

---

# Genetic Diversity as an Objective in Multi-Objective Evolutionary Algorithms

**Andrea Toffolo**

andrea.toffolo@unipd.it

Department of Mechanical Engineering, University of Padova  
Via Venezia 1, 35131 Padova, ITALY

**Ernesto Benini**

ernesto.benini@unipd.it

Department of Mechanical Engineering, University of Padova  
Via Venezia 1, 35131 Padova, ITALY

---

## Abstract

A key feature of an efficient and reliable multi-objective evolutionary algorithm is the ability to maintain genetic diversity within a population of solutions. In this paper, we present a new diversity-preserving mechanism, the Genetic Diversity Evaluation Method (GeDEM), which considers a distance-based measure of genetic diversity as a real objective in fitness assignment. This provides a dual selection pressure towards the exploitation of current non-dominated solutions and the exploration of the search space. We also introduce a new multi-objective evolutionary algorithm, the Genetic Diversity Evolutionary Algorithm (GDEA), strictly designed around GeDEM and then we compare it with other state-of-the-art algorithms on a well-established suite of test problems. Experimental results clearly indicate that the performance of GDEA is top-level.

## Keywords

Evolutionary algorithms, genetic diversity, Pareto optimality, multi-objective optimization, empirical comparison.

## 1 Introduction

Multi-objective optimization problems (MOOPs) often involve the simultaneous concurrence of conflicting objectives. This interaction results in a set of compromised solutions, that is known as the Pareto-optimal set. Since none of the solutions in this set can be considered as better than the others with respect to all the objectives, the goal of multi-criteria optimization is to find as many Pareto-optimal solutions as possible. Multi-objective evolutionary algorithms (MOEAs) are the most powerful tools for fulfilling this requirement due to their inherent parallelism and their ability to exploit the similarities among solutions by recombination. The search for widely different Pareto-optimal solutions is not however an easy task. Selection pressure may cause premature convergence of the population towards restricted portions of the Pareto-optimal set or even towards solutions that do not belong to it. Therefore, the maintenance of genetic diversity within the population is mandatory in order to find true and diverse Pareto-optimal solutions.

Until now, with the noticeable exception of the Strength Pareto approach (Zitzler and Thiele, 1999), the proposed MOEAs generally use two distinct methods to address convergence towards the Pareto-optimal set and to preserve genetic diversity. First, a technique is used to assign individuals a fitness value and, then, another technique

corrects the assignments to protect diversity, often introducing artificial parameters that need to be set somehow a priori in unknown fitness landscapes. In these algorithms, selection pressure is simply addressed towards the Pareto-optimal set, and its “greedy” effects are softened by the introduction of diversity-preserving mechanisms (e.g., fitness sharing).

The new diversity-preserving evaluation method presented in this paper, the Genetic Diversity Evaluation Method (GeDEM), introduces genetic diversity as a *real* objective in the optimization process. The challenge of GeDEM is to provide a dual selection pressure addressed towards both the Pareto-optimal set and the maintenance of diversity within the population. This is achieved by valuing the current non-dominated solutions as well as the most “diverse” ones, that is the most distant according to a certain metric. The former help exploitation in the most promising regions of the search space, the latter are necessary to keep the exploration of the search space alive. However, “greedy” convergence and maintenance of diversity (i.e., exploitation and exploration) are conflicting objectives of the search process, and thus *evaluation is performed considering the maximization of genetic diversity as one of the objectives of a two-criteria optimization problem, the other objective being the maximization of the ranks scored with respect to the objectives of the original MOOP.*

The paper is structured as follows. Section 2 presents the definitions of MOOPs and Pareto concepts. Then, Section 3 gives a brief analysis of the strategies used by reference MOEAs to maintain genetic diversity within the population. In Section 4, the GeDEM is introduced and described in detail. Section 5 provides an illustration of a multi-objective algorithmic framework, which is specifically designed around GeDEM, the Genetic Diversity Evolutionary Algorithm (GDEA). Finally, in Section 6 we present a systematic comparison between GDEA and other state-of-the-art MOEAs, following the guidelines by Zitzler et al. (2000), and then we describe our experimental results.

