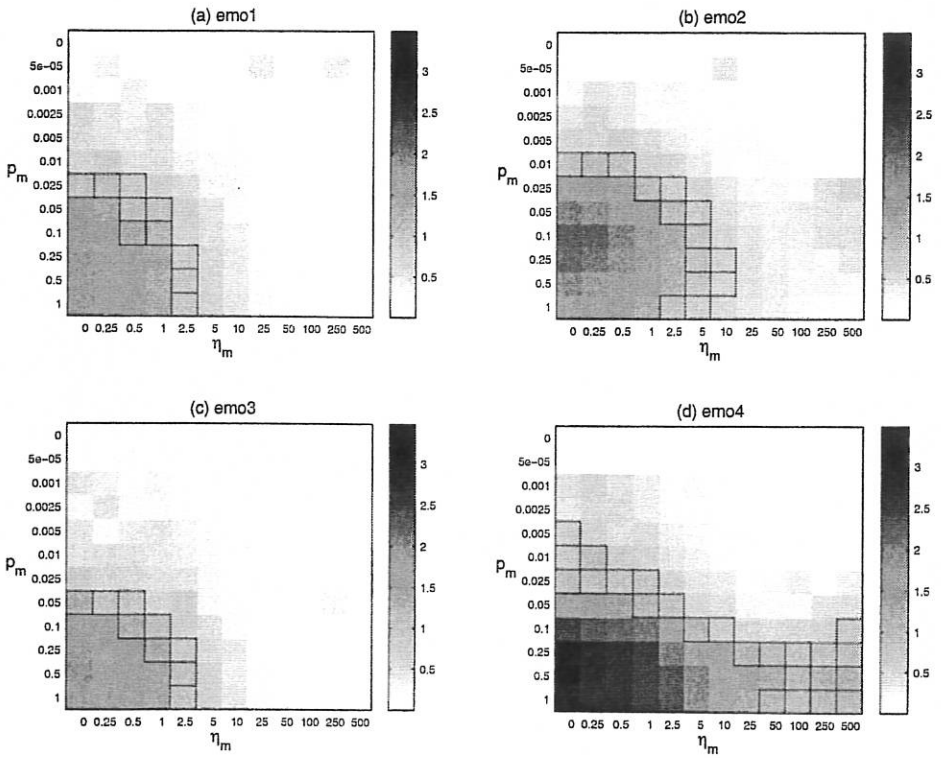

Figure 5.6: Mutation spread maps for DTLZ2(12).

(Equation 4.1), and $\mathbf{x} \in [0\ 1]$, a selection-free bias exists towards poor convergence-variable values. Also, as argued in Section 3.1, if the relationship between child location and parent location is weak then the effect of any selection-for-variation mechanism (diversity-promoting or otherwise) will be largely attenuated. Thus, performance of the non-elitist emo1 and emo2 algorithms is seen to be largely similar in this region of configuration-space. This result is evident from the associated map areas in Figures 5.1 through 5.6.

Non-random selection-for-survival mechanisms do impact on search performance under highly exploratory variation. The elitism in emo3 provides a reference to compare new solutions to. Thus, if the largely directionless $v\{s_v(P[t])\}$ search produces dominated children these will be eliminated. In the non-elitist emo1 and emo2 this would not be the case. However, since the probability of producing a non-dominated child relative to its parent increases with M, the improvements that arise from emo3 reduce in turn. The diversity-promoting s_s mechanism will be active in emo4 because the proportion of the population that is non-dominated is high, as shown in Figure 5.7. This bias towards remote solutions (explained in Section 3.2) leads to further deterioration in proximity values for high M. From Figure 5.6, it is clear that the corresponding spread values are also too large.

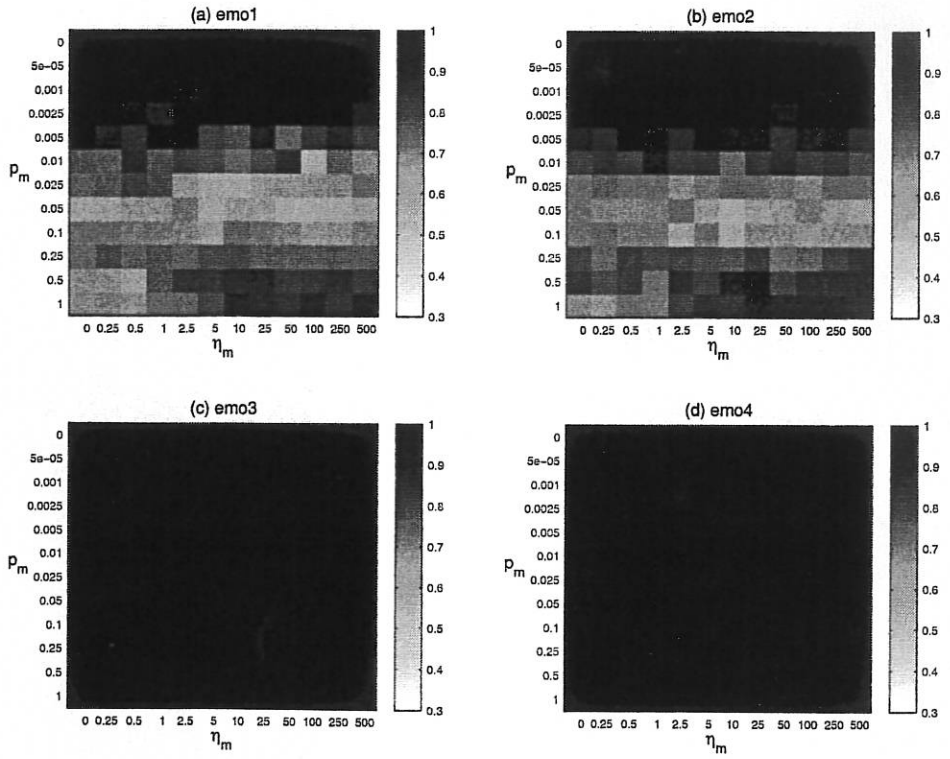

Figure 5.7: Proportion of the population that is non-dominated for DTLZ2(6).

5.3.2 Highly exploitative configurations

In the upper-most map regions of configuration-space, exploration levels are very low and exploitation is the dominant process. Proximity indicator profiles, generated over the course of the evolution, show very slow convergence towards the global Pareto front. As evident from the spread response maps of Figures 5.2, 5.5, and 5.6, the achieved spread values are very small in this region, indicating that the approximation set represents a diminutive area of objective-space.

Mutation success rates are very low in this region of configuration-space as shown in Figure 5.8, but they are not zero across all generations. Thus, long-term progress can be achieved on a small number of superior solutions. There is no observable bias towards remote solutions for two key reasons: (i) in the initial population s_v and s_s comparisons tend not to be based on density because the proportion of locally non-dominated solutions in the population is low (at the natural proportion for DTLZ2, depicted in Figure 4.1 for a random sample of 1000 decision vectors), and (ii) as the evolution progresses, non-dominated solutions become present in the population multiple times because of the mainly copying nature of variation. As explained in Section 3.2, since the density estimator is a version of first-nearest-neighbour, the densities of all solutions will be at a maximum. This renders the density comparison stages of s_v and s_s redundant, even when implemented.

5.3.3 Intermediate trade-offs between exploration and exploitation

In the region of high mutation probabilities coupled with small mutation perturbations, mutation operates effectively on the DTLZ2 problem structure to provide high success rates (see Figure 5.8). In these conditions good convergence towards the global front can be achieved. Also, since the diversity-promoting parts of s_s and s_v will be operative in `emo2` and `emo4` because of the high proportion of the population that is non-dominated, reasonable spread can also be obtained without risk of progressive movement towards poor-proximity areas. As M is increased, the success rates decrease (see Figure 5.9), thus making convergence more difficult and preservation of remote areas with poor proximity more likely for diversity-promoting schemes.

Good convergence to the true front is also achieved in regions of intermediate mutation probability coupled with large mutation perturbations. The nature of the variation is different to that previously described. In the above case, small-scale perturbations were applied to many elements of a candidate solution, whereas in the current situation large-scale perturbations are applied to few elements. Success probabilities are lower for this latter type of variation, as indicated by the associated region of Figure 5.8, but the magnitude of success will be higher leading to a similar overall convergence rate. Preservation and exploitation of remote, poor proximity, solutions is also mitigated in this area by the large number of copied non-dominated solutions.

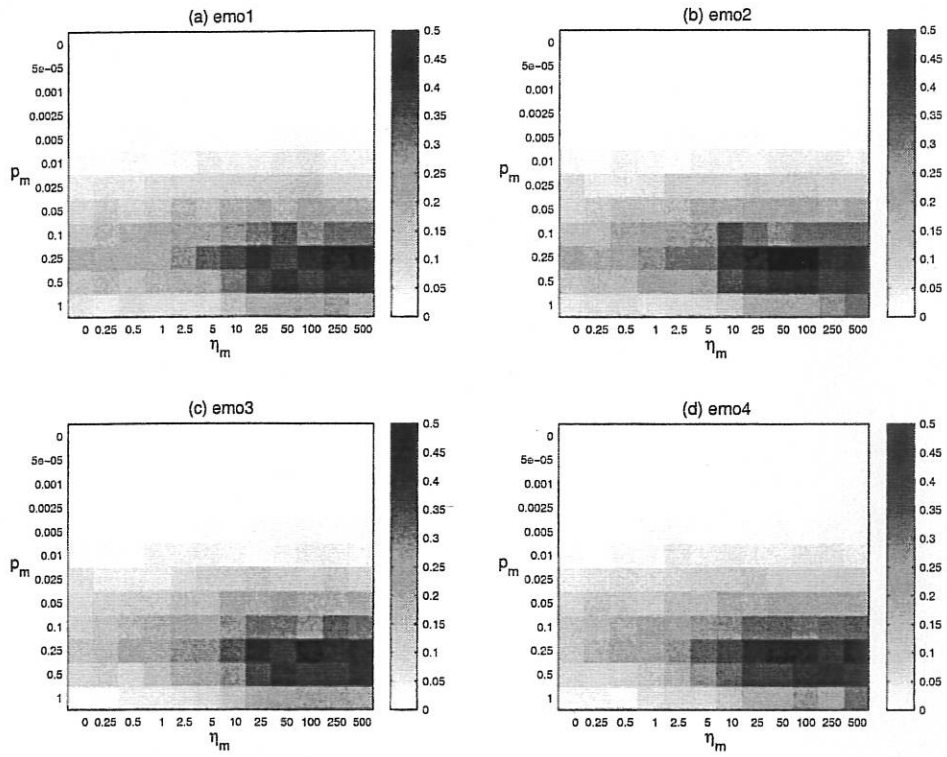

Figure 5.8: Proportion of child values that dominate their parent for DTLZ2(3).

