



This is a repository copy of *On the evolutionary optimisation of many conflicting objectives*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/74688/>

Monograph:

Purshouse, R.C. and Fleming, P.J. (2003) *On the evolutionary optimisation of many conflicting objectives*. Research Report. ACSE Research Report no. 917 . Automatic Control and Systems Engineering, University of Sheffield

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

On the Evolutionary Optimisation of Many Conflicting Objectives

Robin C. Purshouse* *Member, IEEE*, and Peter J. Fleming
Department of Automatic Control and Systems Engineering
University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK.

Submitted October 2003

Abstract

This inquiry explores the effectiveness of a class of modern evolutionary algorithms, represented by Non-dominated Sorting Genetic Algorithm (NSGA) components, for solving optimisation tasks with many conflicting objectives. Optimiser behaviour is assessed for a grid of mutation and recombination operator configurations. Performance maps are obtained for the dual aims of proximity to, and distribution across, the optimal trade-off surface. Performance sweet-spots for both variation operators are observed to contract as the number of objectives is increased. Classical settings for recombination are shown to be suitable for small numbers of objectives but correspond to very poor performance for higher numbers of objectives, even when large population sizes are used. Explanations for this behaviour are offered via the concepts of dominance resistance and active diversity promotion.

Index terms: multi-objective optimisation, many objectives, dominance resistance, diversity promotion, density estimation.

*Robin Purshouse acknowledges the support of the UK Engineering and Physical Sciences Research Council.

I Introduction

Much of the research into multi-objective evolutionary 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 [1]. Hence, there is a clear need to extend evolutionary multi-objective optimisation (EMO) research into the realm of *many-objectives* [2]. Thus, the terminology *evolutionary many-objective optimisation* (EMO) is proposed herein to refer to EMO problems of increased scale.

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 [3]. 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 optimisation 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* [4] exists for each objective then performance comparisons between solutions can be based on the notion of *Pareto dominance*. For further details refer to [1]. In these conditions, the optimal solution to an M -objective task, in which all objectives conflict, is an $(M - 1)$ -dimensional hypersurface. The number of samples required to represent the surface at a fixed resolution is exponential in M . Even if such an *approximation set* [5] 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 in [1]. 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 in [6], the potential for an exclusively Pareto-based solution to the many-objective optimisation problem remains a matter of some interest. Indeed, if the resolution of the obtained approximation set is regarded as a function of some maximum limit imposed on the size of the set (such as the population size of an MOEA), then there is no *a priori* restriction

that prevents the achieved set from being 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 in [7] to stimulate research into many-objective optimisation. In the first known study of its kind, this test suite was used in [8] to investigate the scalability of three contemporary MOEAs: *NSGA-II* [9], *PESA* [10], and *SPEA2* [11]. The implementation of PESA was found to produce approximation sets with good overall proximity of solutions to the global Pareto front but with a poor distribution when solving for many conflicting objectives. By contrast, the approximation sets obtained for the implementations of NSGA-II and SPEA2 exhibited a good distribution of solutions but with poor proximity. However, since a single design-space instance of each algorithm was used, and each algorithm is itself a complicated structure of basic EMO components, it is not immediately clear from the study which components and processes are critical from the many-objective optimisation perspective. In particular, it is not clear why PESA should produce approximation sets of a fundamentally different character to NSGA-II and SPEA2.

In this paper, in accordance with the principles developed in [12], MOEA processes are considered at a component level rather than the composite approach adopted in [8]. The observed EMO behaviour can subsequently be explained in terms of fundamental search components and processes. In addition, results are generated for a *map* of variation operator configuration settings. This permits analysis to be made in terms of the *exploration-exploitation* (e-e) trade-offs in EMO [13] and for regions of high performance configurations (*sweet-spots*) to be identified [14].

This paper develops the initial analysis first reported in [15]. The class of MOEAs studied in this paper is introduced in Section II, for which the associated selection and variation processes are considered in detail. The empirical framework of the inquiry is developed in Section III, which draws on concepts from [12] and [16]. The optimisation task used in simulations is described at this stage, together with the indicators used to measure approximation set performance. The presentation of results is also considered. Simulation results for mutation-based algorithms are presented in Section IV. The results for recombination-based optimisers are included in Section V. The combined results from

these two sections are then analysed in Section VI, with respect to the underlying processes described in Section II. In [17] it is suggested that a potential technique for obtaining good EMO results is to increase the population size of an MOEA as an, ideally exponential, function of the number of objectives. This is unlikely to be a feasible approach in practice because of the computational resources demanded. However, the potential benefits of this technique remain a matter of interest and a population sizing study is thus explored in Section VII.

II Algorithms Considered

II.A Introduction

The search mechanism, often described in terms of exploration and exploitation, employed by the class of evolutionary optimiser considered in this inquiry can be summarised by Equation 1.

$$P[t + 1] = s_s(v(s_v(P[t])), P[t]) \quad (1)$$

$P[t]$ 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. s_v , known simply as the *selection* operator in genetic algorithm (GA) terminology, 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 s_s stage to determine the composition of the subsequent population. The s_s operator is known as the *reinsertion* operator in GA terminology.

The different multi-objective evolutionary optimisers proposed in the literature can generally be categorised according to the manner by which selection-for-variation and selection-for-survival are performed. A taxonomy of the various techniques is offered in [16]. The representation and variation operators are problem-specific but interact heavily with the selection operators (mainly in terms of the exploration-exploitation trade-off) to form the complete algorithm.

Each MOEA is comprised of a number of different selection mechanisms. It can often prove challenging to correctly determine which mechanisms and combinations of mechanisms are chiefly responsible for the observed algorithm performance. Thus deconstructed algorithms are used in the

study to determine the fundamental processes that are responsible for the observed many-objective behaviour of MOEAs.

NSGA-II components have specifically been chosen for this inquiry [9]. These components tend to be computationally and conceptually simple when compared to other algorithm processes in the literature. Also, the NSGA-II has been heavily studied by the EMO community. The *NSGA* [18] family of processes has strong similarities with other popular MOEAs, such as the *SPEA* [19] and *MOGA* [20] families.

NSGA-II, like many other MOEAs, uses Pareto dominance and density estimation considerations in its selection processes. These are discussed in detail in Section II.B to follow. Two algorithms are considered: standard NSGA-II and a decomposition of NSGA-II (denoted as D2 and D1 respectively).

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

1. The selection mechanisms operate on a global population model, for which s_s respects an arbitrarily chosen specific integer upper bound on the population size.
2. Selective bias is primarily based on Pareto dominance.
3. Selective bias *may* be secondarily based on density estimation.
4. If two solutions have equal properties from both a dominance and density perspective then selection should not favour one solution over the other.

Many of the selection mechanisms proposed in the literature share the above four properties. These include early algorithms, such as MOGA, NSGA, and *NPGA* [21], and also contemporary MOEAs such as SPEA, SPEA2, and the state-of-the-art MOGA described in [12].

An example of a methodology that is *not* represented by the processes considered in this chapter is the ϵ -dominance selection-for-survival concept developed in [22]. In this scheme, all locally non-dominated solutions can be retained (within the restrictions imposed by the threshold ϵ), and the population size is increased dynamically to account for this. The theoretical upper bound on population size is known but a specific size cannot be specified independently of the selection mechanism. Thus, property 1 listed above is different for this scheme.

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

The selection-for-survival method in the *PAES* [23] algorithm does not respect property 4. When a new locally non-dominated solution is considered for inclusion in $P[t+1]$, if the population size limit is reached then the solution can only be included if it has a density estimate that is lower than that of any solutions in the partially updated $P[t+1]$. If these latter solutions all have the minimum possible density estimate (each resides in a separate hyperbox) then the new solution cannot be included even if it is also unique to its associated hyperbox. Thus, despite being equal in terms of both dominance and density, the new solution is maximally biased against (the probability of selection is zero) when compared to the solutions in the partially updated $P[t+1]$. Note that the PAES method can be made to respect property 4 by changing the $<$ requirement on density to \leq , as was done in PESA.

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

II.B Selection Mechanisms

Algorithms D1 and D2 are introduced here. Recall that we are striving to study MOEA processes at a component level. This is why, for example, Algorithm D2 includes a diversity-preserving operator, while Algorithm D1 does not. We are seeking to isolate and understand the impact of this operator when applied to many-objective problems.

II.B.1 Algorithm D1

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

At the s_s stage of D1, the $P[t]$ population and $v(s_v(P[t]))$ population are combined. This new population is then ranked according to the non-dominated sorting method [18]. The new population, $P[t + 1]$ is then formed by including the best of the ranks until the population size limit is breached. Solutions that share the current rank under consideration in this situation are selected for inclusion on a random basis. This is equivalent to the $(\mu + \lambda)$ evolution strategy approach.

II.B.2 Algorithm D2

Algorithm D2 extends the selective capabilities of D1 by including supplementary density considerations in both s_v and s_s . The resulting algorithm is identical to the standard NSGA-II algorithm proposed in [9].

As in D1, binary tournament selection is used at the s_v stage. Of two solutions selected at random from the current population (with replacement), the selected solution is the one that dominates the other. If the solutions are non-dominated with respect to each other, then the solution with the smallest density estimate is selected. If the density estimates are also tied then a solution is chosen at random. In [9] this method is defined as the *crowded-comparison operator*.

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

Any density estimator can potentially be used to provide the density information required at the s_v and s_s stages of D2. This inquiry uses the *crowding distance* estimator proposed for use in the original NSGA-II [9]. In this estimator, density is calculated as the mean side length of the hypercube

formed using the first-nearest-neighbour (1-NN) values in each objective as vertices. In this inquiry, the boundary condition for an objective is set to the maximum non-boundary value calculated for that objective (rather than being set to infinity as in the original method proposed in [9]). This ensures that the estimator is unbiased for the equilibrium condition of a perfectly distributed approximation set.