## 2 Multiobjective Optimization and Pareto Concepts

In general, a MOOP requires the minimization of  $m$  objective functions which depend on  $n$  decision variables and are subject to  $k$  constraints. Mathematically, an MOOP minimizes the components of a vector  $F(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x}))$ , where  $\mathbf{x}$  is an  $n$ -dimensional decision variable vector  $\mathbf{x} \in N$ , subject to  $g_i(\mathbf{x}) \leq 0, i = 1, \dots, k$ . The evaluation function  $F : N \rightarrow M$  maps vectors  $\mathbf{x} = (x_1, \dots, x_n)$  of the decision variable space  $N$  to vectors  $\mathbf{y} = (y_1, \dots, y_m)$  of the objective function space  $M$  (Figure 1).

The solution to a MOOP is a (possibly uncountable) set of decision variable vectors in  $N$ : the components of the corresponding vectors in  $M$  represent the best trade-offs in the objective function space.

For the reader’s convenience, the rest of the section is devoted to the mathematical definitions of the key concepts concerning Pareto optimality; examples of these concepts can be found elsewhere (Van Veldhuizen, 1999).

**Pareto Dominance:** A vector  $\mathbf{u} = (u_1, \dots, u_m)$  is said to dominate  $\mathbf{v} = (v_1, \dots, v_m)$  (denoted by  $\mathbf{u} \prec \mathbf{v}$ ) if and only if  $\forall i \in \{1, \dots, m\}, u_i \leq v_i \wedge \exists j \in \{1, \dots, m\} : u_j < v_j$ .  $\mathbf{u}$  is also said to *cover*  $\mathbf{v}$  ( $\mathbf{u} \preceq \mathbf{v}$ ) if and only if  $\mathbf{u} \prec \mathbf{v}$  or  $\mathbf{u} = \mathbf{v}$ .

**Pareto Optimality:** A solution  $\mathbf{x} \in N$  is said to be Pareto optimal with respect to the whole set  $N$  if and only if there is no other solution  $\mathbf{x}' \in N$  for which  $F(\mathbf{x}')$  dominates  $F(\mathbf{x})$ .

**Pareto Optimal Set:** For a given MOOP evaluation function  $F : N \rightarrow M$ , the Pareto optimal set ( $POS$ ) is defined as the subset of all the Pareto optimal vectors in the decision

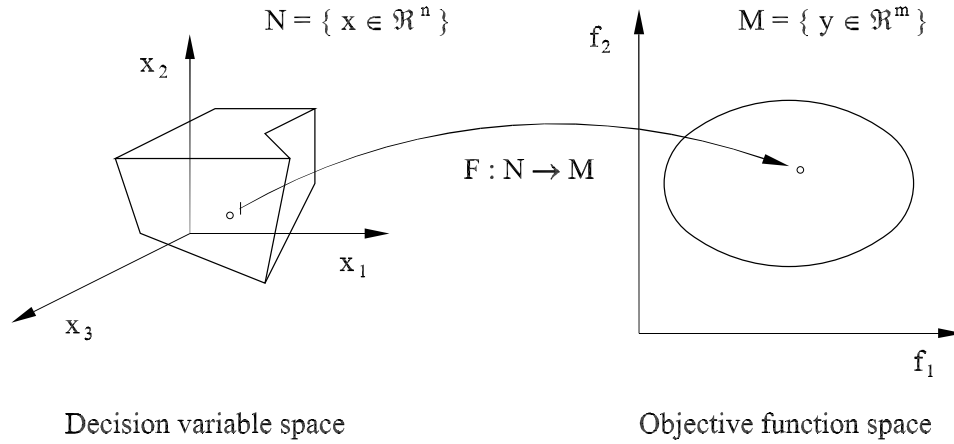


Figure 1: Description of an MOOP evaluation function

variable set:

$$POS := \{x \in N : \nexists x' \in N : F(x') \prec F(x)\} \quad (1)$$

**Pareto Front:** For a given MOOP evaluation function  $F : N \rightarrow M$  and Pareto optimal set  $POS$ , the Pareto front ( $PF$ ) is defined as the set of the vectors mapped from  $POS$  to  $M$  by  $F$ :

$$PF := \{y = F(x) = (f_1(x), \dots, f_m(x)) : x \in POS\} \quad (2)$$

### 3 Multiobjective Evolutionary Algorithms and the Problem of Diversity Preservation

The search process for the solutions to a MOOP has two main objectives:

1. drive the search towards the true Pareto-optimal set/front;
2. prevent premature convergence and distribute the solutions along the set/front itself.