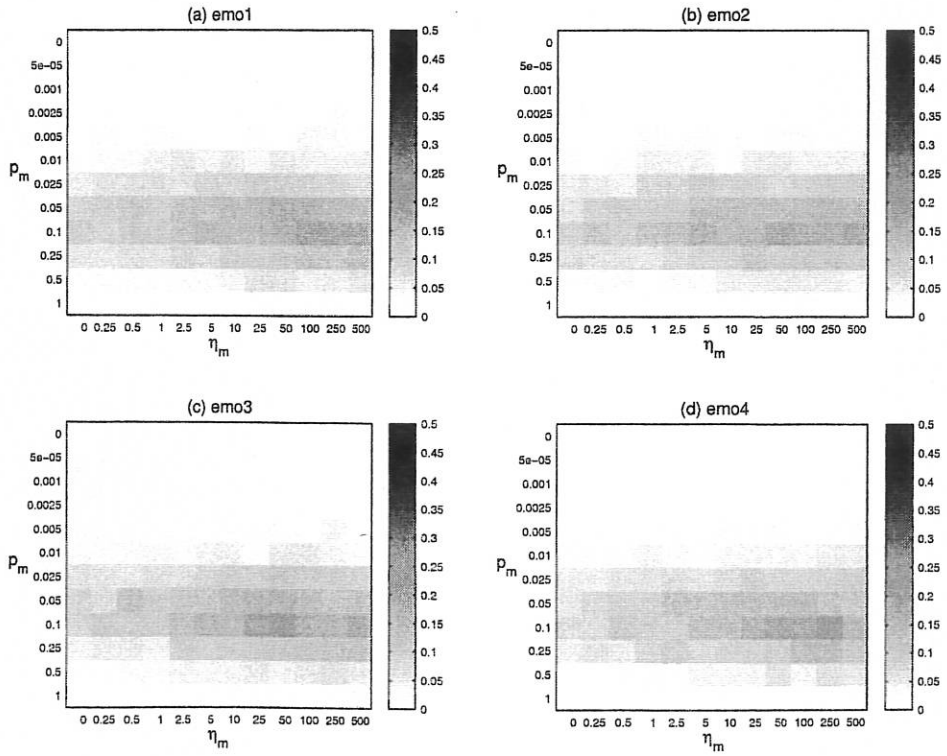

Figure 5.12: Proportion of child values that dominate their parent for DTLZ2(12).

6 Results and Analysis: Recombination

6.1 Recombination: Synopsis

The fundamental results to be drawn from the recombination inquiry are summarised in this section. The interested reader can find further details and explanations in the observation-based Section 6.2 and the discussion-based Section 6.3 that follow this synopsis.

- A suitable trade-off between exploration and exploitation for good many-objective performance can be specified via recombination operator configuration settings.
- For a large set of configuration options, recombination is very good at fulfilling the aim of good proximity for small numbers of objectives. A large sweet-spot for the spread property can be achieved by incorporating explicit diversity-promoting mechanisms.
- Proximity and spread sweet-spots deteriorate rapidly across all algorithms as the number of conflicting objectives to be optimised is increased. Approximation sets with a large spread of solutions but *very* poor proximity are produced for an increasingly large choice of SBX configurations.
- The classical choice of SBX operator settings for NSGA-II relate to good performance for small numbers of conflicting objectives, but become increasingly inappropriate as the number of objectives is increased. For large numbers of objectives, the proximity of the evolved approximation set is demonstrated to be almost as poor as can be generated anywhere in objective-space. This behaviour occurs when recombination success rates are low and the density-promoting selection components are active, and is linked to the self-adaptive nature of the SBX exploratory power.
- The inclusion of decision-variable exchange in the SBX procedure amplifies and extends the behaviour observed in the absence of exchange.

6.2 Observations[5]

6.2.1 6 objectives

DTLZ2(6) proximity response maps for the four selection schemes, when combined with SBX variation, are shown in Figure 6.1. As shown in part (a) of the figure, a small region of good proximity for emo1 is observed for intermediate internal crossover rates, such as p_{ic} = 0.05, coupled with high recombination perturbations (which correspond to low values of η_c such as 0.5). A broken band of good proximity is visible, stretching from this region towards the area of high p_{ic} and intermediate η_c. This band is seen to strengthen and thicken in all directions for the elitist emo3 (Figure 6.1c). The region of good proximity for the diversity-promoting emo2 is of a similar size to emo1 but is shifted towards values of lower p_{ic} in configuration-space, as shown in Figure 6.1b. This region expands for emo4 to include both higher and lower probabilities of internal crossover for larger variation operator perturbations.

Poor proximity is evident across all four algorithms in regions of high internal crossover probability coupled with large variation perturbations. In Figure 6.1b, especially poor proximity is observed for the diversity-promoting emo2 in regions of high crossover probability and intermediate variation perturbation, such as [p_{ic} = 0.5, η_c = 5]. This behaviour is further exacerbated in this region for the elitist, diversity-promoting emo4 as indicated in

[5] For details on the types of response map presented in this section, and for details of how to view the maps, refer to Section 4.5.

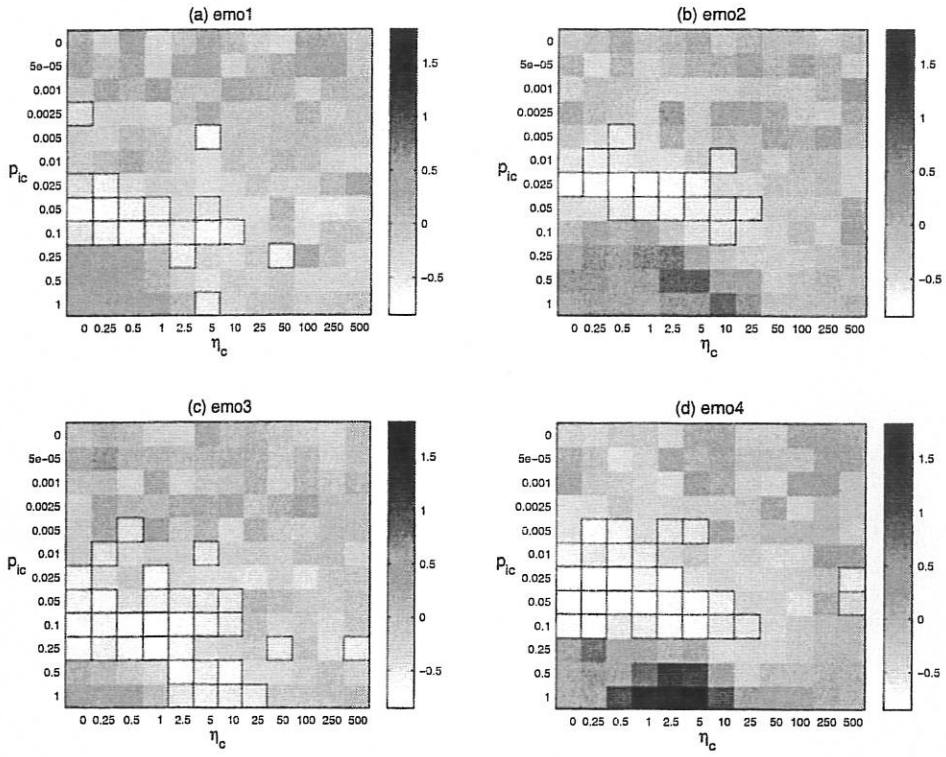

Figure 6.1: Recombination proximity maps for DTLZ2(6).

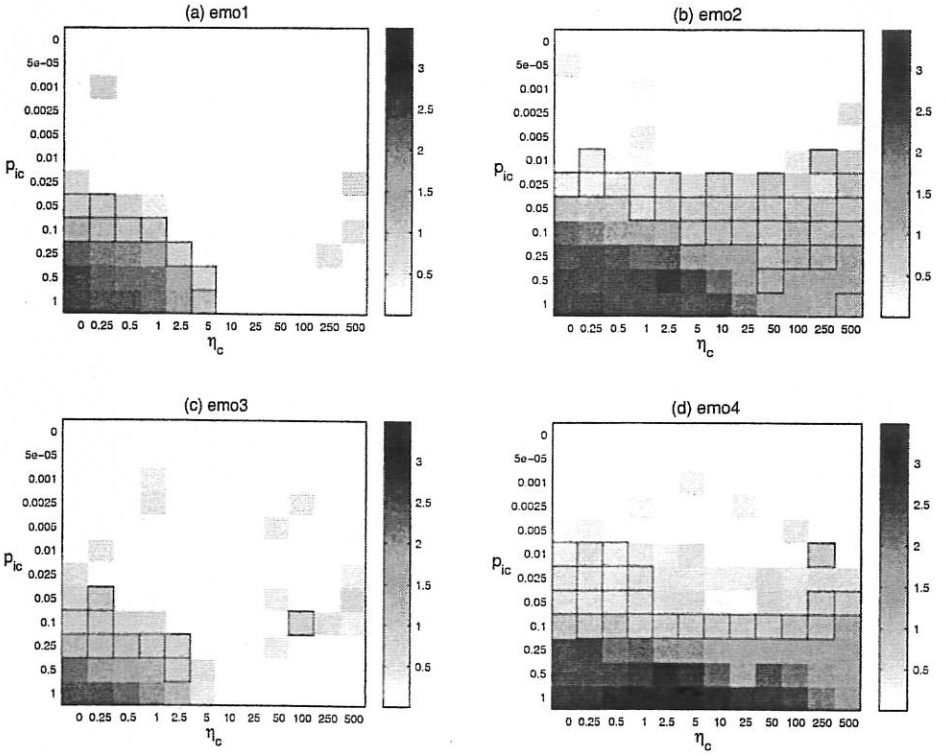

Figure 6.2: Recombination spread maps for DTLZ2(6).

Figure 6.1d. The algorithm suffers from severe divergence in this area of configuration-space. In other regions of configuration-space, across all four selection schemes, the level of convergence appears subject to random noise.

Spread response maps obtained for the six-objective instance of DTLZ2 are shown in Figure 6.2. Very small spread values (coloured white) are widespread in all map areas, except for the bottom-left region, for the algorithms without a diversity promotion mechanism: emo1 (Figure 6.2a) and emo3 (Figure 6.2c). In these regions, the approximation set has reduced to a single candidate solution. A thin band of good spread is apparent for emo1 and emo3, stretching from regions of intermediate p_{ic} coupled with large recombination perturbations to regions of intermediate η_c coupled with high p_{ic}. This band is coincident with a portion of the good proximity area shown in Figures 6.1a and 6.1c. Spread indicator values are seen to become increasingly large in the direction of increasing p_{ic} and $1/\eta_c$. The approximation sets span large regions of (non-optimal) objective-space in these cases.

Spread response maps for the diversity-promoting algorithms, emo2 and emo4, are shown in Figures 6.2b and 6.2d respectively. In the upper-half of both maps (corresponding to low to intermediate values of p_{ic}) spread values are very small, again indicating convergence to a highly limited area of objective-space such as a single candidate solution. Good spread values are seen in a mainly horizontal band of intermediate internal crossover probability. This band is more established in regions of high recombination perturbation [$p_{ic} = 0.05$, $\eta_c = 0.25$] for the elitist emo4, and in areas of low perturbation [$p_{ic} = 0.025$, $\eta_c = 50$] for the non-elitist emo2.

Spread indicator values that are too large are witnessed across all four selection schemes in the region of high p_{ic} coupled with large expected variation perturbation. The region is noticeably larger for the diversity-promoting algorithms, where it extends to cover all values of recombination magnitude. Spread values are particularly high for the elitist emo4, especially in areas of intermediate η_c such as [$p_{ic} = 1.0$, $\eta_c = 5$].

The classical choice of [$p_{ic} = 0.5$, $\eta_c = 15$], with $p_c = 1.0$, from the EMO literature only corresponds to near-optimal proximity in the elitist emo3. Rather poor proximity is witnessed for the other schemes, especially emo4 (NSGA-II). In the optimisers without an explicit diversity-promotion scheme, the approximation set spans only a very small area of objective-space with respect to the true trade-off surface. By contrast, spread values are too large for the diversity-promoting algorithms.