Crowding distance is a low complexity estimator with limited accuracy. It should be noted that the relative effectiveness of the method when compared to other estimators has been questioned by some researchers [16, 24]. However it is argued that the behaviour of D2 when compared to D1 will be generally indicative of the effect of introducing any explicit diversity-promoting mechanism, *when such a mechanism is active*.

The activity — or otherwise — of the diversity promotion mechanisms is an important point. Some estimators, such as fitness sharing and simple hypergrid counting, operate within a defined local neighbourhood. If the neighbourhood is unsuitable for the current distribution of solutions then density information is lost. For example, in a hypergrid scheme, if all the solutions are contained within one hyperbox then they all have the same density estimate. This is also the case if each solution resides in its own hyperbox (regardless of the relative distances between occupied hyperboxes). This behaviour requires the careful selection of neighbourhood size: a task that has been automated in approaches such as that described in [25].

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

II.C Representation and Variation

This inquiry will simulate the many-objective performance of the selection mechanisms described in the previous sub-section using a real-parameter function optimisation problem (described later in Section III.B). Since the decision variables in the task are real-valued, the choice of a real-valued genotype representation follows naturally because such a representation is intuitively close to the problem space and also offers (claimed) performance advantages over a binary representation [26]. Thus each element of the genotype corresponds directly to a decision variable in the problem.

Two frequently studied variation operators for real representations are described in this section. The first is a one-parent mutation operator that produces a single child solution, whilst the second is a two-parent recombination operator that produces two children. Both operators have associated parameters that allow levels of exploration and exploitation to be controlled. The selection mechanisms in Section II.B can then be studied for a variety of exploration-exploitation trade-off settings. The interest is not primarily in which settings are good for the problem considered, but rather in how the different selection methods perform with respect to each other under various exploration-exploitation conditions when simultaneously optimising various numbers of conflicting objectives. The two operators are described in detail below.

II.C.1 Mutation

The *polynomial mutation* operator is used in this inquiry [27]. 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 [7, 8]. Variable-wise mutation is performed according to a probability distribution function centred over the parent value. The operator is defined in Equation 2, where b_i is the parent value for the i th decision variable, u_i and l_i are the upper and lower bounds on the i th 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 i th decision variable.

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

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

Mutation is applied independently to each element of each candidate solution with probability p_m . Thus, the probability of mutating a candidate solution of n decision variables is defined as shown in Equation 3.

$$p(\text{mutate}) = 1 - (1 - p_m)^n \quad (3)$$

II.C.2 Recombination

The *simulated binary crossover* (SBX) operator is also considered in the inquiry as an alternative to polynomial mutation [28]. 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 [7, 8], and is defined in Equation 4, where $b_{1,i}$ and $b_{2,i}$ are the parent values for the i th 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 i th decision variable.

$$\left. \begin{aligned} c_{1,i} &= 0.5 [(1 + \beta_i) b_{1,i} + (1 - \beta_i) b_{2,i}] \\ c_{2,i} &= 0.5 [(1 - \beta_i) b_{1,i} + (1 + \beta_i) b_{2,i}] \\ \beta_i &= \begin{cases} (2r_i)^{1/(\eta_c+1)} & \text{if } r_i < 0.5, \\ [1/(1 - r_i)]^{1/(\eta_c+1)} & \text{otherwise.} \end{cases} \end{aligned} \right\} \quad (4)$$

SBX generates child values from a probability distribution, with standard deviation derived from the distance between parent values and a distribution parameter η_c . The distance determines the overall magnitude of the distribution, whilst η_c determines the shape of the distribution. To generate child values for a decision variable, the distribution is centred over each parent and a random value is generated from the distribution to create one child. The second child is generated symmetrically to the first child about the mid-point between the parents. The child values for a decision variable are then exchanged between the complete child solutions with probability p_e . For simplicity $p_e = 0$ in the study. Note that child values that are generated outside the range of a decision variable are repaired 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 that variation is to be applied to the selected parents), then the probability of applying recombination to a pair of candidate solutions (each comprised of n decision variables) can be expressed as shown in Equation 5

$$p(\text{recombine}) = p_c [1 - (1 - p_{ic})^n] \quad (5)$$

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 3. This factor will assist in making comparisons between the results obtained using the two variation operators.

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

The DTLZ test suite encompasses many problem characteristics, such as multimodality, discontinuity, and distributional bias. For small numbers of objectives, such as three, it is a straightforward task to obtain a good sample-based approximation of the DTLZ2 global Pareto front using an EMO optimiser [7]. 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 IV and V, 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.

III.C Performance Indicators

In the context of this inquiry, performance is regarded as 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 [13]. Ideally, as illustrated in Figure 1, the optimiser should obtain solutions that are Pareto optimal (are of distance zero from the global front), that extend across the full DM region of interest (ROI), and that are as near uniformly distributed as the true surface permits. In this paper, the ROI is assumed to be the entire Pareto front.

Many performance indicators have been proposed for EMO and a discussion of these can be found in [29] and [30]. Following the approach of [31], a *functional* approach is taken in this paper, in which specific unary indicators are used to evaluate specific aspects of performance.

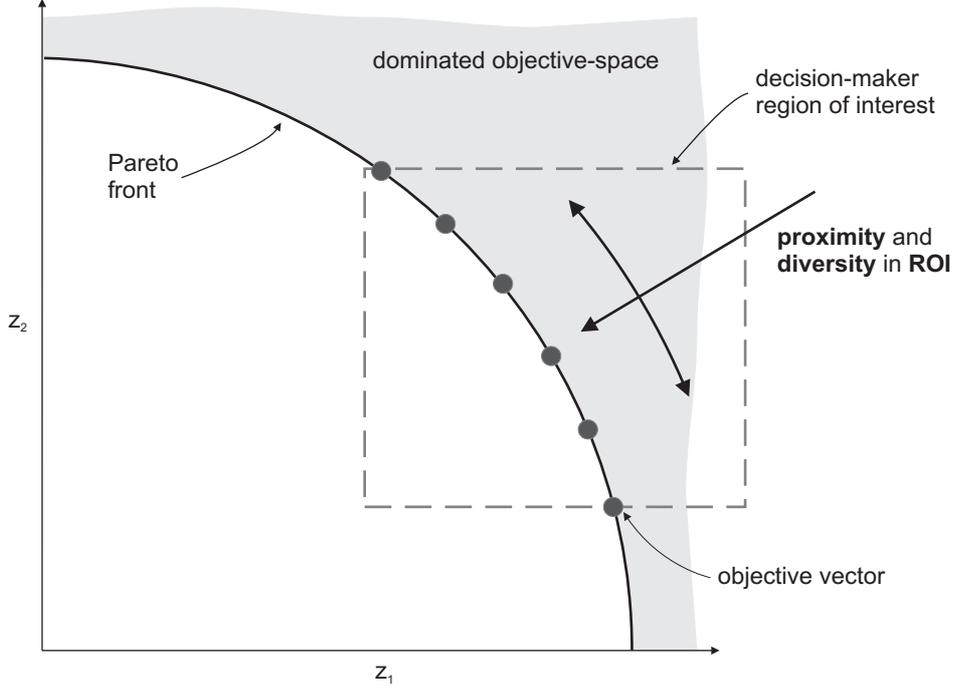


Figure 1: The ideal solution to a multi-objective optimisation problem

III.C.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, an objective vector for DTLZ2 will respect Equation 7. 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 8. This is essentially the same as the *generational distance* metric [32], for the case of a continuous globally optimal reference set, \mathbf{Z}_* .

$$1 \leq \left[\sum_{m=1}^M (z_m)^2 \right]^{1/2} \quad (7)$$

$$I_P = \text{median}_{\mathbf{z}_A \in \mathbf{Z}_A} \left\{ \left[\sum_{m=1}^M (z_{A_m})^2 \right]^{1/2} - 1 \right\} \quad (8)$$

III.C.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: the identified non-dominated vectors should span

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

To achieve high quantisation of the non-dominated set, it would be advantageous to express both aspects of distribution within a single indicator. This approach has been implemented in the Δ metric [9]. Using this approach it can be somewhat unclear which aspect of the distribution — extent or uniformity — is responsible for the observed indicator value. Thus in this study, in order to manage the complexity of the inquiry, only the extent of solutions is considered further.

Spread Indicator: The study uses a variant of the *maximum spread* indicator [33]. 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 9.

$$D = \left[\sum_{m=1}^M \left(\max_{z_A \in Z_A} \{z_{A_m}\} - \min_{z_A \in Z_A} \{z_{A_m}\} \right)^2 \right]^{1/2} \quad (9)$$

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 10. 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.

$$I_S = D / \left[\sum_{m=1}^M \left(\max_{z_* \in Z_*} \{z_{*m}\} - \min_{z_* \in Z_*} \{z_{*m}\} \right)^2 \right]^{1/2} \quad (10)$$

III.D Investigative Framework

The inherent high dimensionality of many-objective optimisation presents both conceptual and computational challenges to the analysis of algorithm behaviour. Thus, the inquiry framework is aimed toward exploratory data analysis rather than statistically significant performance comparison. Following the methodology in [16], single replication results are generated for a wide variety of configuration

instances (each representing a particular exploration-exploitation trade-off setting) to yield a response map in optimiser design-space. The use of multiple replications is still regarded as preferable, but this is computationally impractical for this inquiry. Note that spatial similarity between optimiser responses arguably provides some support for statistical confidence (or otherwise) in the observed behaviour. In [16], a polynomial model *response surface* is also fitted to the raw data map to obtain a smooth representation. However, for the purposes of this inquiry, the raw data itself proves sufficient to expose the spatial relationships in the data. For further details on response surfaces, refer to [34].

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

Responses for algorithms incorporating mutation are presented in Section IV. These are measured over 1000 generations and are generated for a variety of values of M with the population size fixed at 100. Similar results for recombination are presented in Section V. Results for both the variation operators are subsequently analysed in Section VI. In [17] it was suggested that a key method for coping with large M is to increase the population size, since this will tend to reduce the proportion of the population that is non-dominated and thus provide improved dominance-based discrimination. Whilst this approach is unlikely to be practical in many real-world applications, where the computational cost of evaluating a candidate solution may be very high, results for various population sizes are considered in Section VII.