Therefore MOEAs have to face two major issues in order to accomplish these tasks, that is, respectively:

1. how to perform fitness assignment and selection (using aggregating, nonPareto population-based or Pareto-based approaches);
2. how to maintain genetic diversity within the population of solutions.

Since a simple EA tends to converge towards a single solution and to lose some others during the search process (Mahfoud, 1995), the introduction of a method aiming at preserving diversity in the population is mandatory. These methods can be distinguished into niching techniques, which also promote the formation and maintenance of stable

sub-populations (niches), and non-niching techniques. Fitness sharing (Holland, 1975; Goldberg and Richardson, 1987) is the most frequently used niching technique. Each individual has its fitness divided by the number of individuals that are less distant than a specified parameter  $\sigma_{share}$  (the niche radius). Distance can be defined either in the decision variable space (genotypic or phenotypic) or in the objective function space (phenotypic only). The most common non-niching technique is restricted mating, in which the mating between two individuals is permitted only if their distance is less than a specified parameter  $\sigma_{mate}$  in order to avoid the formation of lethal individuals.

In the following, we briefly analyze the strategies used by reference MOEAs to promote diversity. For comprehensive overviews of evolutionary approaches to multi-objective optimization the reader is referred to the following more specific studies (Deb, 2001; Coello Coello et al., 2002).

- In the Vector Evaluated Genetic Algorithm (VEGA) (Schaffer, 1984; Schaffer, 1985), the search is guided in multiple directions by switching among the objectives during the selection phase. The mating pool is divided into equal-sized parts, each one being filled with individuals that are chosen according to a different objective.
- Hajela and Lin's Genetic Algorithm (HLGA) (Hajela and Lin, 1992) uses the weighted-sum method to assign fitness. The weights are variable, encoded in the genotype, and are evolved themselves to search for multiple solutions simultaneously. Diversity within weight combinations is maintained by phenotypic fitness sharing.
- Fonseca and Fleming's Genetic Algorithm (FFGA) (Fonseca and Fleming, 1993), the Niche Pareto Genetic Algorithm (NPGA) (Horn and Nafpliotis, 1993; Horn et al., 1994) and the Non-dominated Sorting Genetic Algorithm (NSGA) (Srinivas and Deb, 1994) are all Pareto-based EAs that use the concept of fitness sharing in different ways. FFGA assigns each solution a rank that equals the number of solutions that dominate it plus one. Phenotypic fitness sharing in the objective function space is then performed to uniformly distribute the solutions along the Pareto front. NPGA uses a niche count during tournament selection. Two individuals are picked at random and compared for domination with a restricted set of size  $t_{dom}$  (also picked at random at the beginning of each selection phase): if one is dominated and the other is not, then the non-dominated individual is selected for reproduction; otherwise, if both are either dominated or non-dominated, the one having the least individuals in its niche wins the tournament (phenotypic sharing in the objective function space is performed). NSGA applies, straightforwardly, the revolutionary non-dominated sorting procedure suggested by Goldberg (1989), adopting a ranking selection method to emphasize current non-dominated individuals and fitness sharing to maintain diversity within the population. Before entering the selection phase, the population is ranked according to Pareto dominance by iteratively extracting non-dominated sets from the unclassified solutions. Fitness sharing is also applied iteratively. The solutions belonging to the first non-dominated set are assigned the same dummy fitness value, giving them an equal reproductive potential and then phenotypic fitness sharing in the decision variable space is performed among the solutions of the set. The procedure is then repeated for the following non-dominated sets, assigning a dummy fitness value that is a little smaller than the worst shared fitness value observed among the solutions belonging to the previous sets. The shared fitness values are finally used in the selection phase.