6.2.2 Varying the number of conflicting objectives

The proximity response maps achieved for the three-objective instance of DTLZ2 are shown in Figure 6.3, whilst maps for the twelve-objective instance are provided in Figure 6.4. Overall, there is clear evidence to show that the proportion of configuration-space that is associated with good proximity performance decreases substantially as the scale of the optimisation task increases. The strong band of good proximity obtained by emo1 for three conflicting objectives (shown in Figure 6.3a) contracts as the number of objectives is increased. For the twelve-objective instance of DTLZ2, it is almost non-existent (Figure 6.4a). For small numbers of conflicting objectives, emo2 provides better proximity in the direction of increasing η_c than emo1, but this benefit is overwhelmed by the general convergence problems encountered by all algorithms in this region, as task size increases. These difficulties would seem to be enhanced for the diversity-promoting algorithms as shown by the strong divergence behaviour experienced in the region with exemplar [$p_{ic} = 0.25$, $\eta_c = 2.5$] in Figures 6.3b, 6.3d, 6.4b, and 6.4d.

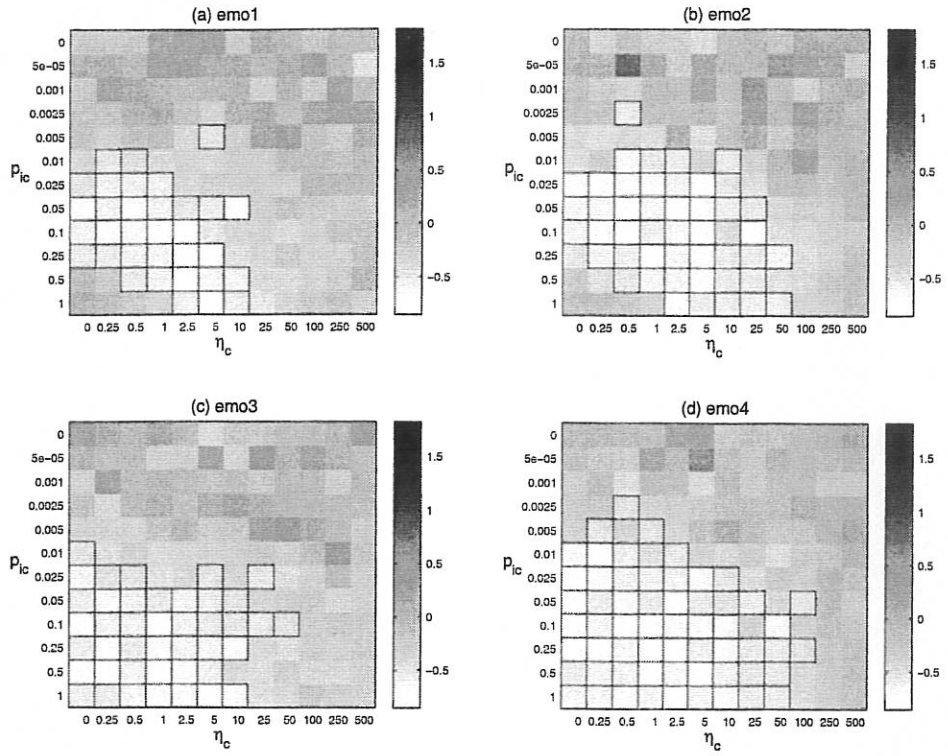

Figure 6.3: Recombination proximity response maps for DTLZ2(3).

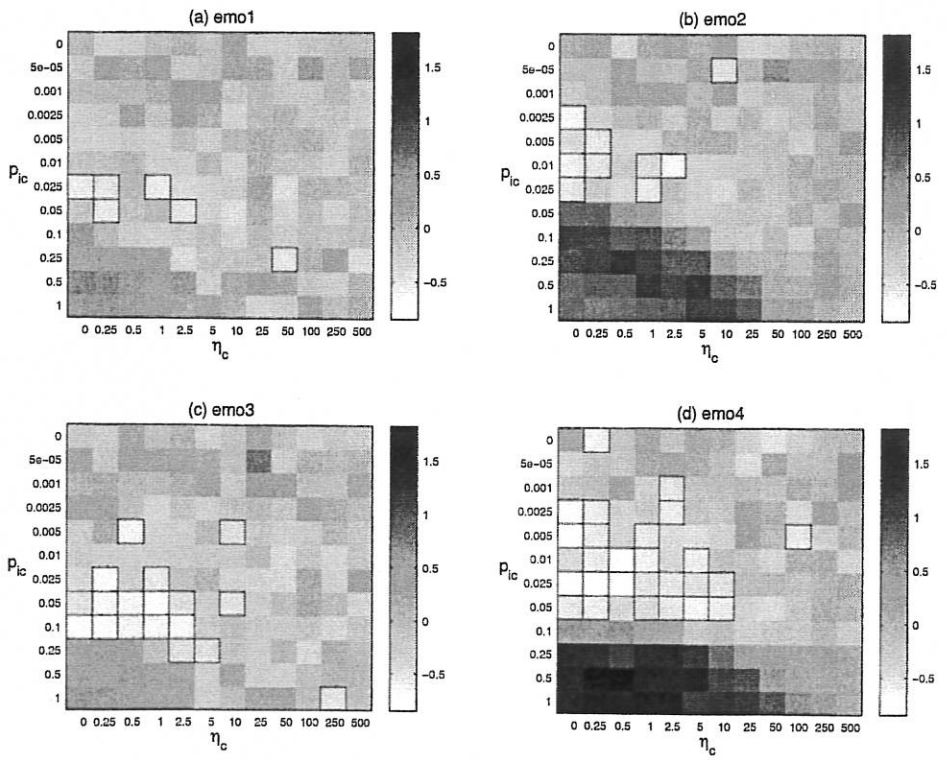

Figure 6.4: Recombination proximity response maps for DTLZ2(12).

Interesting behaviour is demonstrated by the elitist algorithms, emo3 and emo4, as the number of conflicting objectives is increased. For three objectives, as illustrated by comparing Figure 6.3c to Figure 6.3a and Figure 6.3d to Figure 6.3b, the elitist algorithms offer significantly improved proximity over their non-elitist equivalents for regions of high p_{ic} and large recombination perturbation, even including [p_{ic} = 1.0, η_c = 0.0]. The proximity sweet-spot for emo4 (NSGA-II) covers approximately half of the total configuration-space considered, as shown in Figure 6.3d. However as the number of conflicting objectives is increased beyond the levels normally considered by EMO designers, as shown in Figures 6.1 and 6.4, the ability of the elitist algorithms to offer improved proximity over their non-elitist counterparts is limited to regions of intermediate p_{ic} and high $1/\eta_c$. Performance in areas of high p_{ic} (rates considered standard for a recombination operator) becomes increasingly worse than the non-elitist equivalent. Indeed for emo4 solving DTLZ2(12) with configuration [p_{ic} = 1.0, η_c = 0.25], the achieved proximity value is close to the worst possible value obtainable for this optimisation task (Figure 6.4d). Note that performance in map regions of low internal crossover probability together with low recombination magnitude is largely invariant regardless of selection scheme and task size.

The spread response maps generated for the three-objective and twelve-objective instances of DTLZ2 are shown in Figures 6.5 and 6.6 respectively. The region in configuration-space in which the approximation set represents only a small proportion of the global trade-off surface (often the spread is zero because the set contains only a single, unique, candidate solution) is seen to be largely independent of M for each optimiser. It should be noted, though, that there is some evidence of improved spread in regions with exemplar [p_{ic} = 0.05, η_c = 250] for small M for the non-diversity-promoting emo1 and emo3. In the case of these latter algorithms, good spread (evident in the bottom-left map region of high p_{ic} and high $1/\eta_c$) tends to decrease with increasing M (compares Figure 6.5, 6.2, and 6.6) and the benefit of elitism over non-elitism in this area also tends to decrease in these conditions (compare the difference between Figure 6.5a and 6.5c to the difference between Figure 6.2a and 6.2c to the difference between Figure 6.6a and 6.6c).

For small numbers of conflicting objectives, as shown in Figure 6.5, good spread values are more widely distributed in configuration-space for algorithms that incorporate an explicit diversity-promotion mechanism. The results in Figure 6.5d suggest that the NSGA-II sweet-spot for good spread on DTLZ2(3) covers approximately half of all configuration-space considered. As the number of conflicting objectives is increased, the region of good spread reduces to intermediate internal crossover probabilities and the complete range of expected recombination perturbation magnitude. This can be seen by comparing Figure 6.5 to Figure 6.2, and then Figure 6.2 to Figure 6.6. Values of spread that are much too large are particularly evident for emo4. This behaviour is linked to the previously discussed proximity results and is seen to become exaggerated as M increases. Note that the rate of deterioration in spread performance, in terms of the number of sampled configurations that give reasonable performance, would appear to reduce as M increases.

From considering Figure 6.3 together with Figure 6.5 it is clear that there are many recombination operator configurations for which good proximity values can be married with good values for spread for small numbers of conflicting criteria. This is particularly true of the elitist, diversity-promoting, emo4 (NSGA-II). As M is increased, the number of such configurations becomes increasingly few. For the twelve-objective instance of DTLZ2, on the evidence of a single replication, none of the configurations considered can provide good proximity together with good spread for emo1 (see Figures 6.4 and 6.6). One configuration, [p_{ic} = 0.25, η_c = 2.5], offers this benefit for emo3. For the remaining diversity-promoting selection schemes, appropriate configurations can be found in the region of high $1/\eta_c$ and intermediate p_{ic} such as [p_{ic} = 0.025, η_c = 0.0].

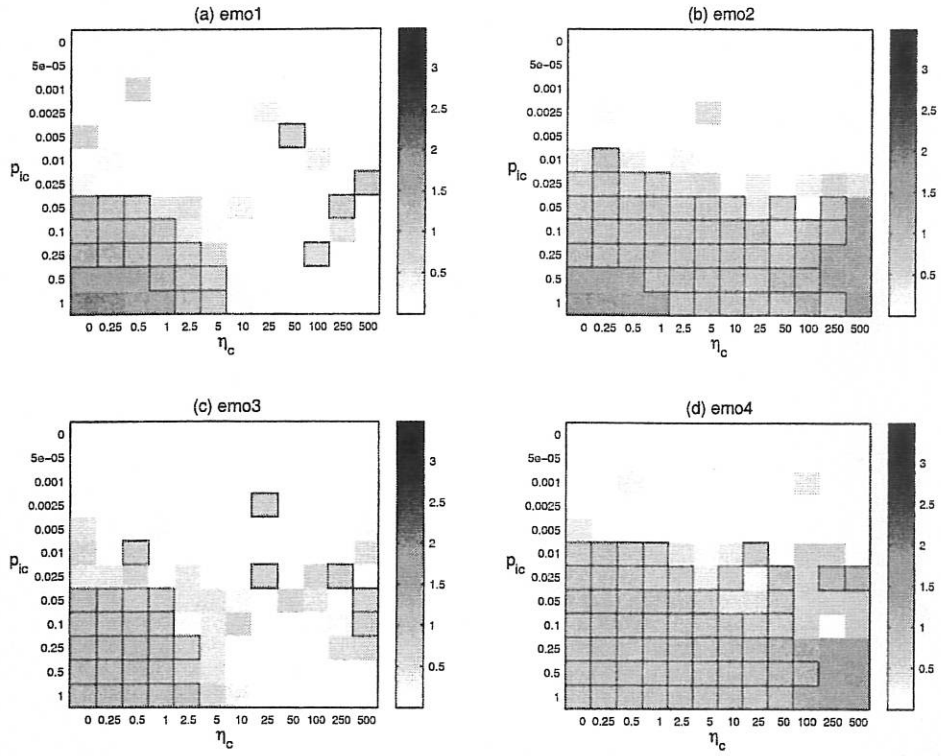

Figure 6.5: Recombination spread response maps for DTLZ2(3).