III.E Introduction to Results

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

III.E.1 Types of Response Map

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

Proximity is measured by subtracting the initial proximity indicator value, computed as described in Equation 8, from the median of the values obtained from the final 100 iterations of the optimisation process. Thus, a proximity value of zero indicates no progress from the initial population, a negative value indicates convergence toward the global trade-off surface, and a positive value indicates divergence away from the true surface.

Spread is measured as the median value of the spread indicator, described in Equation 10, taken over the final 100 iterations of the optimiser. The optimal value of spread is unity. Values less than unity indicate approximation set convergence to a region of objective-space that is smaller than the true trade-off surface. Values larger than unity indicate that the approximation set extends over a greater region than that described by the optimal surface.

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

- **Copy rate.** The proportion of child solutions that are direct copies of (at least) one of their parents.
- **Resistance rate.** The proportion of child solutions that are non-dominated with respect to all

of their parents.

- **Success rate.** The proportion of child solutions that dominate at least one parent and are not copies of any parent.

III.E.2 Presentation of Results

An example presentation of response maps for the proximity indicator, I_P is shown in Figure 2. Part (a) of the figure, on the left, shows the response map for the algorithm D1 that does not include any specific diversity promotion mechanisms. Part (b) of the figure, on the right, shows the corresponding map for the algorithm D2 that does include such mechanisms. The results have been collected for algorithms that use recombination: the range of expected recombination magnitudes, η_c , is shown on the horizontal axis. The vertical axis shows the range of recombination application probabilities, p_{ic} . If the algorithm had been using mutation, the same ranges would be shown for η_m and p_m respectively. Note that variation magnitude is related to the inverse of the distribution parameter for both recombination and mutation, and thus discussions relating to this will generally refer to large values of $1/\eta_{c,m}$ rather than small values of $\eta_{c,m}$.

Performance for each $\{p_{ic}, \eta_c\}$ setting is indicated by a grey-scale square at the appropriate location. Lighter shades of grey indicate better proximity, as shown by the shaded-bar of indicator values to the right of each map. Considering Figure 2b, a range of good proximity is thus evident for intermediate values of p_{ic} in the range $[0.005 \ 0.1]$. Conversely, a region of poor proximity is evident in the region of intermediate $1/\eta_c$ coupled with high p_{ic} , such as $\{p_{ic} = 0.5, \eta_c = 5\}$.

The grid squares highlighted by a solid boundary correspond to configurations that exceed a pre-defined performance threshold. For proximity, the performance threshold is -0.5. The configuration $\{p_{ic} = 0.0025, \eta_c = 25\}$ is one such example in Figure 2a. For the spread response maps, performance is highlighted in the range $[0.75 \ 1.25]$ (within 25% of optimal).

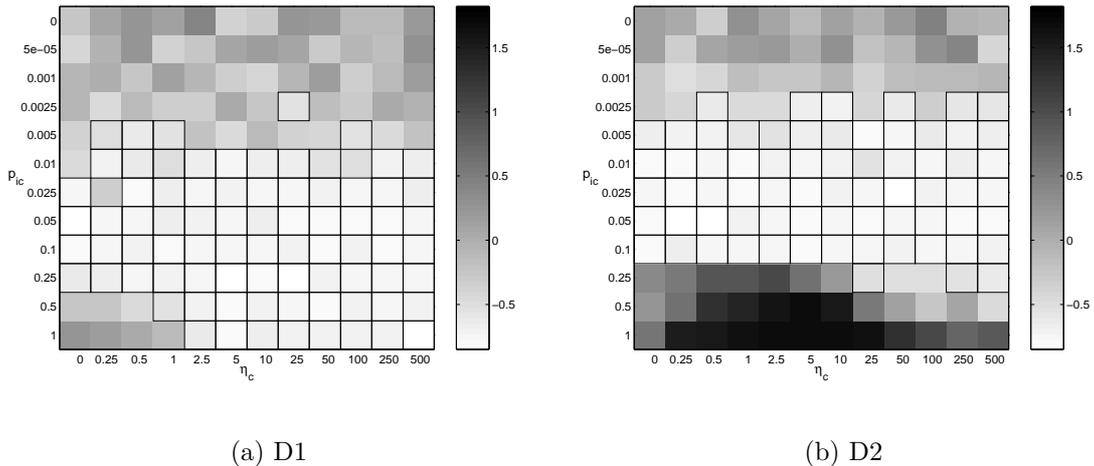


Figure 2: Example D-class proximity response maps

IV Results: Mutation

The three-objective proximity map for algorithm D1 is shown in Figure 3a. The corresponding results for algorithm D2 are shown in Figure 3b. Both algorithms produce approximation sets with good proximity for a large region of mutation configuration choices. However, in the region of high p_m and large $1/\eta_m$, the obtained proximity values are essentially unchanged from those obtained for the initial population. This behaviour is slightly more extensive for D2 than D1.

As the number of objectives is increased, the proximity sweet-spots contract for both D1 and D2. Proximity results for DTLZ2(6) are presented for D1 in Figure 3c, and for D2 in Figure 3d. Some reduction in the sweet-spot is observed for D1, particularly for high p_m and large $1/\eta_m$ configurations. The behaviour is more pronounced in the case of D2: the band of good proximity is noticeably thinner in regions of (i) low p_m coupled with small $1/\eta_m$ and (ii) high p_m coupled with large $1/\eta_m$. In these latter configurations, there is evidence that D2 is producing approximation sets with a worse proximity than that obtained for the initial population of the optimiser.

Further contraction in the proximity sweet-spot is observed as the number of conflicting objectives is continued to be increased. Results for a 12-objective instance of DTLZ2 are shown for D1 and D2 in Figure 3e and Figure 3f respectively. There is only a small reduction in the sweet-spot for D1, but the sweet-spot for D2 has continued to decrease substantially. For this latter algorithm, in regions of

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

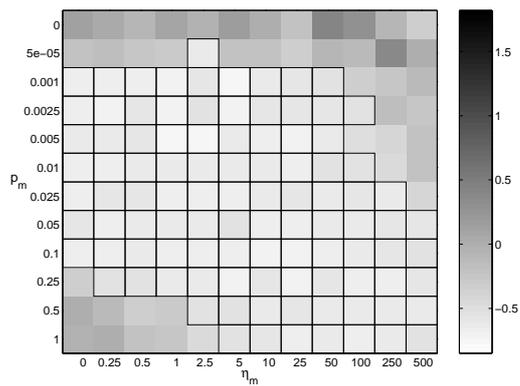
Spread results for the three-objective instance of DTLZ2 are shown for algorithm D1 in Figure 4a and for algorithm D2 in Figure 4b. The sweet-spot for D1 is limited to a band stretching from configurations of intermediate p_m combined with large $1/\eta_m$, such as $\{p_m = 0.05, \eta_m = 0\}$ to configurations of high p_m coupled with intermediate $1/\eta_m$, such as $\{p_m = 1, \eta_m = 10\}$. In the region where high p_m is combined with large $1/\eta_m$, spread values are larger than unity. This indicates that the approximation set is spread widely through objective-space. In regions of low p_m and small $1/\eta_m$, spread values are close to zero. This indicates that the approximation set represents a region of objective-space that is much smaller than optimal. The sweet-spot is considerably larger for D2, although the latter two observations for D1 also hold in this case. Good spread values are identified in larger regions at each end of the band previously identified for D1.

As the number of conflicting objectives is increased, the regions of good spread contract for both D1 and D2. Results for DTLZ2(6) are shown for D1 and D2 in Figure 4c and Figure 4d respectively. General thinning of the bands identified for DTLZ2(3) is evident for both algorithms. Large magnitudes of spread can be seen for D2 in regions of high p_m coupled with large $1/\eta_m$.

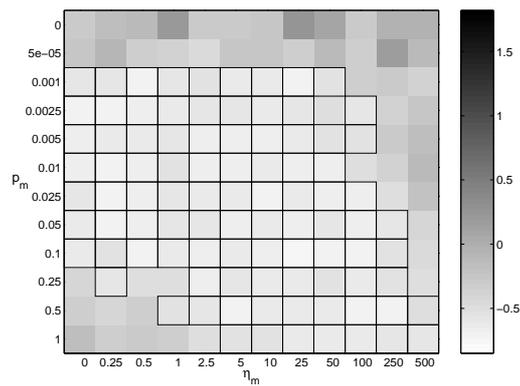
Spread results for the 12-objective instance of DTLZ2 are presented for D1 in Figure 4e and for D2 in Figure 4f. Further contraction of the sweet-spots is shown for both algorithms, but the level of deterioration is much lower between DTLZ2(6) and DTLZ2(12) than that observed between DTLZ2(3) and DTLZ2(6).

EMO algorithms are required to produce approximation sets with both good proximity and good diversity. From the results in Figure 3 and Figure 4 it is evident that mutation operator configurations that produce approximation sets of this quality can be identified for both D1 and D2 for all values of M considered. However, the number of such configurations reduces with increasing M for both D1 and D2.