- The Strength Pareto Evolutionary Algorithm (SPEA) (Zitzler and Thiele, 1999) is an elitist Pareto-based strategy that uses an external secondary population to store the non-dominated solutions found so far. The striking idea of this EA is that the Pareto-based fitness assignment method is itself a niching technique that does not require the concept of distance. The fitness (strength) of a non-dominated solution stored in the external population is proportional to the number of individuals covered, while the fitness of a dominated individual is calculated by summing the strength of the non-dominated solutions that cover it. This fitness assignment criterion results in the definition of a niche that can be identified with the portion of the objective function space covered by a non-dominated solution. Both the population and the external non-dominated set participate in the selection phase (the smaller the fitness, the higher the reproduction probability). The secondary population is updated every generation and pruned by clustering if the number of the non-dominated individuals exceeds a predefined size.

## 4 The Genetic Diversity Evaluation Method

### 4.1 Genetic Diversity as an Objective

As we pointed out at the beginning of Section 3, the multi-objective optimization process has two objectives: the convergence to the Pareto-optimal set and the maintenance of genetic diversity within the population. The basic idea of GeDEM is to actually use these objectives during the evaluation phase and to rank the solutions with respect to them, emphasizing the non-dominated solutions as well as the most genetically different. This results in a selection pressure driving the search simultaneously towards Pareto-optimal and diverse solutions, or, from an equivalent point of view, towards the exploitation of the current non-dominated solutions and to the exploration of the search space.

In fact, an individual's worth is not considered to rely only upon its objective function values, but also on its ability, when mating with other individuals, to generate offspring that are dispersed enough in the decision variable space, in order to explore the latter efficiently and exhaustively. The individuals having a fairly different genotype from the rest of the population have this ability, and this can be measured by a distance metric. An individual that is distant from all the others has more chances, when mating, to produce offspring in regions of the search space not covered by the current population. Common distance metrics that can be used to express this concept mathematically are the (normalized) Euclidean distance in the decision variable or objective function space and the Hamming distance in the string space (i.e., the same distance metrics used in fitness sharing). Possible measures of the diversity of an individual include the sum of its distances from all the other individuals, or the minimum distance from another individual.

Since the two objectives of the optimization process are themselves conflicting, GeDEM actually performs a two-objective non-dominated sorting procedure to assign fitness. The ranks of the solutions are determined maximizing (i) the ranks scored with respect to the objectives of the original MOOP, the non-dominated solutions having the highest rank, and (ii) the values assigned to each individual as a measure of its genetic diversity, calculated according to the chosen distance metric. Figure 2 illustrates a typical population and how GeDEM ranks the solutions. The individuals that help exploitation and exploration are shown, while the true "lethals" are the individuals having a poor rank and some neighbors around them.

The calculation of reciprocal distances among the individuals requires  $\mathcal{O}(n^2)$  oper-

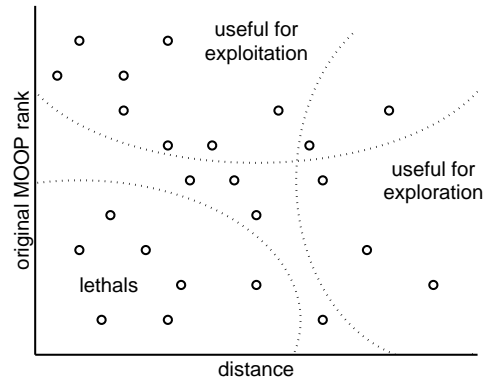


Figure 2: GeDEM fitness assignment

ations and the determination of Pareto optimality is an  $\mathcal{O}(n^2)$  algorithm: therefore, the computational complexity overhead of GeDEM is still  $\mathcal{O}(n^2)$ , as for the fitness sharing technique.

#### 4.2 Genetic Diversity Is a Special Objective

When a GeDEM is applied, we have to make a slight modification to the definition of Pareto dominance given in Section 2. In fact, the solutions having the same original MOOP ranks should not dominate one another, independent of the values of their distance-based diversity measure. Therefore, the definition of dominance used in the non-dominated sorting procedure performed by GeDEM is: Vector  $\mathbf{u} = (rank_u, dist_u)$  dominates vector  $\mathbf{v} = (rank_v, dist_v)$  if and only if  $rank_u < rank_v \wedge dist_u \leq dist_v$ . Figure 3 shows the difference induced by the two definitions in the determination of the non-dominated set of a sample population.