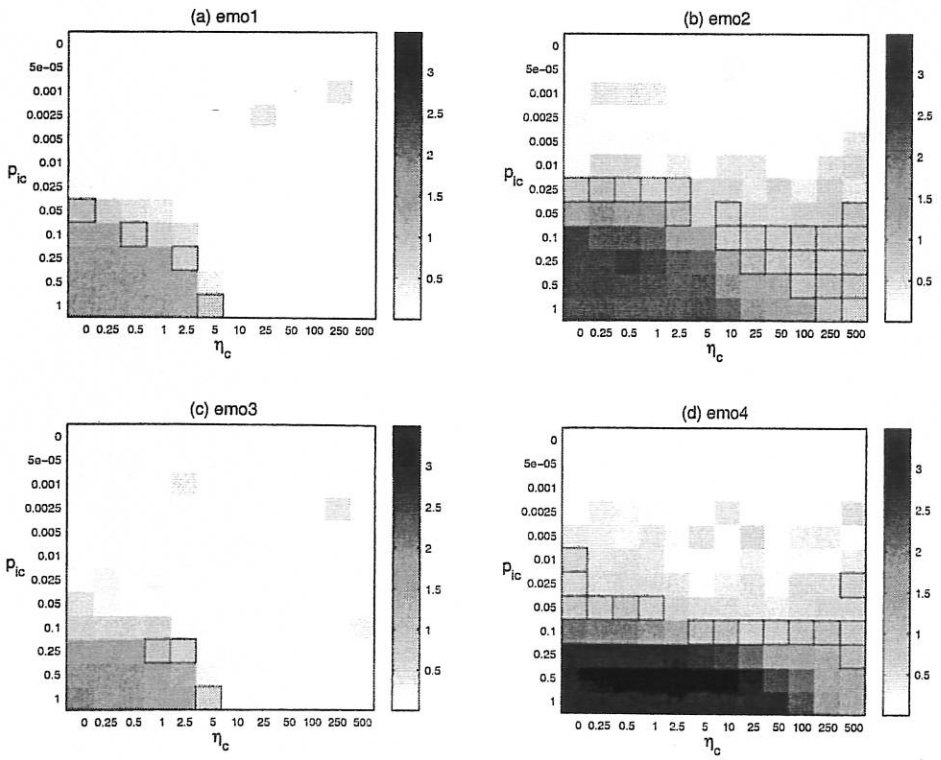

Figure 6.6: Recombination spread response maps for DTLZ2(12).

The standard SBX parameter settings in the literature for real parameter optimisation tasks such as DTLZ2 are [p_{ic} = 0.5, η_c = 15] with p_c = 1.0 (Deb *et al*, 2002b). The response maps generated in this study indicate that these settings are appropriate for small M instances of DTLZ2 (for any of the selection schemes considered), but would appear increasingly unlikely to provide acceptable approximation sets in terms of proximity and distribution as the scale of the task increases. As indicated by Figure 6.1d and Figure 6.2d, even for the six-objective instance of DTLZ2 the standard parameter settings would be expected to produce an approximation set with a worse proximity indicator value than that obtained from a random sample of decision-space, spread widely over non-optimal regions of objective-space.[6] As M increases, the quality of the solution set is demonstrated to deteriorate still further.

6.2.3 SBX with exchange

In the implementation of the SBX recombination operator described above, child values are obtained directly from the self-adaptive probability distributions. In some studies in the literature that consider this operator, the child values for a decision variable are then exchanged with non-zero probability p_e (usually set to 0.5). Researchers have reported that this extra mechanism, labelled herein as *SBXe*, leads to improved performance. DTLZ2(6) proximity response maps for SBXe are shown in Figure 6.7. The corresponding spread response maps are shown in Figure 6.8.

In terms of good proximity, the key difference between the response maps for SBX (Figure 6.1) and SBXe (Figure 6.7) is that the behaviour observed for regions of large expected perturbation magnitudes (defined strictly in terms of the SBX distribution) for SBX continues into areas with smaller η_c. Map regions where proximity was discovered to be particularly poor for SBX remain so after the addition of exchange, and divergent behaviour is seen to increase still further in the diversity-promoting algorithms (Figure 6.7b and Figure 6.7d).

Locations in SBX configuration-space that were found to offer good performance in terms of the spread indicator remain largely unchanged after the incorporation of exchange, as can be seen by comparing Figure 6.8 to Figure 6.2. The main discrepancy between the two sets of maps is that spread values are considerably larger in the region of high p_{ic} combined with intermediate η_c for the diversity-promoting algorithms in the case of SBXe.

As the number of conflicting objectives is varied, the maps for SBXe change (in a relative fashion) in the same manner as for SBX without exchange. Visual results are not presented to help constrain the size of the paper. In summary, the classic parameter settings are found to be again adequate for only small numbers of objectives.

6.3 Discussion

In terms of the trade-off between exploration and exploitation, and its relationship to the configuration map, the general *type* of behaviour exhibited by the algorithms under recombination is similar to that under mutation. However, the magnitude of the behaviour is observed to be much larger in the case or recombination (as is evident from comparing Figure 5.4 to Figure 6.4 for example). In Section 6.3.1, reasons are offered for this difference between the variation operators. The incorporation of exchange also affects optimiser responses. An explanation for this difference is proposed in Section 6.3.2.

---

[6] The boundary between overall good and poor performance for the classical operator settings has been found to lie between 4 and 5 objectives for DTLZ2.

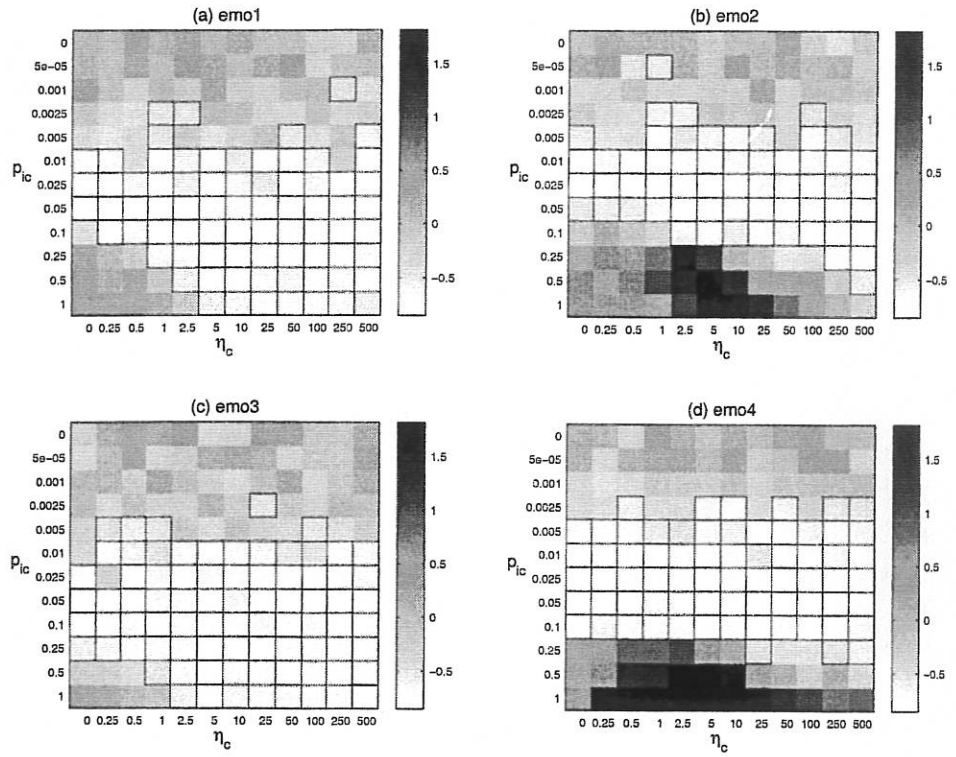

Figure 6.7: Recombination (with exchange) proximity maps for DTLZ2(6).

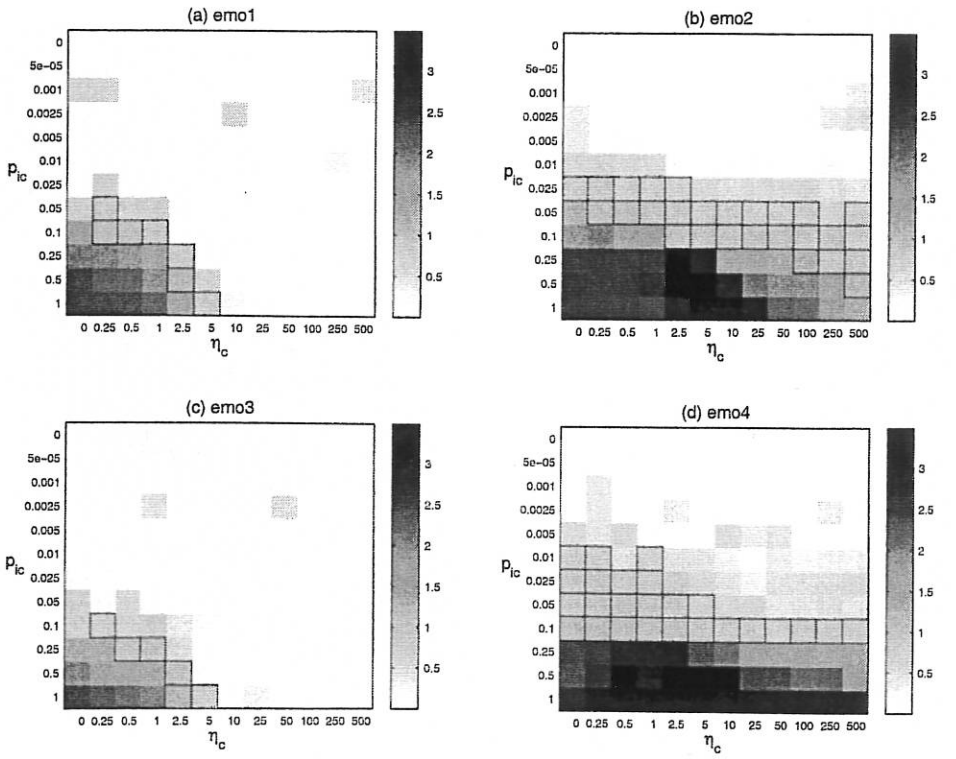

Figure 6.8: Recombination (with exchange) spread maps for DTLZ2(6).