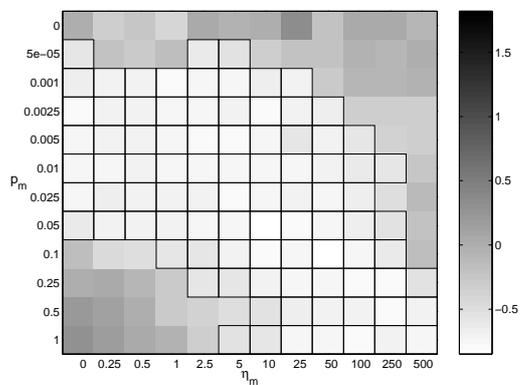
Good configurations for D1 are located in a different area of the map to those for D2. Configurations with relatively high p_m and relatively large $1/\eta_m$, such as $\{p_m = 0.25, \eta_m = 2.5\}$, produce approximation sets with both good proximity and good diversity for D1. However, these settings



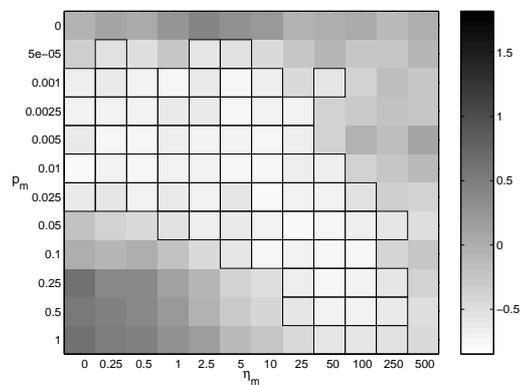
(a) D1: $M = 3$



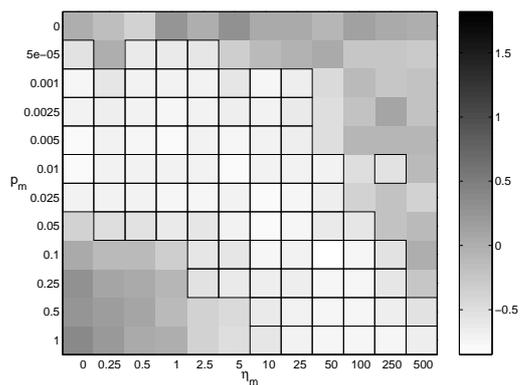
(b) D2: $M = 3$



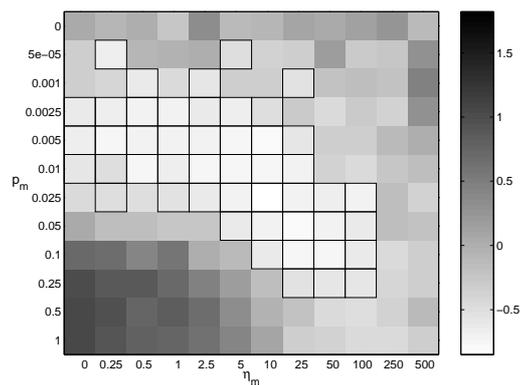
(c) D1: $M = 6$



(d) D2: $M = 6$

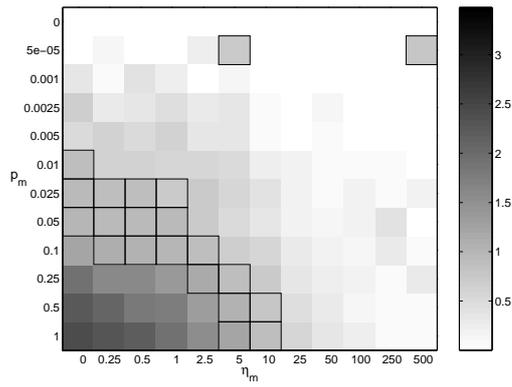


(e) D1: $M = 12$

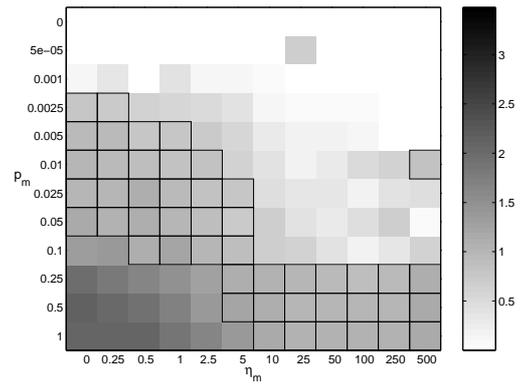


(f) D2: $M = 12$

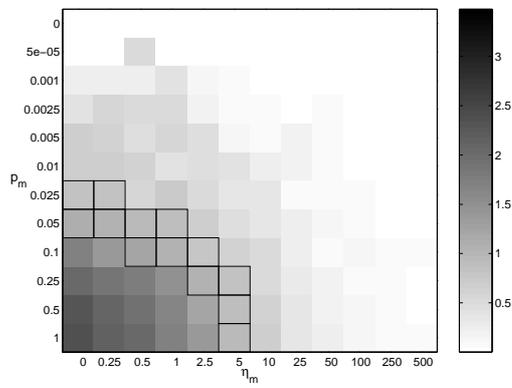
Figure 3: Proximity response maps for mutation



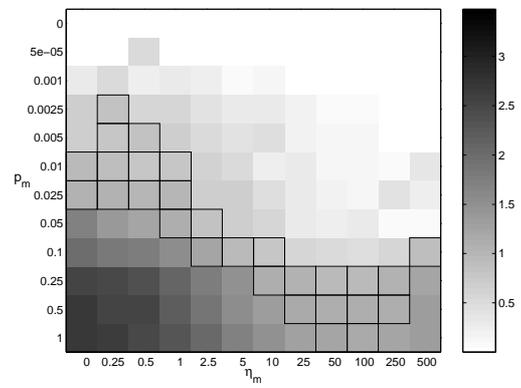
(a) D1: $M = 3$



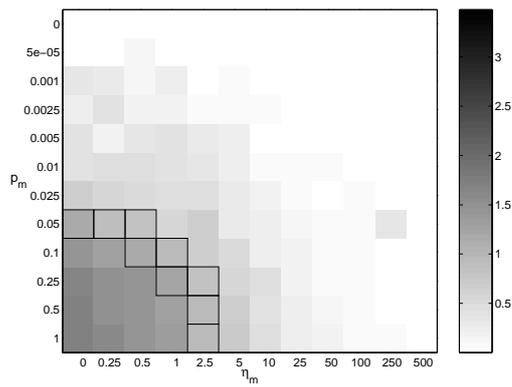
(b) D2: $M = 3$



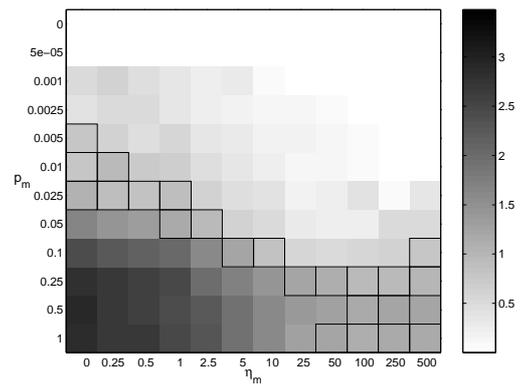
(c) D1: $M = 6$



(d) D2: $M = 6$



(e) D1: $M = 12$



(b) D2: $M = 12$

Figure 4: Spread response maps for mutation

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

Polynomial mutation has been used as the variation operator when solving DTLZ2 in previous studies in the literature. Both [7] and [8] used an algorithm identical to D2 with $p_m = 1/n$ and $\eta_m = 20$ (in combination with SBX recombination). From Figure 3 and Figure 4, it can be seen that this configuration produces approximation sets with good proximity, but only over a very small section of the trade-off surface, for both D1 and D2 when mutation is considered alone.

V Results: Recombination

The proximity response maps for the D1 and D2 algorithms incorporating the SBX recombination operator are shown in Figure 5. For the three-objective instance of DTLZ2, a substantial region of good proximity is evident for intermediate to high p_{ic} when combined with intermediate to large $1/\eta_c$. The map for D1 is provided in Figure 5a, whilst the equivalent for D2 is shown in Figure 5b. The sweet-spot for D2 is seen to extend further into regions of lower p_{ic} and smaller $1/\eta_c$ than that for D1.

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

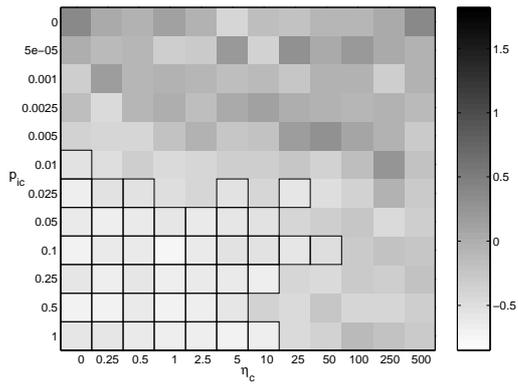
The D1 proximity response map for DTLZ2(12) is shown in Figure 5e, whilst the corresponding

map for D2 is depicted in Figure 5f. The sweet-spot for D1 has reduced still further to a region of intermediate p_{ic} . The D2 sweet-spot is slightly larger than that for D1 and is located in a region of slightly lower p_{ic} . In regions of high p_{ic} , the proximity results are very poor for D2. The region of divergence identified for DTLZ2(6) has expanded considerably to include expected variation magnitudes in the range from intermediate to high $1/\eta_c$.

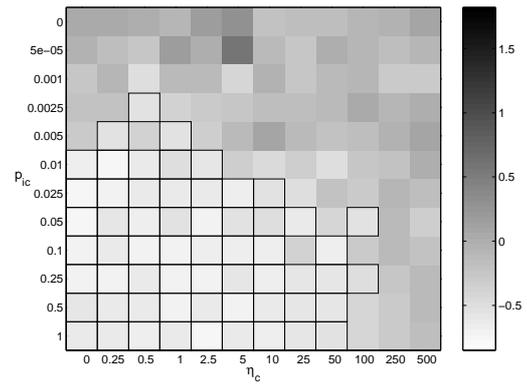
Spread response maps for algorithms D1 and D2 optimising the three-objective instance of DTLZ2 are shown in Figure 6. As indicated in Figure 6a, the spread sweet-spot for D1 encompasses configurations of intermediate to high p_{ic} coupled with intermediate to large $1/\eta_c$. The majority of other SBX configurations produce approximation sets with very small spread measures, thus indicating that the set represents only a very small region of objective-space. D2 offers a more substantial region of good spread performance, as shown in Figure 6b. The D2 sweet-spot extends further into configurations of lower p_{ic} and smaller $1/\eta_c$. However, for low values of p_{ic} , spread values are close to zero.