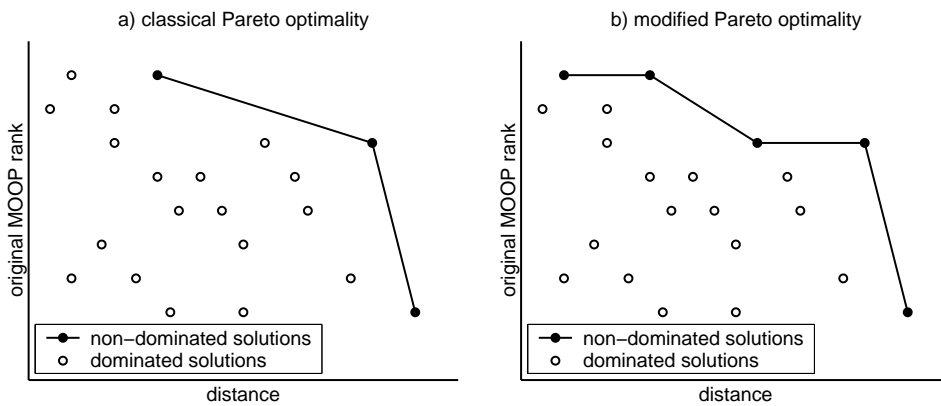


Figure 3: Modification of the Pareto optimality definition in the application of GeDEM

### 4.3 Clones and Elitism

The features of GeDEM strongly suggest the adoption of certain devices in order to achieve the best performance from it.

First of all, something should be said about “clones”. Clones are individuals having the same genotype, therefore they represent the same solution, and have the same objective function values and the same value of diversity. Their presence deeply affects the calculations made to determine the genetic diversity of each individual, because the ranks assigned by GeDEM are altered, especially those assigned to clones themselves. In fact, two identical solutions would be assigned a worse rank than the same non-cloned solution. Thus, the exploitation of the true relationships among the solutions requires the suppression of the offspring that are identical to other members of the population before the offspring undergo the evaluation process. The choice of removing clones also avoids a useless re-evaluation of the objective functions for the same solution, so that a greater number of points in the search space are investigated at the same computational effort; removing clones can also be seen as a mechanism that considers the individuals of the old population as a sort of “taboo” list of solutions for the offspring.

If clones are removed, recombination between clones becomes impossible and the probability that an individual is passed without changes to the next generation is considerably lowered. An elitism mechanism is therefore necessary to save the best solutions found so far. Zitzler and Thiele (1999) make use of a secondary population in their SPEA (see Section 3). Van Veldhuizen and Lamont (2000) point out that a secondary population of some sort is a necessity in MOEAs. Zitzler et al. (2000) find that the best-performing MOEAs make explicit use of elitist strategies (see also Section 6). The use of an elite set, however, raises new questions about its relationship with the population: Which solutions are inserted and how long are they kept in the elite set? When and how are which members of the elite set re-inserted into the population? Other elitist mechanisms, such as the  $(\mu + \lambda)$  selection scheme used in the area of evolutionary strategy (Schwefel, 1995; Bäck, 1996) do not involve secondary populations and are easier to implement.

## 5 The Genetic Diversity Evolutionary Algorithm

The Genetic Diversity Evolutionary Algorithm (GDEA) is a framework that is strictly designed around GeDEM to exalt its characteristics. Some of the design choices follow from the basic features of GeDEM (e.g., the replacement of clones, the use of an elitist strategy), the others are inspired by the will to make things as simple as possible, neither introducing arbitrary parameters nor using sophisticated heuristics.

The structure of GDEA follows the main steps of a  $(\mu + \lambda)$  Evolution Strategy (Schwefel, 1995; Bäck, 1996). Evolution, however, is considered at its genotypic level, with the traditional binary coding of the decision variables.