6.3.1 Comparative analysis of mutation and recombination results

Highly exploratory configurations

For recombination, as was previously demonstrated for mutation in Section 5, good proximity cannot be obtained in configurations of high variation probability and large variation magnitude. Divergence has been shown to be more severe in the recombinatorial case (for example, compare Figure 5.4 to Figure 6.4). In the case of mutation, it was argued that this behaviour arose because of (i) a bias in the variation operator (independent of selection) towards poor convergence-variable values and (ii) active selection of remote non-dominated solutions with poor proximity values (via diversity-promoting mechanisms). In the case of SBX, cause (i) is exaggerated because of the practice of cropping infeasible decision values to the nearest feasible equivalent, as discussed in Section 3.3.2. For DTLZ2, the repaired values correspond to the worst case settings for proximity (Equation 3.1). Many repairs will be required for exploratory configurations, when the expected perturbation (controlled by η_c and the distance between parents) is large relative to the range of the decision variables.

In addition to the increased bias towards poor proximity values under random selection, the self-adaptive nature of the SBX operator is also argued to contribute heavily to the divergence behaviour. As explained in Section 2.3.3, if the parent values are close together then the expected perturbation magnitude is smaller than if the parents were far apart. Thus, given a large proportion of poor convergence-variable values in a population, the search can become increasingly limited to regions that are remote from the global trade-off surface. However, for the distribution-variables, the diversity-promoting mechanisms will maintain a good variety of solutions and so the search will remain of a larger scale for these variables. Thus, new remote solutions will be readily generated via sufficient perturbation to the distribution-variables, whilst convergence-variables remain at poor values.

The dynamic rescaling of expected variation magnitudes is the fundamental difference between SBX recombination and polynomial mutation. The ability of SBX to focus on a particular sub-region of decision-space enables the mechanisms of selection-for-variation even for very large $1/\eta_c$. For mutation, effective selective bias can only be achieved for selection-for-survival in highly exploratory environments.

Highly exploitative configurations

For configurations that involve low variation probability or low variation perturbation, the approximation set tends to converge to a very small region of objective-space (corresponding to small spread indicator values) if diversity-promoting mechanisms are not active. Spread values are particularly small when recombination is used, as shown by the examples in Figure 6.6. This behaviour arises because of the tendency of SBX to converge to the mid-point between parent solutions, given a limited variety of parent material. By comparison, polynomial mutation is a variation operator that will respect the multi-point characteristics of the search regardless of the diversity in the parent population.

Intermediate trade-off between exploration and exploitation

Good optimiser performance can be achieved using either mutation or recombination for intermediate variation probabilities and perturbations. Configurations have been identified for which recombination results are very good, and outperform anything that can be achieved using mutation. [p_{ic} = 0.025, η_c = 0.5] is one such example for emo4 solving DTLZ2(6) as shown in Figures 6.1d and 6.2d. As described in the highly exploratory analysis above, this behavioural discrepancy is due to the self-adaptive nature of the SBX search. If good convergence-variable values can be identified then SBX can focus the search on these regions regardless of the initial distribution parameter setting. By contrast, for mutation the expected perturbation is largely fixed for a particular setting of η_m (it only depends on the distance to

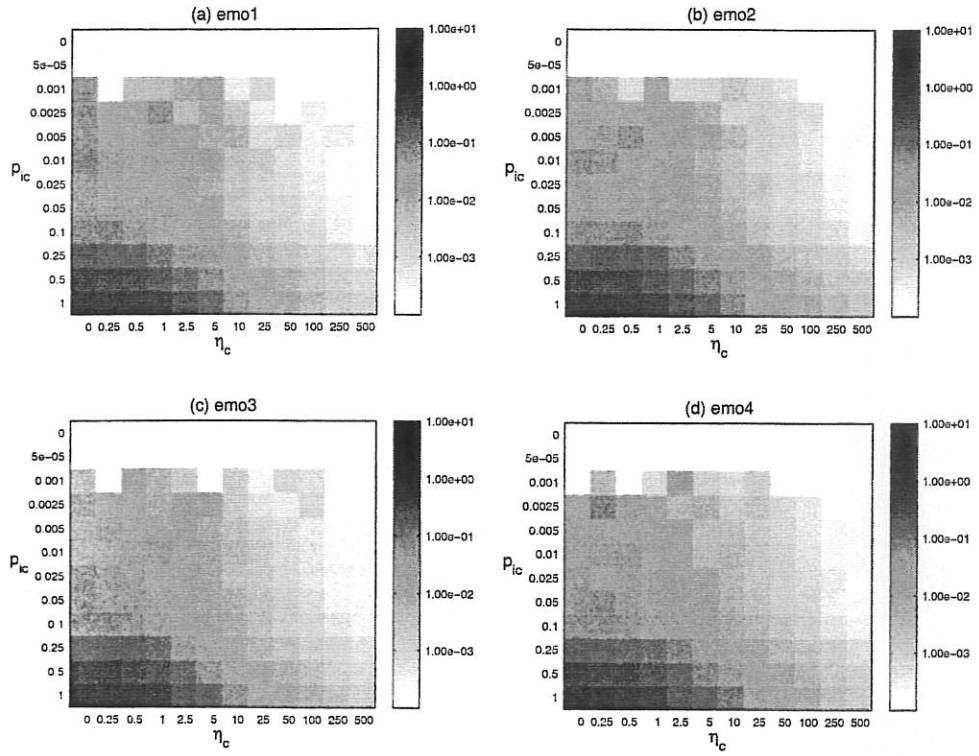

Figure 6.9: Median objective-space perturbations induced by SBX for DTLZ2(6).

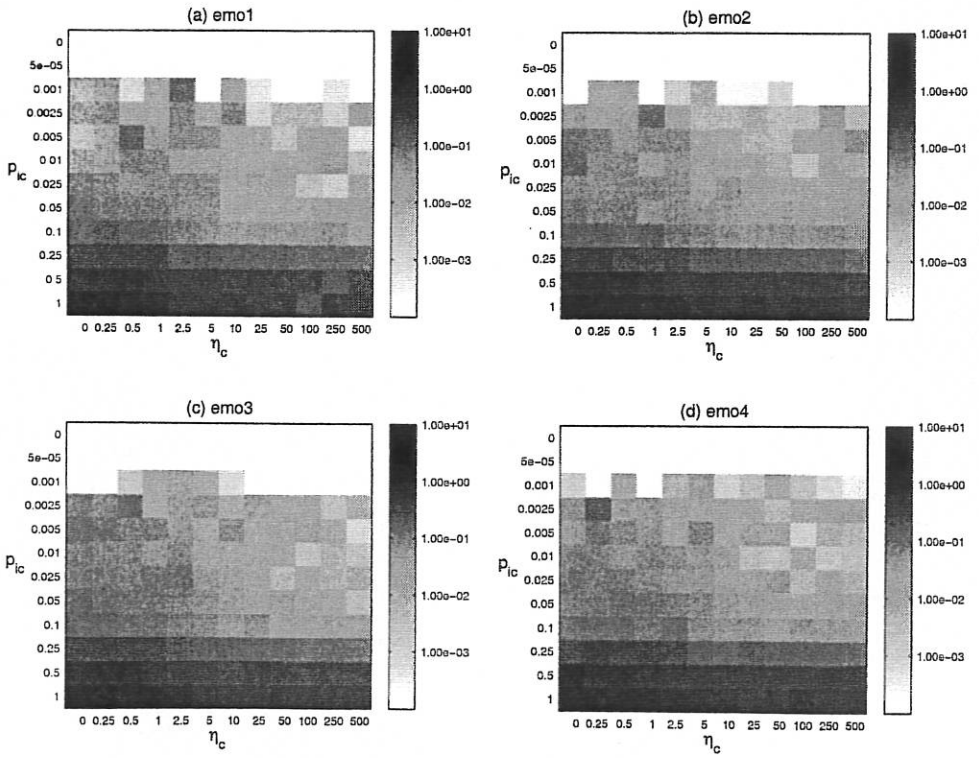

Figure 6.10: Median objective-space perturbations induced by SBXe for DTLZ2(6).

the feasible boundary in decision-space relative to η_m). Thus, SBX can retain reasonable variation success probabilities over a wide range of required search resolutions. It is more difficult to achieve good success rates under increasing search localisation for the non-adaptive (fixed η_m) polynomial mutation distribution.

6.3.2 Analysis of the incorporation of exchange

As demonstrated in Section 6.2.3, algorithm responses from the left-hand side of the configuration map for SBX without exchange are (approximately) reproduced on the right-hand side of the map for SBX with exchange. The maps of median magnitudes of the empirical variation perturbation for SBX are given in Figure 6.9. The corresponding maps for SBXe are shown in Figure 6.10. It is clear that the incorporation of exchange eliminates much of the effect of the SBX distribution parameter, η_c, on the expected magnitude of perturbation. Thus, response behaviour becomes largely invariant of η_c for SBXe.

7 Population sizing study

7.1 Introduction

Deb (2001) proposes the use of large population sizes as a potential method for achieving good evolutionary many-objective optimisation results. The rationale behind this is that the proportion of the population that is non-dominated in a given sample reduces as the size of the sample increases, allowing improved discrimination based on Pareto dominance. This was verified in Figure 3.1 for the DTLZ2 task. In general, large samples will provide better quality representations of whichever sectors of objective-space are currently the foci.

Unfortunately, in practice, the use of large population sizes is prohibitive because of the computational resources required to evaluate potential solutions and process them during an iteration of the optimiser. The first of these issues compels many real-world EMO applications, such as that documented by Thompson *et al* (1999), to limit the population size to fewer than 50 individuals. This is such a serious issue for computational search and optimisation techniques in general that 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 in evolutionary algorithms, known as *fitness inheritance*, also considers how to reduce evaluation requirements (Smith *et al*, 1995).

At present, population sizes for generational MOEAs tend to lie between approximately 10 and 1000 (Coello *et al*, 2002). Thus, in this section of the inquiry, the six-objective DTLZ2 task from the previous sections is re-solved for the same emo algorithms with population sizes of 10 and 1000. Note that more noise would generally be expected for small population sizes because of the small sample sizes considered. The results for mutation are presented in Section 7.2, followed by the results for recombination in Section 7.3. The population sizing study concludes with the key EMO design issues to be drawn from the analysis.

7.2 Mutation[7]

The DTLZ2(6) proximity response maps for the emo algorithms with a population size of 10 (P10) are shown in Figure 7.1. The maps for a population size of 1000 (P1000) are provided in Figure 7.2, whilst the maps for P100 were previously given in Figure 5.1. The proximity results in general, and the sweet-spots in particular, are seen to be largely invariant over the

[7] For details on the types of response map presented in this section, and for details of how to view the maps, refer to Section 4.5.