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

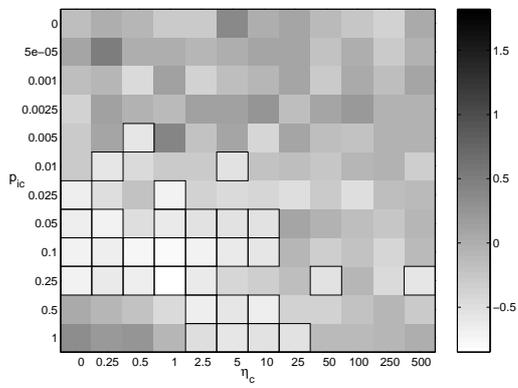
The spread response map for the 12-objective instance of DTLZ2 is shown for D1 in Figure 6e, and for D2 in Figure 6f. The sweet-spot for D1 is now highly limited, $\{p_{ic} = 0.25, \eta_c = 2.5\}$ being one such configuration. The band of good spread identified for D2 solving DTLZ2(6) has become thinner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



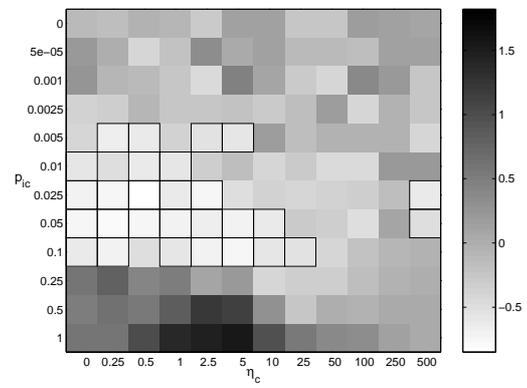
(a) D1: $M = 3$



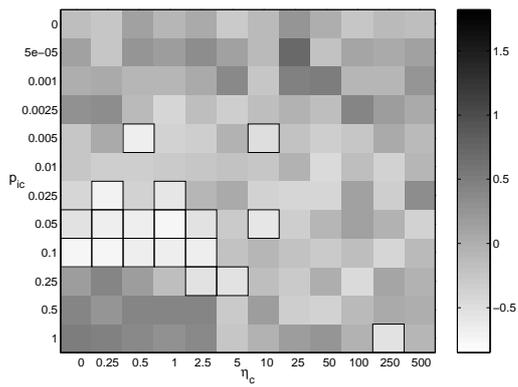
(b) D2: $M = 3$



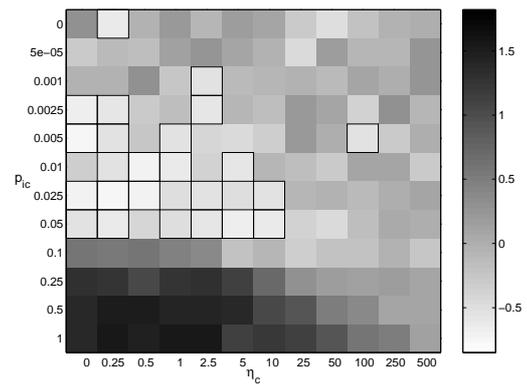
(c) D1: $M = 6$



(d) D2: $M = 6$

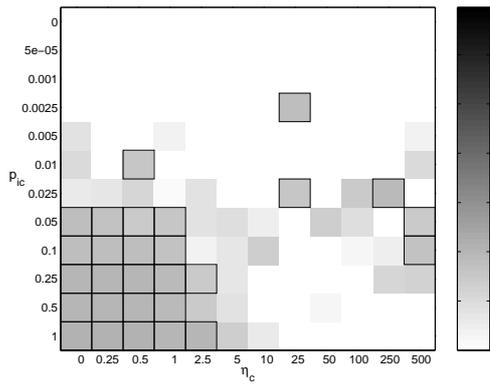


(e) D1: $M = 12$

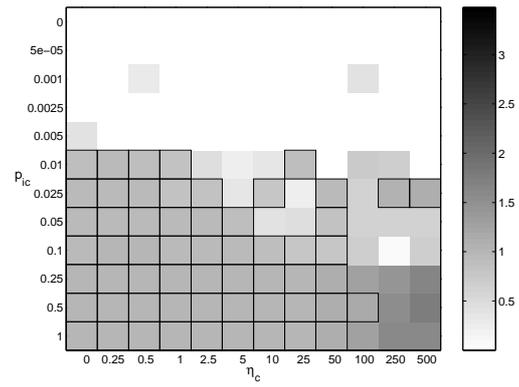


(f) D2: $M = 12$

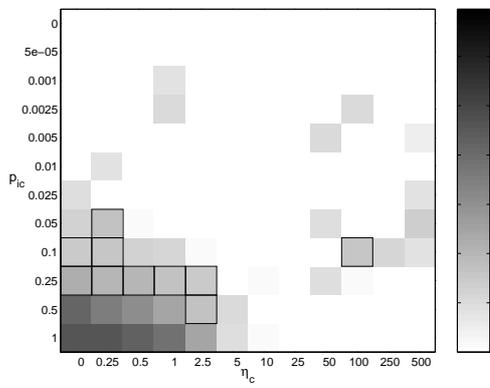
Figure 5: Proximity response maps for recombination



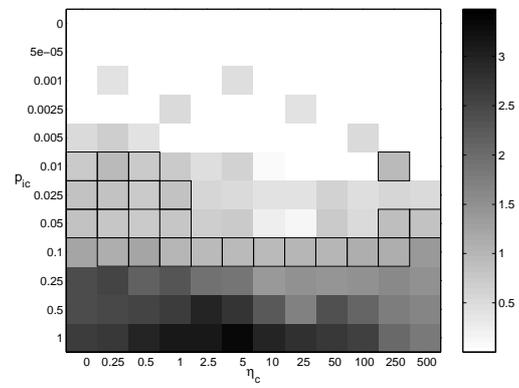
(a) D1: $M = 3$



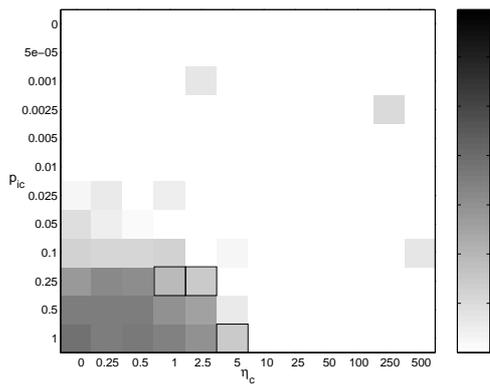
(b) D2: $M = 3$



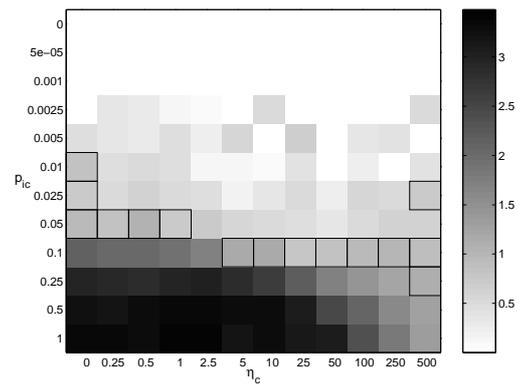
(c) D1: $M = 6$



(d) D2: $M = 6$



(e) D1: $M = 12$



(f) D2: $M = 12$

Figure 6: Spread response maps for recombination

Through considering Figure 5 and Figure 6 together, recombination operator configurations that produce approximation sets with both good proximity and good spread can be identified for both D1 and D2 for all values of M assessed. For $M = 3$, the sweet-spots for proximity and spread co-locate nicely for both algorithms. In particular, the evidence suggests that a wide range of configuration choices for D2 will lead to approximation sets that are good from the perspective of both axes of performance. As the number of conflicting objectives is increased, the selection of configurations that produce overall good performance decreases for both D1 and D2.

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

Like the polynomial mutation operator, the SBX recombination operator has also been previously used in the D2 algorithm on DTLZ2 [7, 8]. The configuration used in this work was $\{p_{ic} = 0.5, \eta_c = 15\}$ with $p_c = 1$ (the operator was also used in conjunction with polynomial mutation). From the results shown in Figure 5 and Figure 6 this configuration is demonstrated to be appropriate for the three-objective instance of DTLZ2 for D2. However, as the number of objectives is increased, the settings appear to become increasingly unacceptable. A D2 optimisation of DTLZ2(12) would be predicted to offer an approximation set with poor proximity, with solutions spread widely throughout objective-space.

VI Analysis

VI.A Introduction

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

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

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

VI.B Active Diversity Promotion Mechanisms

EMO researchers have previously observed that, for a fixed population size, the proportion of the population that is locally non-dominated increases rapidly with increasing M [35, 17]. Empirical results for DTLZ2, for a variety of population sizes and M -objective instances, are shown in Figure 7 (where each measurement is the mean of 1000 random sample sets). For a population size of 100, the proportion increases from approximately 0.3 for $M = 3$ to approximately 0.8 for $M = 12$. Note that this is only the *initial state* for the optimisers. Future proportions, as experienced by the optimisers, depend on the effect of repeated applications of s_v , v , and s_s .

D1 and D2 both discriminate in favour of non-dominance at the s_s stage. The effect of this is to quickly increase the proportion of non-dominated solutions in the population, as shown for various M for algorithm D2 in Figure 8. For all values of M the proportion is observed to rise rapidly to 1.0, regardless of the initial state. Thus, from the perspective of the selection-for-variation mechanism, the majority of solutions are equivalent for D1. s_v will be random in this case. However, for D2, the secondary diversity enhancement mechanism is now activated.