- Step 1:* An initial population of  $\mu$  individuals is generated at random.
- Step 2:* A mating pool of  $2\lambda$  individuals is formed, each individual having the same probability of being selected.
- Step 3:*  $\lambda$  offspring are generated by crossover. Some bits of the offspring are also randomly mutated with a probability  $p_{mut}$ .
- Step 4:* The whole population of  $\mu + \lambda$  individuals is checked to discover possible clones.

These clones are removed and replaced with new randomly generated individuals (this is done to encourage the exploration of the search space and also to have the algorithm evaluate, for convenience, new  $\lambda$  different offspring every generation; still the occurrence of clones' birth is not so frequent if clones are removed generation after generation).

- Step 5:* The objective function values of the  $\mu + \lambda$  individuals are evaluated and the non-dominated sorting procedure by Goldberg (1989) is performed to assign the ranks to the solutions according to the objectives of the MOOP.
- Step 6:* The whole population of  $\mu + \lambda$  individuals is processed to determine the value of the distance-based genetic diversity measure for each individual.
- Step 7:* GeDEM is applied according to the ranks scored in Step 5 and the values of the diversity measure assigned in Step 6. The non-dominated sorting procedure by Goldberg (1989) is used to assign the ranks.
- Step 8:* The best  $\mu$  solutions among parents and offspring, according to the ranks assigned in Step 7 by GeDEM, are selected for survival and the remaining  $\lambda$  are eliminated.
- Step 9:* If the maximum number of generations is reached then stop, else go to Step 2.

## 6 Comparison with Other Multiobjective Evolutionary Algorithms

A systematic comparison of various MOEAs was recently provided by Zitzler et al. (2000). They compared the performance of eight algorithms (VEGA, HLGA, FFGA, NPGA, NSGA, SPEA, a random search algorithm and a single-objective EA using weighted-sum aggregation) on six test problems featuring the characteristics that may cause difficulties in converging to the Pareto-optimal front and in maintaining diversity within the population (Deb, 1999): convexity, non-convexity, discrete Pareto fronts, multimodality, deception and biased search spaces. The aim was to assess which EAs are best fit for which problem feature and to determine the specific advantages and drawbacks of each technique. Contrary to what one might expect, the results showed that a hierarchy of algorithms exists, in which the first and second places are held by SPEA and NSGA, respectively. Further investigations demonstrated that an elitist variant of NSGA (the so-called NSGA-II) equals the performance of SPEA.

In this section, GDEA is tested on the same problems according to the same methodology. The performance is then compared according to the same metric with the simulation results of SPEA and NSGA obtained by Zitzler et al. (2000), available for public use at <http://www.tik.ee.ethz.ch/~zitzler/testdata.html>. The six test functions, the methodology and the metric of performance used in the comparison are briefly recalled in the following for easy reference.

### 6.1 Test Functions

Each of the six test functions  $\mathcal{T}_1, \dots, \mathcal{T}_6$  introduced by Zitzler et al. (2000) is a two-objective minimization problem that involves a distinct feature among those identified by Deb (1999). All the test functions are constructed in the same way, according to the guidelines in Deb (1999):

$$\begin{aligned} \text{Minimize} \quad & \mathcal{T}(\mathbf{x}) = (f_1(x_1), f_2(\mathbf{x})) \\ \text{subject to} \quad & f_2(\mathbf{x}) = g(x_2, \dots, x_n)h(f_1(x_1), g(x_2, \dots, x_n)) \\ \text{where} \quad & \mathbf{x} = (x_1, \dots, x_n) \end{aligned} \quad (3)$$



The function  $g$  controls the search space lateral to the Pareto-optimal front, while the function  $f_1$  controls the search space along the Pareto-optimal front. The function  $h$  determines the shape of the front itself. This method makes it possible to investigate the problem features separately in order to assess whether an MOEA has the ability to converge to the true Pareto-optimal set and to find diverse Pareto-optimal solutions under particular conditions.

- Test function  $\mathcal{T}_1$  has a convex Pareto-optimal front:

$$\begin{aligned} f_1(x_1) &= x_1 \\ g(x_2, \dots, x_n) &= 1 + 9 \cdot \sum_{i=2}^n x_i / (n-1) \\ h(f_1, g) &= 1 - \sqrt{f_1/g} \end{aligned} \quad (4)$$

where  $n = 30$  and  $x_i \in [0, 1]$ . The Pareto-optimal front corresponds to  $g(\mathbf{x}) = 1$ .