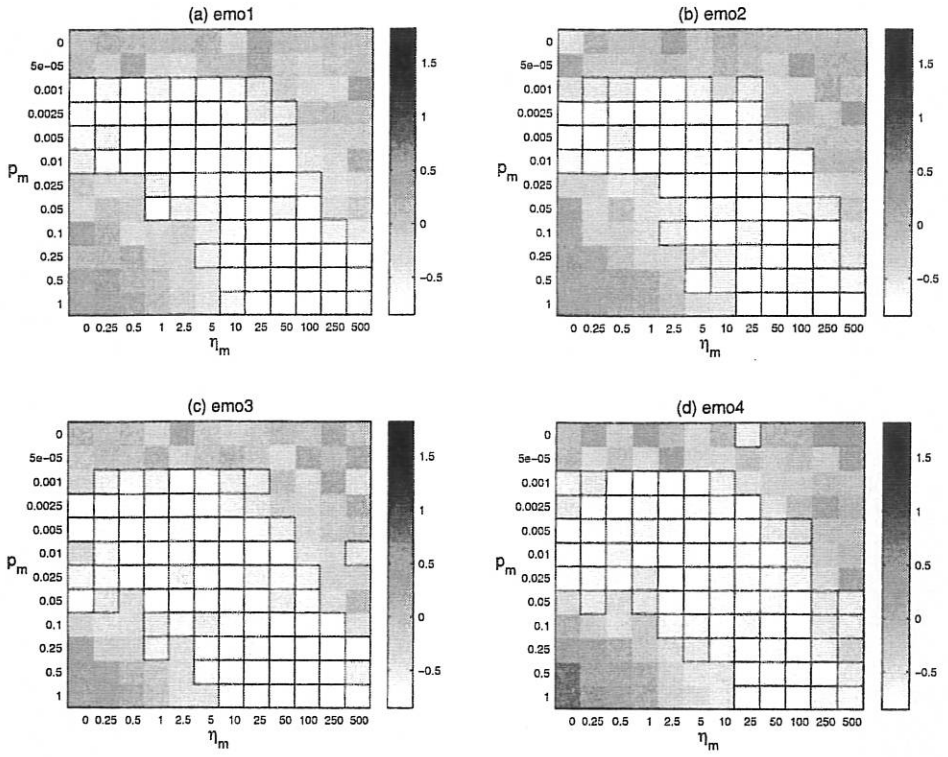

Figure 7.1: Mutation proximity maps for DTLZ2(6) with P10.

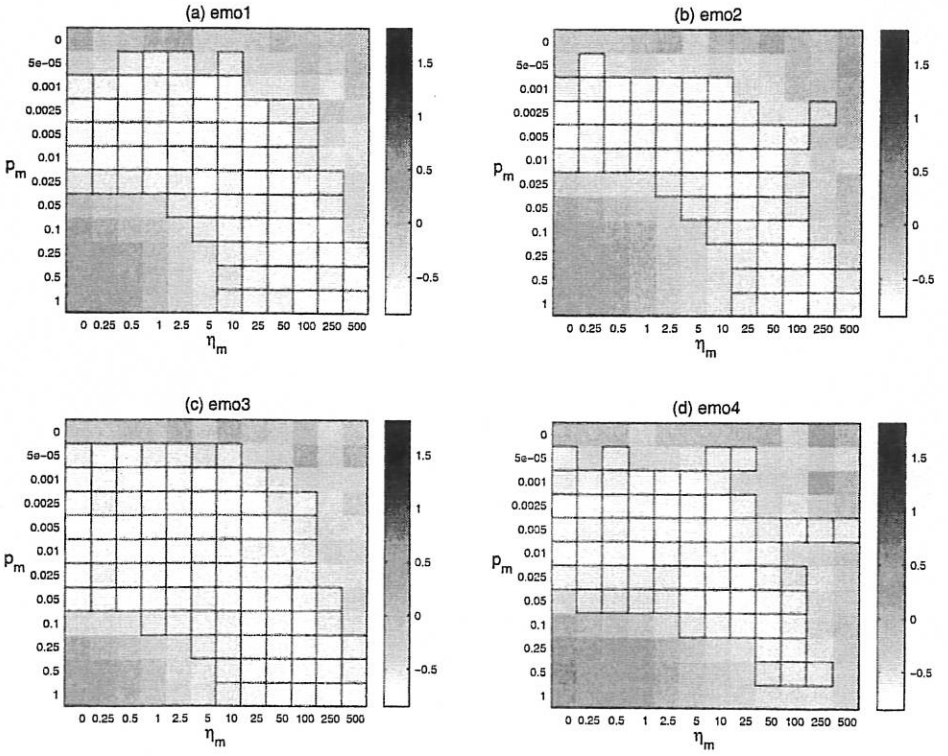

Figure 7.2: Mutation proximity maps for DTLZ2(6) with P1000.

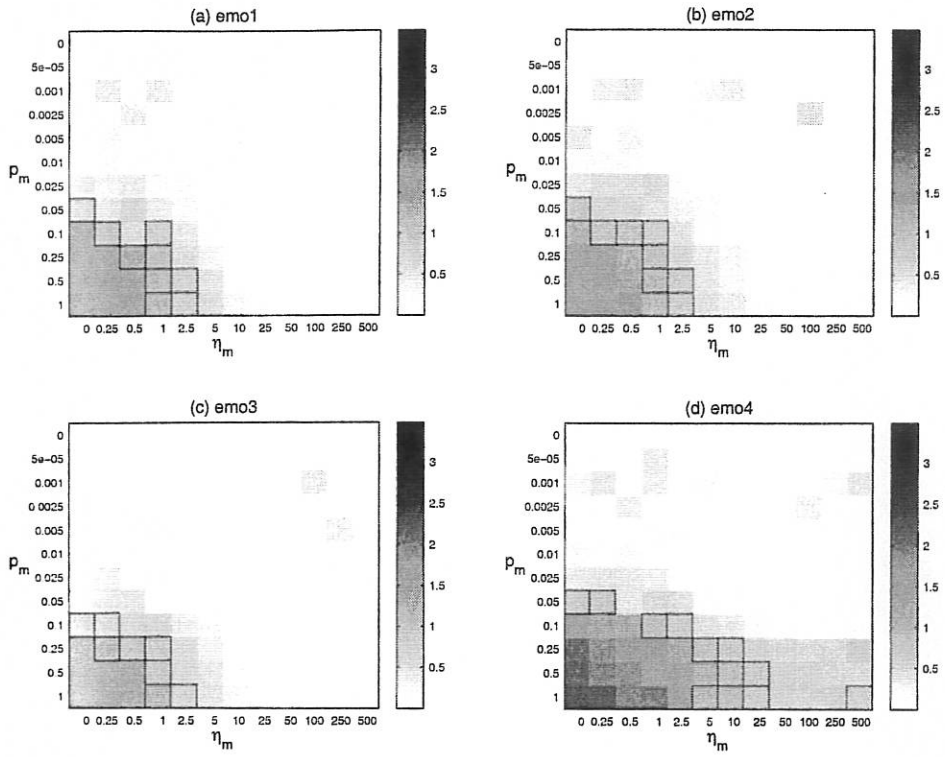

Figure 7.3: Mutation spread maps for DTLZ2(6) with P10.

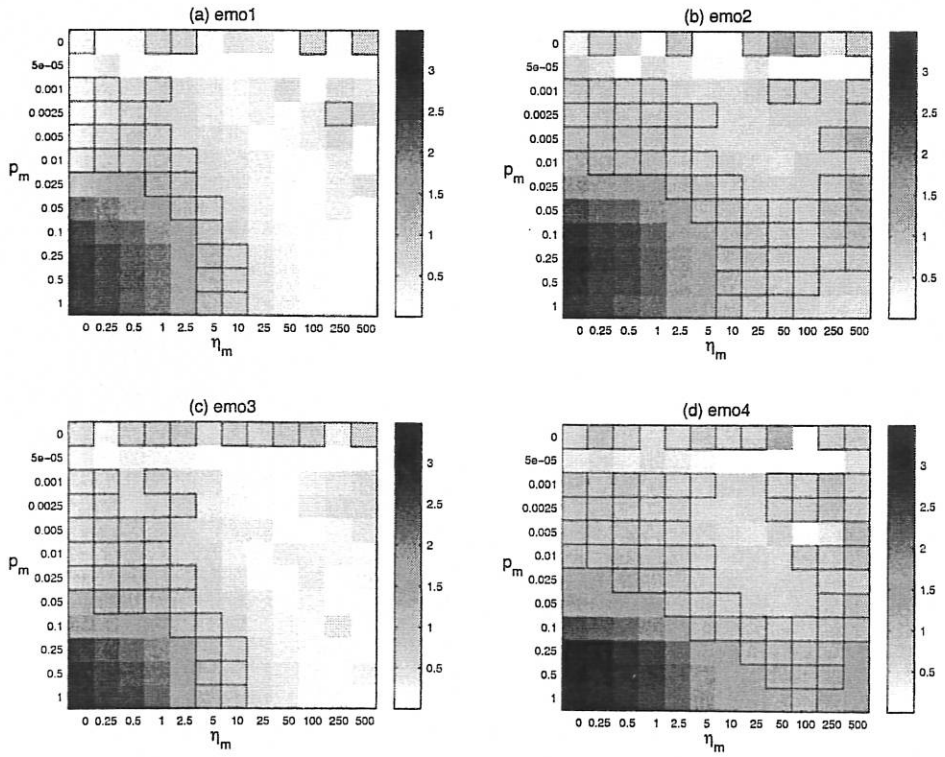

Figure 7.4: Mutation spread maps for DTLZ2(6) with P1000.

population sizes considered. This lack of effect can arguably be attributed to the fact that the mutation operator success rate (in terms of the child dominating the parent) is independent of the other population members, given that the parent has been selected for variation.

The associated spread response maps are provided in Figure 7.3 for P10 and in Figure 7.4 for P1000. The maps for P100 were previously shown in Figure 5.2. The spread values are generally seen to become larger with increasing population size. Also, the sweet-spot for spread becomes considerably larger in the diversity-promoting emo2 and emo4 algorithms for large population sizes. These results are likely to be reflective of the better quality sampling of objective-space (especially in terms of diversity) due to the large sample sizes used.

7.3 Recombination

The proximity response maps for the recombination-based emo algorithms with a population size of 10 solving the six-objective DTLZ2 task are shown in Figure 7.5. The corresponding maps for P1000 are shown in Figure 7.6, whilst the P100 maps were previously given in Figure 6.1. The proximity results are very poor for small population sizes and are seen to improve with increasing population size. Since the SBX operator relies on the dissimilarity between parents to drive variation, sufficient diversity in the mating pool is required for an effective search. The level of diversity will increase with increased population size. For small sample sizes, too little genetic material is available and the optimisation process will quickly stagnate. Note that the general location of the proximity sweet-spot in configuration-space remains similar as population size is increased. The sweet-spot is seen to swell in the direction of smaller variation magnitude.

DTLZ2(6) spread response maps for P10 and P1000 are provided in Figures 7.7 and 7.8 respectively. The maps for P100 are shown in Figure 6.2. As with the mutation results of the previous section, spread values are seen to increase from very low values to larger values across configuration-space. Spread values are very small for the small population size shown in Figure 7.7. This indicates convergence of the population on to a very small region of decision-space, in many cases a single point. Large regions of spread within 25% of the optimal value of unity are evident for the large population size shown in Figure 7.8. Again the good sample diversity provided by larger populations, that is required for exploration via the SBX operator, is argued to be responsible for the discrepancy in this aspect of performance.

7.4 Summary

Single-parent variation operators such as polynomial mutation do not require population diversity to advance the search. A multi-point parallel search is undertaken, where individual solutions compete with others to promote their genetic material into the next generation. Thus, configurations for an approximation set with good proximity to the global trade-off surface are largely independent of population size. However, since the number of search points is limited for small population sizes, it can prove difficult to achieve the desired spread. Note that there is some evidence to suggest that the classical polynomial mutation operator settings for NSGA-II (emo4), [$p_m = |\mathbf{x}|^{-1}$, $\eta_m = 20$], will be a good choice for achieving both good proximity and good spread for small population sizes.