The operation of the density estimator used in D2 is illustrated by the schematic in Figure 9. The estimator assigns larger crowding distances (or, equivalently, smaller densities) to:

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

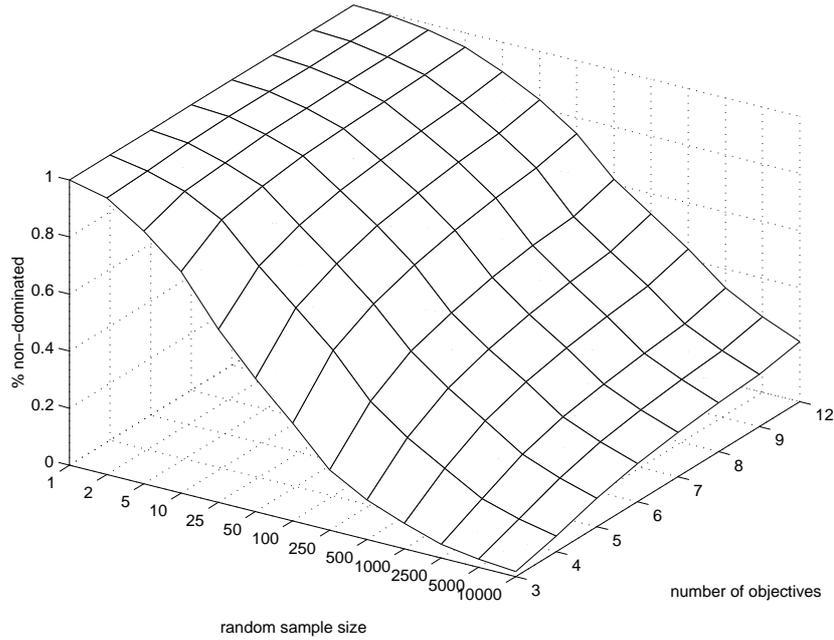


Figure 7: Estimated non-dominated proportion of random samples of DTLZ2(M) solutions

- immediate neighbours of remote solutions.

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

The above selective bias will present difficulties if the variation operators have a low expectation of success (defined in Section III.D as the probability of a child dominating its parent or parents). Success rates for both recombination- and mutation-based D2, for the three- and twelve-objective instances of DTLZ2, are shown in Figure 11. Note that the results for D1 are very similar and are thus not reproduced herein. The results for both variation operators show that success rates tend to decrease with increasing M . Therefore, one of the factors required to explain the difficulties encountered by diversity-promoting MOEAs is shown to become more apparent with increasing M .

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

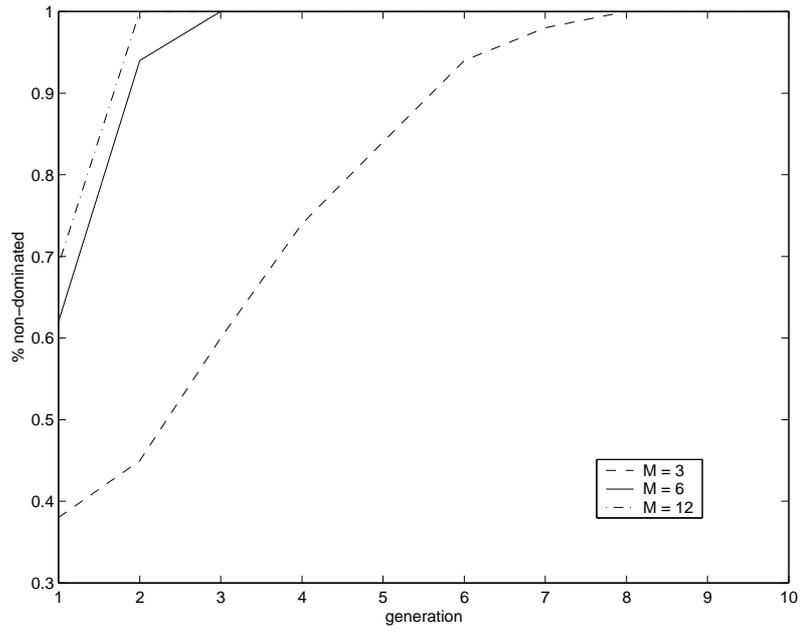


Figure 8: Example observed proportion of the population that is locally non-dominated during the initial stages of optimisation via mutation-based D2

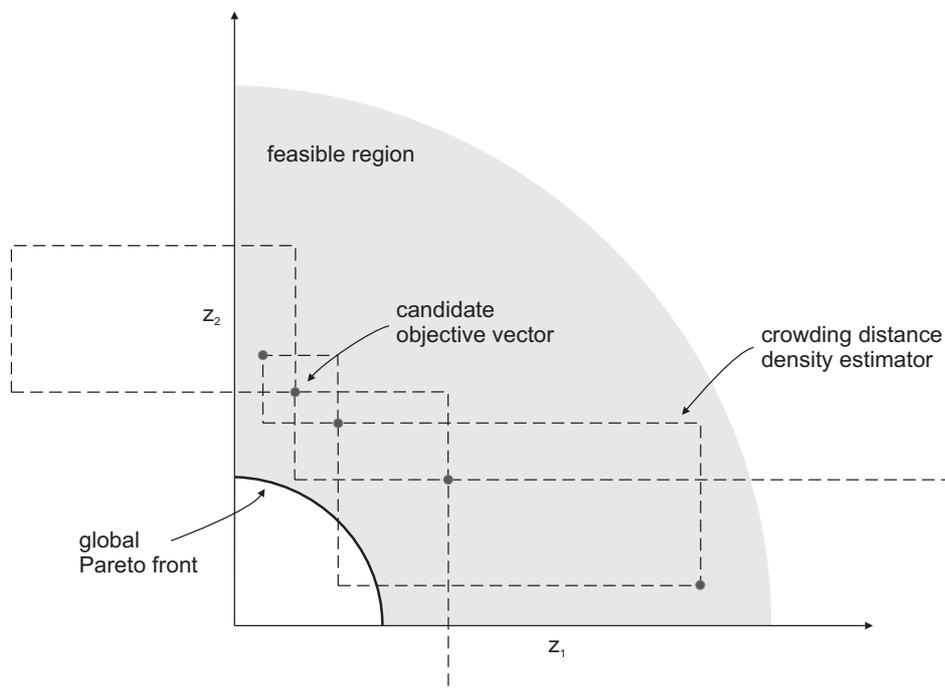
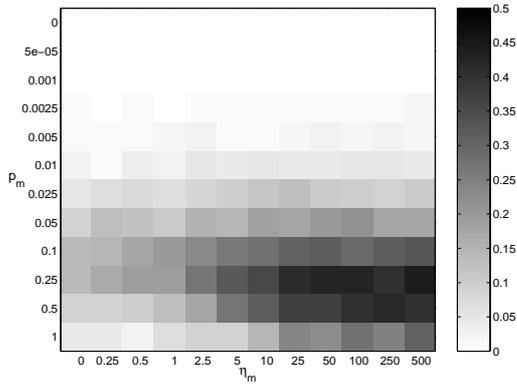
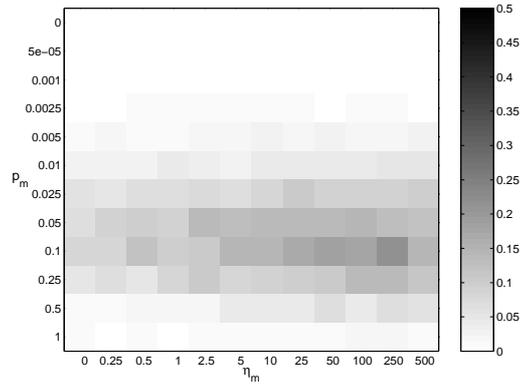


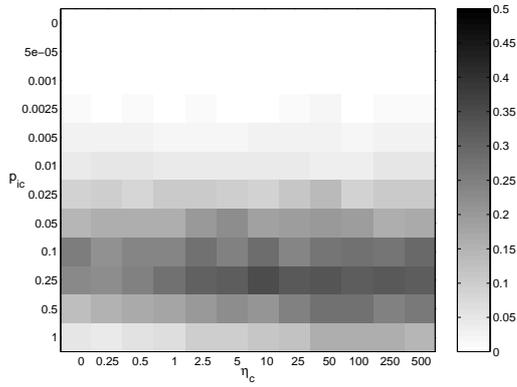
Figure 9: Crowding distance density estimator



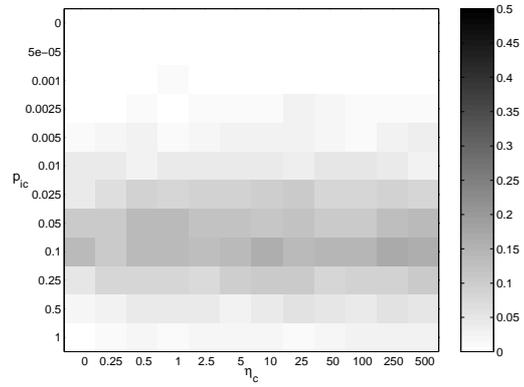
(a) Mutation: $M = 3$



(b) Mutation: $M = 12$



(c) Recombination: $M = 3$



(d) Recombination: $M = 12$

Figure 10: Proportion of children that dominate their parents for D2

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

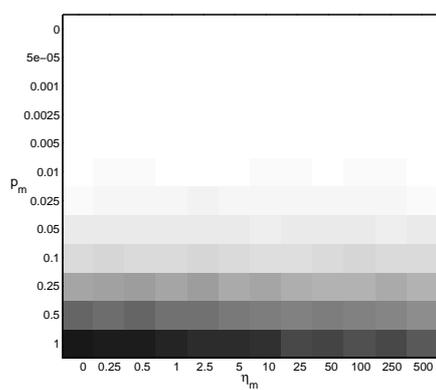
Dominance resistance was first identified, and introduced as terminology, in [36] in the context of real-parameter representations and associated variation operators for a specific class of multi-objective problems. This behaviour was also encountered in [7] for a set of real-parameter *constraint surface* tasks that are scalable in the number of conflicting objectives. These researchers also identified that the level of dominance resistance could increase with the dimension of objective-space.

Response maps that show the proportion of children that are non-dominated but are not equal to their parents are provided in Figure 11. For DTLZ2(3), a large proportion of such solutions is indicated for high operator application rates for both mutation and recombination (Figure 11a and Figure 11c respectively). As the number of objectives is increased, depicted in Figure 11b and Figure 11d, the high proportion of non-dominated children is observed to extend much further toward lower variation application probabilities. This empirically demonstrated dominance resistance supports the heuristic notion that a solution has more opportunity to be locally non-dominated as the number of dimensions of objective-space is increased.