- Test function  $\mathcal{T}_2$  has a non-convex Pareto-optimal front:

$$\begin{aligned} f_1(x_1) &= x_1 \\ g(x_2, \dots, x_n) &= 1 + 9 \cdot \sum_{i=2}^n x_i / (n-1) \\ h(f_1, g) &= 1 - (f_1/g)^2 \end{aligned} \quad (5)$$

where  $n = 30$  and  $x_i \in [0, 1]$ . The Pareto-optimal front corresponds to  $g(\mathbf{x}) = 1$ .

- Test function  $\mathcal{T}_3$  has a discontinuous Pareto-optimal front consisting of several convex parts:

$$\begin{aligned} f_1(x_1) &= x_1 \\ g(x_2, \dots, x_n) &= 1 + 9 \cdot \sum_{i=2}^n x_i / (n-1) \\ h(f_1, g) &= 1 - \sqrt{f_1/g} - (f_1/g) \sin(10\pi f_1) \end{aligned} \quad (6)$$

where  $n = 30$  and  $x_i \in [0, 1]$ . The Pareto-optimal front corresponds to  $g(\mathbf{x}) = 1$ .

- Test function  $\mathcal{T}_4$  involves a multi-modal problem:

$$\begin{aligned} f_1(x_1) &= x_1 \\ g(x_2, \dots, x_n) &= 1 + 10(n-1) + \sum_{i=2}^n (x_i^2 - 10 \cos(4\pi x_i)) \\ h(f_1, g) &= 1 - \sqrt{f_1/g} \end{aligned} \quad (7)$$

where  $n = 10$ ,  $x_1 \in [0, 1]$  and  $x_2, \dots, x_n \in [-5, 5]$ . The true Pareto-optimal front is convex and corresponds to  $g(\mathbf{x}) = 1$ .

- Test function  $\mathcal{T}_5$  involves a deceptive problem, in which  $x_i$  are binary strings:

$$\begin{aligned} f_1(x_1) &= 1 + u(x_1) \\ g(x_2, \dots, x_n) &= \sum_{i=2}^n \left( \begin{cases} 2 + u(x_i) & \text{if } u(x_i) < 5 \\ 1 & \text{if } u(x_i) = 5 \end{cases} \right) \\ h(f_1, g) &= 1/f_1 \end{aligned} \quad (8)$$

where  $n = 11$ ,  $x_1 \in \{0, 1\}^{30}$  and  $x_2, \dots, x_n \in \{0, 1\}^5$ .  $u(x_i)$  is the number of 1s in  $x_i$  (unitation). The true Pareto-optimal front is convex and corresponds to  $g(\mathbf{x}) = 10$ .

- Test function  $\mathcal{T}_6$  features non-uniformity of the search space:

$$\begin{aligned} f_1(x_1) &= 1 - \exp(-4x_1) \sin^6(6\pi x_1) \\ g(x_2, \dots, x_n) &= 1 + 9 \cdot \sum_{i=2}^n x_i / (n-1) \\ h(f_1, g) &= 1 - (f_1/g)^2 \end{aligned} \quad (9)$$

where  $n = 10$  and  $x_i \in [0, 1]$ . The true Pareto-optimal front is non-convex and corresponds to  $g(\mathbf{x}) = 1$ .

## 6.2 Methodology

The methodology used in Zitzler et al. (2000) is strictly followed. GDEA is executed 30 times on each test function. However, it is not possible to start GDEA from the same 30 initial populations used by the other EAs, because these data are not available. Since GDEA includes a pure elitist strategy, the non-dominated solutions in the final population are considered as the outcome of a simulation run (on-line performance), while Zitzler et al. (2000) take into account the non-dominated solutions among all solutions generated during a simulation run (off-line performance). Each run ends after 250 generations. The population size  $\mu$  is set to 100 and the size of offspring population  $\lambda$  is set to 80, in order to perform the same number of fitness evaluations. The mutation probability  $p_{mut}$  is set to 0.01. The  $n$  decision variables are encoded in a single chromosome. Each parameter is represented by 30 bits, except for the deceptive test function  $\mathcal{T}_5$  in which  $x_2, \dots, x_n$  are represented by 5 bits. The measure of genetic diversity used by GeDEM in this study is the minimum normalized Euclidean distance in the decision variable space from another individual.