Multi-parent variation operators, of which SBX recombination is an example, *do* require sufficient population diversity to advance the search. Since this diversity is unlikely to be available for small sample sizes, evolutionary optimisers using SBX require a larger population that the equivalent algorithm equipped with mutation to achieve good proximity values. Large population sizes are also required for good spread values. Under a fixed population size of 100 individuals, the classical settings for SBX, [$p_c = 1.0$, $p_{ic} = 0.5$, $\eta_c = 15$], were shown in Section 6 to become an inappropriate choice for optimisation problems with

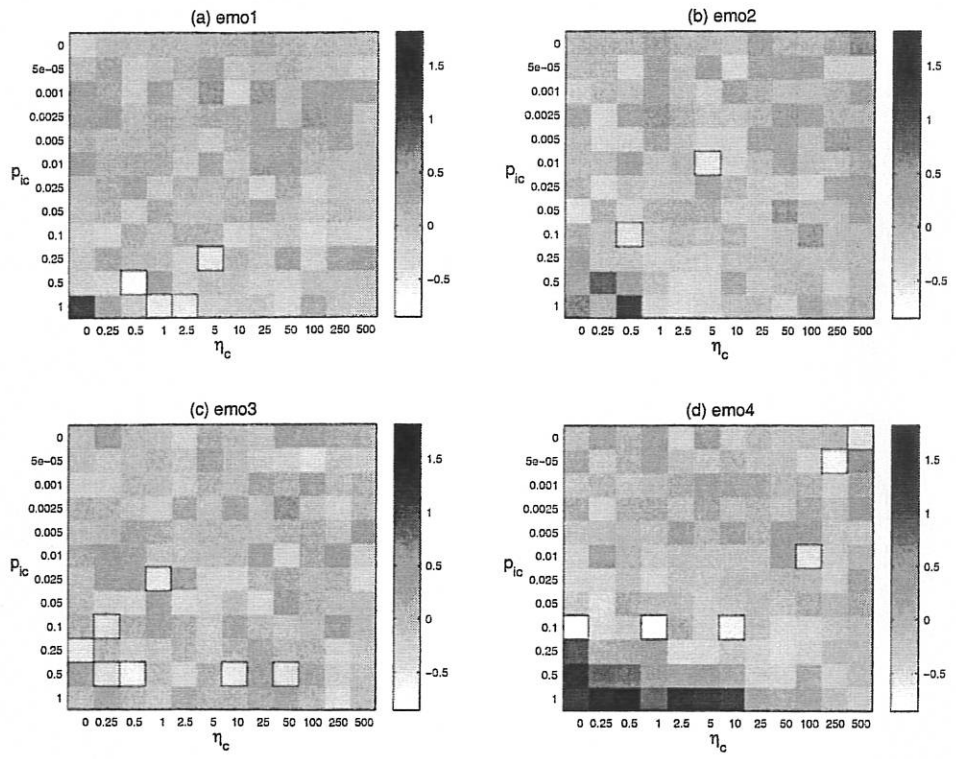

Figure 7.5: Recombination proximity maps for DTLZ2(6) with P10.

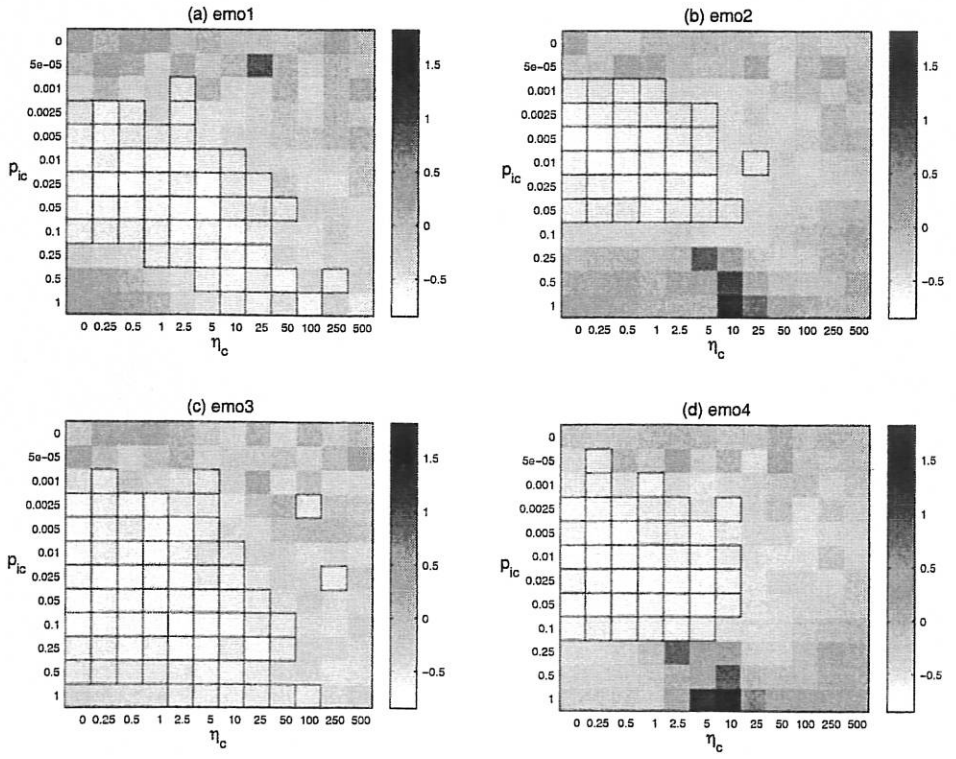

Figure 7.6: Recombination proximity maps for DTLZ2(6) with P1000.

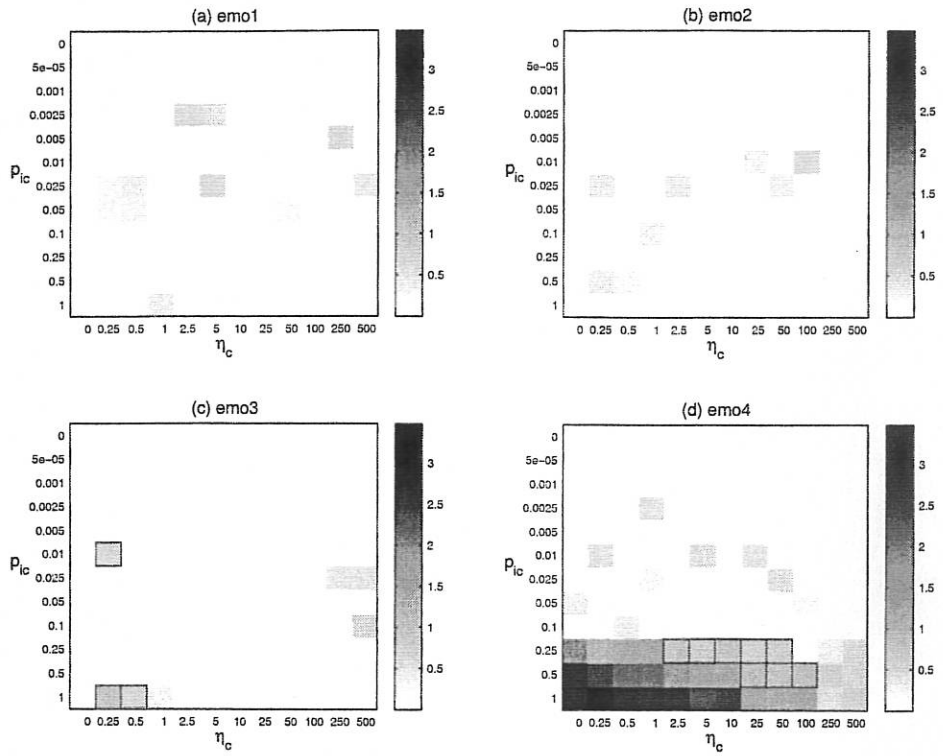

Figure 7.7: Recombination spread maps for DTLZ2(6) with P10.

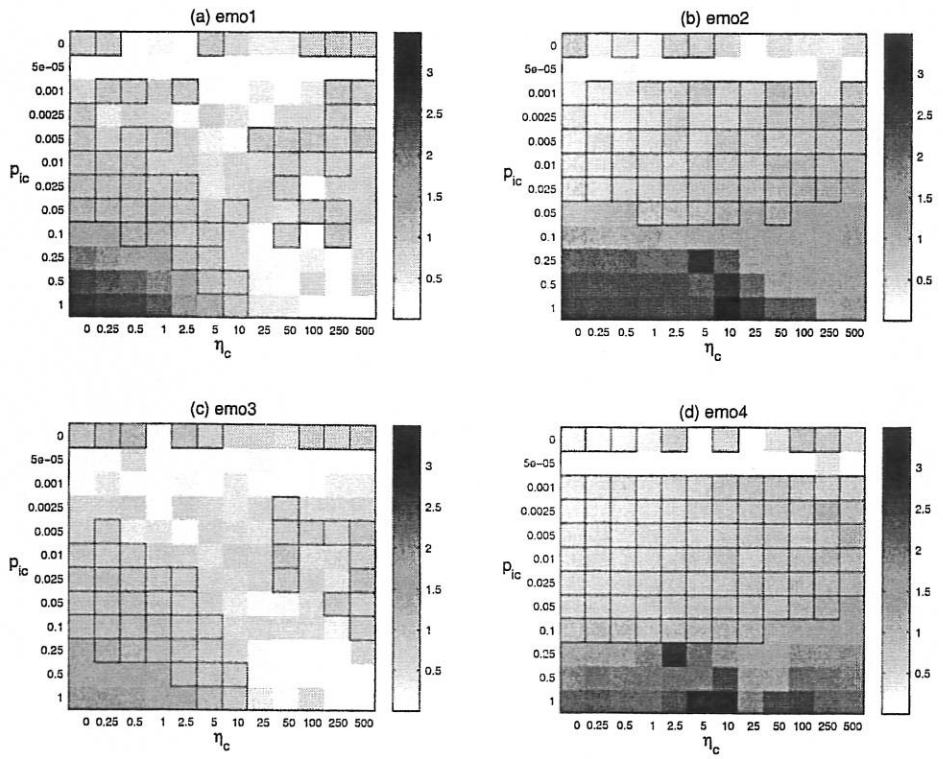

Figure 7.8: Recombination spread maps for DTLZ2(6) with P1000.

more than a small number of conflicting objectives. This remains the case, even for a population size of 1000.

8 Conclusion

8.1 Outcomes of the inquiry

The development of a comprehensive understanding of the many-objective optimisation process, where conflict is experienced between the objectives, is critical to both assessment of the suitability of modern evolutionary algorithms in these conditions and the development of new EMO methods.

This inquiry has considered a decomposition of the NSGA-II selection mechanisms in order to identify the algorithm components, and combinations of components, that are responsible for the observed behaviour. Optimiser responses have been generated for many locations in operator configuration-space. This has allowed regions of good performance (sweet-spots), and other levels of performance, to be identified for each optimiser. These results were obtained for the DTLZ2 optimisation task using an exploratory (rather than statistically robust) experimental method, and thus any extrapolation of the conclusions must be done with care.

Simulation results have been obtained for a set of selection mechanisms that are frequently used in the EMO domain, and for two popular variation operators for real-parameter optimisation. The analysis of these results, in the context of the selection and variation mechanisms that drive the evolutionary optimisation process, has produced the following key observations, hypotheses, and recommendations:

- All algorithms have reasonably large proximity sweet-spots for small numbers of conflicting objectives. Algorithms with diversity-promoting mechanisms produce larger spread sweet-spots. For both proximity and spread, superior performance is associated with the related elitist mechanisms.