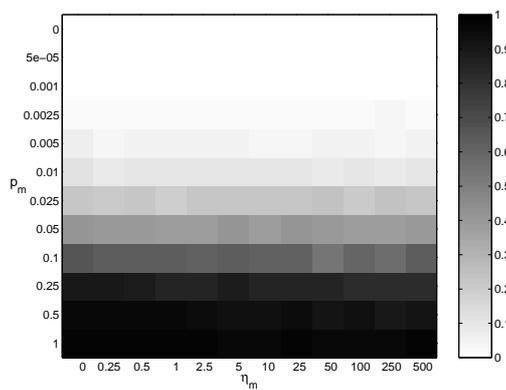
VI.C Inactive Diversity Promotion Mechanisms

For configurations corresponding to a low probability of variation operator application (the upper regions of the response maps presented in this paper), a large proportion of the children produced during variation are identical copies of their parents. Empirical results for D2, for both mutation and recombination, are shown in Figure 12. Note, again, that the results for D1 are very similar to these and are thus omitted.

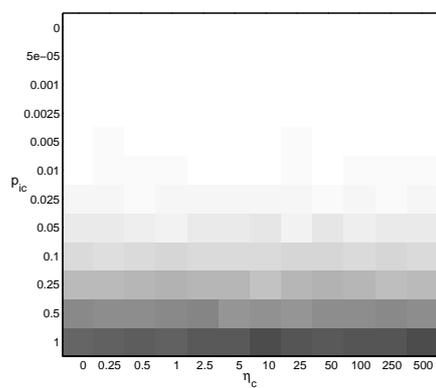
Since the density estimator used in D2 is a variant of first-nearest-neighbour, and equality in decision-space corresponds to equality in objective-space for the deterministic DTLZ2, solutions with



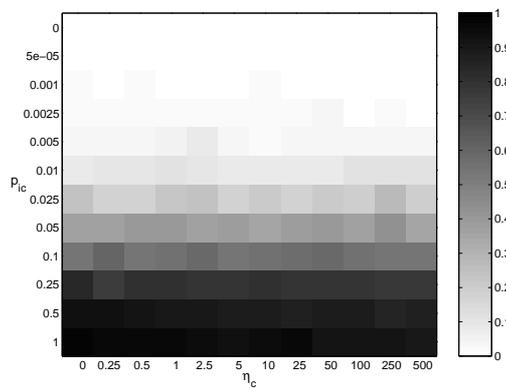
(a) Mutation: $M = 3$



(b) Mutation: $M = 12$

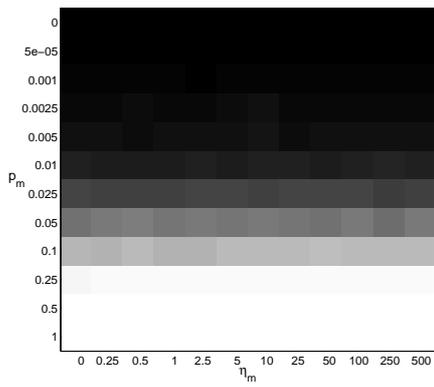


(c) Recombination: $M = 3$

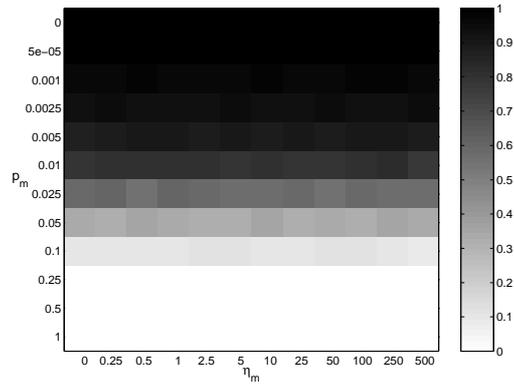


(d) Recombination: $M = 12$

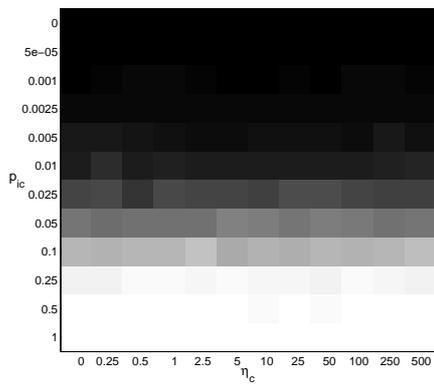
Figure 11: Proportion of children that are non-dominated (but not equal) to their parents for D2



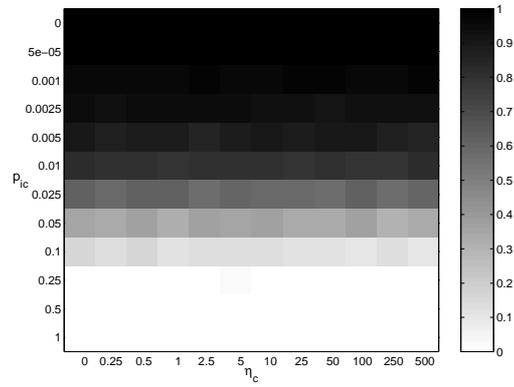
(a) Mutation: $M = 3$



(b) Mutation: $M = 12$



(c) Recombination: $M = 3$



(d) Recombination: $M = 12$

Figure 12: Proportion of children that are equal to their parents for D2

copies will have the maximum possible density estimate (corresponding to a crowding distance of zero). For configurations of low p_{ic} or p_m , this neutralises any density-dependent selection mechanisms because the densities of most solutions are identical. Thus, for the case of equivalence under dominance, selection is entirely random for both D1 and D2. Thus, little difference is evident between these two algorithms under the aforementioned configurations. This can be verified by comparing the upper region of the D1 response maps in Figures 3, 4, 5, and 6 to the corresponding D2 equivalents.

VI.D Comparison Between Mutation and Recombination

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

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

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

selection mechanisms in D2 also ensure that genetic diversity remains intact for the distribution variables of DTLZ2. Thus, large-scale exploration is still possible for these variables. So SBX is capable of simultaneously performing exploitation of poor proximity and exploration of diversity, leading to approximation sets spread widely in objective-space with very poor overall proximity.

VII Effect of Population Size

VII.A Introduction

In [17], the use of large population sizes is proposed as a potential method for achieving good many-objective optimisation results, since this will reduce the proportion of non-dominated solutions in the population and thus provide improved Pareto-based discrimination.

In practice, the use of large population sizes is often prohibitive in real-world applications because of the computational resources required to evaluate and process potential solutions. Thus in many applications, such as that documented in [37], the population size is generally limited to fewer than 50 individuals.

The often high cost of solution evaluation is a serious issue for computational search and optimisation techniques in general, rather than just affecting evolutionary algorithms. As a result, a research field known as *meta-modelling* has arisen that is devoted to the development and deployment of approximation models in solution evaluation [38]. A related topic that is specific to evolutionary algorithms, known as *fitness inheritance*, also considers how to reduce evaluation requirements [39]; this has been used in the context of MOEAs [40].

Given the above qualifications, the identification of the benefits — in terms of approximation set quality — that can be obtained for larger population sizes remains a matter of interest. In [41], the six-objective experiments in Section IV and Section V are repeated for population sizes 10, 100 and 1000 to produce new results for algorithms D1 and D2 using both mutation and recombination. An analysis of the observations arising from these experiments follows in the next section.

VII.B Analysis

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

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

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

The improved spread indicator results obtained in this study [41] can also be argued to be attributable to the richness of decision-space sampling inherent to larger population sizes. For a larger population, there is a greater probability of sampling any given area of the search space and of potentially making multiple samples from one region. Thus, the spread is likely to be naturally superior for a larger population, and the loss of one promising solution through sampling errors is less likely to

result in a region of interest becoming completely unrepresented. This argument applies to both variation operators considered in the inquiry. Also, as explained above, SBX requires diversity to facilitate exploratory behaviour. Thus, initial diversity is required in order to search for further diversity.

VIII Conclusions

VIII.A General Conclusions

This study has considered the behaviour of NSGA-II in many-objective optimisation through analysis of the selection and variation processes that underpin the overall methodology. NSGA was selected as it is a widely-used member of a family of algorithms that evaluate fitness using dominance comparisons and density estimation without partitioning objective-space. Thus, generalised observations of NSGA-II performance should be transferable to related algorithms, depending on the extent to which they share selection-for-variation and selection-for-survival mechanisms, and also how the exploration-exploitation trade-off within the variation operators relates to exploration-exploitation trade-off in selection.

It has been shown that the behaviour of MOEAs can change dramatically with the number of conflicting objectives to be optimised. In particular, the behaviour observed for an algorithm configuration for a small number of objectives cannot be generalised to an arbitrary (larger) number of objectives. Further, the sweet-spot corresponding to good quality approximation sets contracts as the number of conflicting objectives increases.

Through consideration of algorithm processes at a component level, the diversity-promoting selection mechanisms have been identified as highly influential to optimiser outcome. MOEA diversity promotion mechanisms can prove harmful for many-objective optimisation. In MOEA, the primary convergence-based operator uses the relative concept of Pareto dominance. If the number of non-dominated solutions is large, then selection is based solely on diversity, generally regarded as a secondary selection operator. However, obtaining a good diversity is not a difficult task in itself, especially in many-objective space and the best diversity is often associated with very poor proximity values. Thus, if the current solutions are dominance resistant, then the many-objective search may

evolve away from the true trade-off surface, with widespread dispersal of solutions in non-optimal objective-space.

Results obtained suggest that active diversity promotion can pose a serious challenge to obtaining an approximation set with good proximity to the global trade-off surface. But such a mechanism is positively required in EMO, since without it an EA will tend to experience genetic drift and thus converge on to a smaller region of objective-space than required. Diversity promotion is a crucial process in an MOEA, but its implementation requires some care.

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

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

VIII.B Directions for Future Work

The exploratory analysis of conflict in many-objective optimisation has produced some noteworthy findings of potential interest to the EMO research community. However, only a single real-parameter

function optimisation problem has been used during simulations. Thus, despite the realisation that the factors that are believed to underpin the observed behaviour could occur outside of this problem, the results cannot be generalised further at this stage. Similar studies are therefore required for other classes of many-objective problem, such as the *multi-objective quadratic assignment problem* class developed in [42] which encompasses both the *travelling salesman problem* and the *graph partitioning problem*. Analysis is also required of conflict scalability in real-world scenarios. Other classes of algorithm also require a detailed many-objective analysis. For example, the ϵ -dominance selection-for-survival mechanism developed in [22] is predicted to become an increasingly popular component of EMO algorithms and is thus a good candidate process for a study of this kind. The effect of alternative diversity promotion mechanisms in the context of fixed population sizes should also be considered.

The problem of dominance resistance has been identified as a key concern when optimising many conflicting objectives. Parallels exist here with the problem of *lethals* identified in single-objective multimodal function optimisation. In [43] it was discovered that the recombination of spatially dissimilar solutions tended to produce children that performed relatively badly in the problem domain. Improved search efficiency was obtained by allowing recombination to occur only between parents located within the same local neighbourhood. The neighbourhood size was calculated using the same methods as those used for fitness sharing. In many-objective optimisation, lethals can perhaps be regarded as locally non-dominated remote solutions with a highly substandard component in one or more objectives. Thus, the incorporation of some form of *mating restriction* may prove fruitful in a many-objective context. Indeed, such a methodology was suggested in the original MOGA [20].

IX Acknowledgements

The authors are very grateful to Salem Adra for his assistance in creating the revised version of this paper.

References

- [1] C. A. C. Coello, D. A. V. Veldhuizen, and G. B. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, New York, 2002.
- [2] M. Farina and P. Amato. On the optimal solution definition for many-criteria optimization problems. In J. Keller and O. Nasraoui, editors, *Proceedings of the 2002 NAFIPS-FLINT International Conference*, pages 233–238, Piscataway, New Jersey, 2002. IEEE Service Center.
- [3] R. C. Purshouse and P. J. Fleming. Conflict, harmony, and independence: Relationships in evolutionary multi-criterion optimisation. In C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, editors, *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, pages 16–30, Berlin, 2003. Springer.
- [4] F. Y. Edgeworth. *Mathematical Psychics. An Essay on the Application of Mathematics to the Moral Sciences*. The London School of Economics and Political Science, London, reprint edition, 1932.
- [5] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.
- [6] J. D. Knowles. *Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization*. PhD thesis, The University of Reading, Reading, UK, 2002.
- [7] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable multi-objective optimization test problems. In IEEE Neural Networks Council, editor, *Proceedings of the 2002 Congress on Evolutionary Computation (CEC 2002)*, volume 1, pages 825–830, Piscataway, New Jersey, 2002. IEEE Service Center.
- [8] V. Khare, X. Yao, and K. Deb. Performance scaling of multi-objective evolutionary algorithms. In C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, editors, *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, pages 376–390, Berlin, 2003. Springer.

- [9] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.
- [10] D. W. Corne, J. D. Knowles, and M. J. Oates. The Pareto envelope-based selection algorithm for multiobjective optimization. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H.-P. Schwefel, editors, *Proceedings of the Parallel Problem Solving from Nature VI Conference*, pages 839–848, Berlin, 2000. Springer-Verlag.
- [11] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the strength Pareto evolutionary algorithm. TIK Report 103, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 2001.
- [12] R. C. Purshouse and P. J. Fleming. Why use elitism and sharing in a multi-objective genetic algorithm? In W. B. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke, and N. Jonoska, editors, *Proceedings of the 2002 Genetic and Evolutionary Computation Conference (GECCO 2002)*, pages 520–527, San Francisco, California, 2002. Morgan Kaufmann Publishers.
- [13] P. A. N. Bosman and D. Thierens. The balance between proximity and diversity in multiobjective evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 7(2):174–188, 2003.
- [14] D. E. Goldberg. The race, the hurdle, and the sweet spot: Lessons from genetic algorithms for the automation of design innovation and creativity. IlliGAL Report 98007, University of Illinois at Urbana-Champaign, Urbana, Illinois, 1998.
- [15] R. C. Purshouse and P. J. Fleming. Evolutionary many-objective optimisation: An exploratory analysis. In *Proceedings of the 2003 Congress on Evolutionary Computation (CEC 2003)*, volume 3, pages 2066–2073, 2003.
- [16] M. Laumanns, E. Zitzler, and L. Thiele. On the effects of archiving, elitism, and density based selection in evolutionary multi-objective optimization. In E. Zitzler, K. Deb, L. Thiele, C. A. C.

Coello, and D. Corne, editors, *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, pages 181–196, Berlin, 2001. Springer-Verlag.

- [17] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, Chichester, UK, 2001.
- [18] N. Srinivas and K. Deb. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3):221–248, 1994.
- [19] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, 1999.
- [20] C. M. Fonseca and P. J. Fleming. Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. In S. Forrest, editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 416–423, San Mateo, California, 1993. Morgan Kaufman Publishers.
- [21] J. Horn and N. Nafpliotis. Multiobjective optimization using the niched Pareto genetic algorithm. IlliGAL Report 93005, University of Illinois at Urbana-Champaign, Urbana, Illinois, 1993.
- [22] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multi-objective optimization. *Evolutionary Computation*, 10(3):263–282, 2002.
- [23] J. D. Knowles and D. W. Corne. Approximating the nondominated front using the Pareto archived evolution strategy. *Evolutionary Computation*, 8(2):149–172, 2000.
- [24] K. Deb, M. Mohan, and S. Mishra. Towards a quick computation of well-spread Pareto-optimal solutions. In C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, editors, *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, pages 222–236, Berlin, 2003. Springer-Verlag.
- [25] C. M. Fonseca and P. J. Fleming. Multiobjective genetic algorithms made easy: Selection, sharing, and mating restriction. In A. M. S. Zalzala, editor, *Proceedings of the First International Confer-*

- ence on *Genetic Algorithms in Engineering Systems: Innovations and Applications (GALESIA 95)*, pages 42–52, Stevenage, UK, 1995. Institution of Electrical Engineers.
- [26] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, Berlin, 1992.
- [27] K. Deb and M. Goyal. A combined genetic adaptive search (GeneAS) for engineering design. *Computer Science and Informatics*, 26(4):30–45, 1996.
- [28] K. Deb and R. B. Agrawal. Simulated binary crossover for continuous search space. *Complex Systems*, 9(2):115–148, 1995.
- [29] J. D. Knowles and D. W. Corne. On metrics for comparing nondominated sets. In IEEE Neural Networks Council, editor, *Proceedings of the 2002 Congress on Evolutionary Computation (CEC 2002)*, volume 1, pages 711–716, Piscataway, New Jersey, 2002. IEEE Service Center.
- [30] C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, editors. *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, volume 2632 of *Lecture Notes in Computer Science*, Berlin, 2003. Springer-Verlag.
- [31] K. Deb and S. Jain. Running performance metrics for evolutionary multi-objective optimization. KanGAL Report 200200, Indian Institute of Technology Kanpur, Kanpur, 2002.
- [32] D. A. V. Veldhuizen. *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. PhD thesis, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, 1999.
- [33] E. Zitzler. *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. PhD thesis, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 1999.
- [34] R. H. Myers and D. C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley, Chichester, 2nd edition, 2002.
- [35] C. M. Fonseca and P. J. Fleming. Multiobjective optimization and multiple constraint handling with evolutionary algorithms — Part II: An application example. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 28(1):38–47, 1998.

- [36] K. Ikeda, H. Kita, and S. Kobayashi. Failure of Pareto-based MOEAs: Does non-dominated really mean near to optimal? In IEEE Neural Networks Council, editor, *Proceedings of the 2001 Congress on Evolutionary Computation (CEC 2001)*, volume 2, pages 957–962, Piscataway, New Jersey, 2001. IEEE Service Center.
- [37] H. A. Thompson, A. J. Chipperfield, P. J. Fleming, and C. Legge. Distributed aero-engine control systems architecture selection using multi-objective optimisation. *Control Engineering Practice*, 7(5):655–664, 1999.
- [38] L. Bull. On model-based evolutionary computing. *Soft Computing*, 3(2):183–190, 1999.
- [39] R. Smith, B. Dike, and S. Stegmann. Fitness inheritance in genetic algorithms. In K. M. George, editor, *Proceedings of the 1995 ACM Symposium on Applied Computing*, pages 345–350, New York, 1995. ACM Press.
- [40] J-H. Chen, D. E. Goldberg, S-Y. Ho, and K. Sastry. Fitness inheritance in multi-objective optimization. In W. B. Landon and E. Cant’ u-Paz and K. Mathias and R. Roy and D. Davis and R. Poli and K. Balakrishnan and V. Honavar and G. Rudolph and J. Wegener and L. Bull and M. A. Potter and A. C. Schultz and J. F. Miller and E. Burke and N. Jonoska, editor, *Proceedings of the 2002 Genetic and Evolutionary Computation Conference (GECCO 2002)*, volume 2, pages 319–326, San Francisco, California, 2002. Morgan Kaufmann.
- [41] R. C. Purshouse and P. J. Fleming. On the evolutionary optimisation of many conflicting objectives. Acse research report, University of Sheffield, Sheffield, UK.
- [42] J. D. Knowles and D. W. Corne. Instance generators and test suites for the multiobjective quadratic assignment problem. In C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, editors, *Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, pages 295–310, Berlin, 2003. Springer.
- [43] K. Deb and D. E. Goldberg. An investigation of niche and species formation in genetic function optimization. In J. D. Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms*, pages 42–50, San Mateo, California, 1989. Morgan Kaufmann Publishers.