- Configuration sweet-spots generally reduce in size as the number of conflicting objectives is increased. Configurations that can simultaneously achieve an approximation set with good proximity and good spread become increasingly limited. Note that optimal proximity, spread, and uniformity should be achievable regardless of population size. The limit on the size of the approximation set should merely affect the resolution of the obtained trade-off surface.

- Approximation sets with very poor proximity values (considerably worse than would be expected from a random sample of decision-space) can be generated by algorithms with diversity-promoting mechanisms when attempting to optimise many conflicting objectives. This behaviour is due to low variation operator success probabilities (in terms of children dominating their parents) coupled with a high percentage of unique non-dominated solutions in the population. This latter factor activates the mechanisms that bias towards remote solutions. In high-dimensional space, these solutions are likely to be poor in terms of proximity to the global surface. The SBX recombination operator exhibits extreme cases of this behaviour because the self-adaptive exploratory power can focus on poor DTLZ2 convergence-variable values, whilst retaining large expected perturbations of the distribution-variables.

- Classical mutation operator configuration settings remain a suitable choice for large numbers of conflicting objectives. The classical settings for recombination are found to

be unsuitable in these conditions. A low probability of application, such as that generally used for mutation, is recommended instead. This should be combined with a large expected magnitude of perturbation.

- The proximity sweet-spots obtained for mutation are largely invariant of population size, within the range of values usually considered for EMO. However, spread results can be improved by using a larger population size. Both aspects of performance for recombination can be achieved through adoption of a larger population.

The results of this inquiry suggest that the most predictable behaviour of an NSGA-type evolutionary optimiser solving a many-objective task can be obtained by deploying only a mutation operator. Further improvements to performance are possible through the use of recombination, but this operator also exhibits the more extreme poor performance if operator settings are chosen incorrectly. The regions of poor performance in configuration-space become more expansive as the number of conflicting objectives is increased.

If diversity-promoting mechanisms are not included in the optimiser then the risk of poor proximity in the approximation set can be avoided, but this is often at the expense of convergence to only a very small portion of the trade-off surface. Thus, it would be prudent to retain these mechanisms but also to monitor the operator success rates and activity of the diversity-based selection on-line. This feedback of the conditions that lead to optimiser difficulties can be used to adjust the search mechanisms during the optimisation process.

In summary, an evolutionary many-objective optimiser is required to produce an approximation to the global trade-off surface with good diversity in areas of good proximity. For this reason, diversity promotion is generally regarded as a secondary selection operator (Bosman and Thierens, 2003). However, since the primary convergence-based operator uses the relative concept of Pareto dominance, if the proportion of non-dominated – and dominance resistant – solutions is large then selection is based solely on diversity. As mentioned by Bosman and Thierens (2003), obtaining a good diversity is not a difficult task in itself, especially in many-objective space. Furthermore, the best diversity is often associated with very poor proximity values. Thus, a many-objective search based on the standard two-objective evolutionary components may actually evolve away from the global trade-off surface unless care is taken.

8.2 Proposals for future work

Research into the critical area of EMO scalability to optimisation tasks with large numbers of conflicting objectives is still in its infancy. Similar studies to this inquiry need to be conducted for other test problems, for other classes of evolutionary optimiser (such as hypergrid-based methods), and for other variation operators. The framework proposed in the inquiry could be extended to include multiple simulation replications for each optimiser configuration and advanced data-mining techniques to explore the massive amounts of generated information.

Deb and Goldberg (1989) discovered that the recombination of spatially-dissimilar parent solutions often produced low-performance children, known as *lethals*. The authors were able to improve EA performance by only allowing recombination to occur between parents located within the same local neighbourhood. This process is known as *mating restriction*. In the context of many-objective optimisation, lethals could be considered to be non-dominated, remote solutions with a highly substandard component in one or more objectives. An exploration of the effect of mating restriction schemes in the context of diversity-promoting mechanisms for many-objective optimisation may prove rewarding.

References

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.

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

COELLO, C.A.C. *et al*, 2002. *Evolutionary algorithms for solving multi-objective problems.* New York: Kluwer Academic Publishers.

CORNE, D.W. *et al*, 2000. The Pareto envelope-based selection algorithm for multiobjective optimization. *In:* M. SCHOENAUER *et al*, ed. *Proceedings of the 2000 parallel problem solving from nature conference (PPSN VI),* Paris 16-20 September 2000, *Springer-Verlag, 839-848.*

DEB, K., 2001. *Multi-objective optimization using evolutionary algorithms.* Chichester: Wiley.

DEB, K. and AGRAWAL, R.B., 1995. Simulated binary crossover for continuous search space. *Complex systems,* 9 (2), 115-148.

DEB, K *et al*, 2002a. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation*, 6 (2), 182-197.

DEB, K. *et al*, 2002b. Scalable multi-objective optimization test problems. *In:* D.B. FOGEL *et al*, ed. *Proceedings of the 2002 congress on evolutionary computation (CEC 2002),* Honolulu 12-17 May 2002. *IEEE Press, 825-830.*

DEB, K. *et al*, 2003. Towards a quick computation of well-spread Pareto-optimal solutions. *In:* C.M. FONSECA *et al*, ed. *Proceedings of the second international conference on evolutionary multi-criterion optimization (EMO 2003),* Faro 8-11 April 2003. *Springer-Verlag, 222-236.*

DEB, K. and GOEL, T., 2001. Controlled elitist non-dominated sorting genetic algorithms for better convergence. *In:* E. ZITZLER *et al*, ed. *Proceedings of the first international conference on evolutionary multi-criterion optimization (EMO 2001),* Zurich 7-9 March 2001. *Springer-Verlag, 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,* 1989. *Morgan Kaufmann, 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.* Kanpur: Indian Institute of Technology Kanpur, (KanGAL Report No. 2002004).

EDGEWORTH, F.Y., 1932. *Mathematical psychics. An essay on the application of mathematics to the moral sciences.* Reprint. London: The London School of Economics and Political Science.

EVERSON, R.M. *et al*, 2002. Full elite-sets for multi-objective optimisation. *In:* I.C. PARMEE, ed. *Proceedings of the fifth international conference on adaptive computing in design and manufacture (ACDM 2002),* Exeter 16-18 April 2002. *Springer-Verlag, 343-354.*

FARINA, M. and AMATO, P., 2002. On the optimal solution definition for many-criteria optimization problems. *In:* J. KELLER and O. NASRAOUI, ed. *Proceedings of the NAFIPS-FLINT international conference 2002,* New Orleans 27-29 June 2002. *IEEE Press, 233-238.*

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,* Urbana-Champaign June 1993. *Morgan Kaufmann, 416-423.*

GOLDBERG, D.E., 1989. *Genetic algorithms in search, optimization and machine learning.* Reading: Addison-Wesley.

GOLDBERG, D.E., 1998. *The race, the hurdle, and the sweet spot: lessons from genetic algorithms for the automation of design innovation and creativity.* Urbana: University of Illinois at Urbana-Champaign, (IlliGAL Report No. 98007).

IKEDA, K. *et al*, 2001. Failure of Pareto-based MOEAs: does non-dominated really mean near to optimal? *In:* J-H. KIM, ed. *Proceedings of the 2001 congress on evolutionary computation (CEC 2001),* Seoul 27-30 May 2001. *IEEE Press, 957-962.*

KHARE, V. *et al*, 2003. Performance scaling of multi-objective evolutionary algorithms. *In:* C.M. FONSECA *et al*, ed. *Proceedings of the second international conference on evolutionary multi-criterion optimization (EMO 2003),* Faro 8-11 April 2003. *Springer-Verlag, 376-390.*

KNOWLES, J.D., 2002. *Local-search and hybrid evolutionary algorithms for Pareto optimization.* Thesis (PhD). University of Reading.

KNOWLES, J.D. and CORNE, D.W., 2003a. Instance generators and test suites for the multiobjective quadratic assignment problem. *In:* C.M. FONSECA *et al*, ed. *Proceedings of the second international conference on evolutionary multi-criterion optimization (EMO 2003),* Faro 8-11 April 2003. *Springer-Verlag, 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.

LAUMANNS, M. *et al*, 2001. On the effects of archiving, elitism, and density based selection in evolutionary multi-objective optimization. *In:* E. ZITZLER *et al*, ed. *Proceedings of the first international conference on evolutionary multi-criterion optimization (EMO 2001),* Zurich 7-9 March 2001. *Springer-Verlag, 181-198.*

LAUMANNS, M. *et al*, 2002. Combining convergence and diversity in evolutionary multi-objective optimization. *Evolutionary computation,* 10 (3), 263-282.

MYERS, R.H. and MONTGOMERY, D.C., 2002. *Response surface methodology: process and product optimization using designed experiments.* 2nd ed. Chichester: Wiley.

PURSHOUSE, R.C. and FLEMING, P.J., 2002a. *Elitism, sharing, and ranking choices in evolutionary multi-criterion optimisation.* Sheffield: University of Sheffield, (ACSE Research Report No. 815).

PURSHOUSE, R.C. and FLEMING, P.J., 2002b. Why use elitism and sharing in a multi-objective genetic algorithm? *In:* W.B. LANGDON *et al*, ed. *Proceedings of the genetic and evolutionary computation conference (GECCO 2002),* New York City 9-13 July 2002. *Morgan Kaufmann, 520-527.*

PURSHOUSE, R.C. and FLEMING, P.J., 2003. Conflict, harmony, and independence: relationships in evolutionary multi-criterion optimisation. *In:* C.M. FONSECA *et al*, ed. *Proceedings of the second international conference on evolutionary multi-criterion optimization (EMO 2003),* Faro 8-11 April 2003. *Springer-Verlag, 16-30.*

RUDOLPH, G., 2001. Evolutionary search under partially ordered fitness sets. *In:* M.F. SEBAALY, ed. *Proceedings of the international symposium on information science innovations in engineering of natural and artificial intelligent systems (ISI 2001),* Dubai 17-21 March 2001. *ICSC Academic Press, 818-822.*

SMITH, R.E. *et al*, 1995. Fitness inheritance in genetic algorithms. *In:* K.M. GEORGE *et al*, ed. *Proceedings of the 1995 ACM symposium on applied computing,* Nashville 1995. *ACM Press, 345-350.*

THOMPSON, H.A. *et al*, 1999. Distributed aero-engine control systems architecture selection using multi-objective optimisation. *Control Engineering Practice*, 7(5), 655-664.

VELDHUIZEN, D.A.V., 1999. *Multiobjective evolutionary algorithms: classifications, analyses, and new innovations*. Thesis (PhD). US Air Force Institute of Technology.

ZITZLER, E., 1999. *Evolutionary algorithms for multiobjective optimization: methods and applications*. Thesis (PhD). Swiss Federal Institute of Technology.

ZITZLER, E. *et al*, 2000. Comparison of multiobjective evolutionary algorithms: empirical results. *Evolutionary computation*, 8 (2), 173-195.

ZITZLER, E. *et al*, 2001. SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization. *In:* K. GIANNAKOGLOU *et al,* ed. *EUROGEN 2001, Evolutionary methods for design, optimization and control with applications to industrial problems,* Athens 12-21 September 2001.

ZITZLER, E. *et al*, 2003. Performance assessment of multiobjective optimizers: an analysis and review. *IEEE transactions on evolutionary computation,* 7 (2), 117-132.

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.