## 6.3 Metric of Performance

Different metrics can be defined to compare the performance of EAs with respect to the different goals of optimization itself (Zitzler et al., 2000): how far is the resulting non-dominated set from the Pareto front, how uniform is the distribution of the solutions along the Pareto set/front, how wide is the Pareto set/front. The metric used here (Zitzler and Thiele, 1999; Zitzler et al., 2000) is based on Pareto dominance and compares two sets of solutions determining the fraction of a set that is covered by the other.

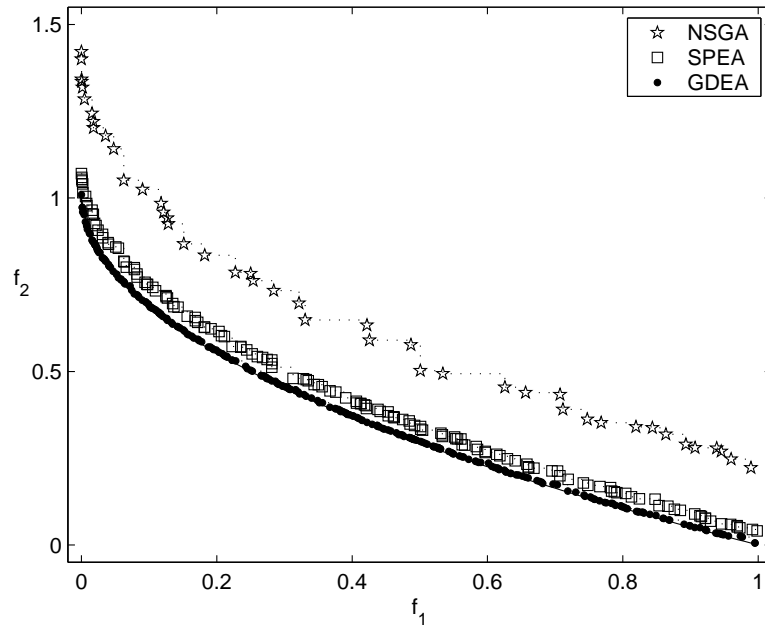
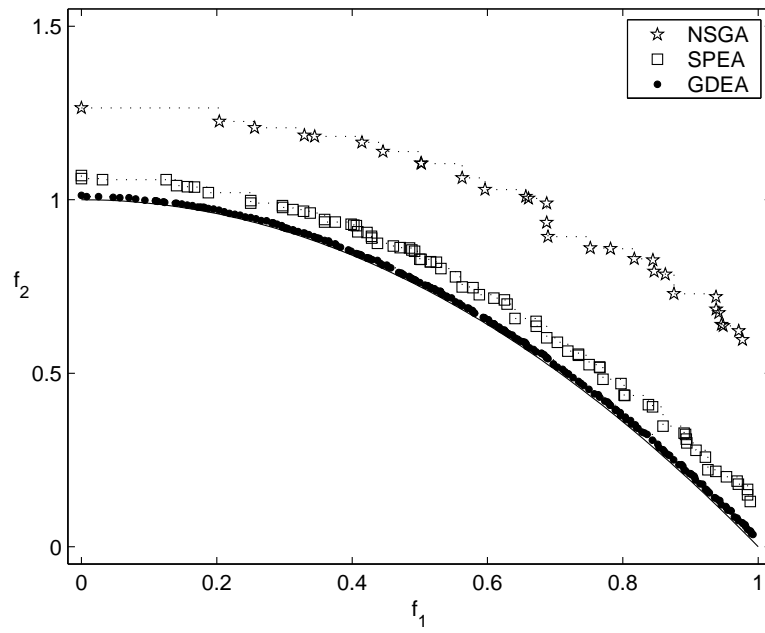
**Coverage of two sets:** Let  $X', X'' \subseteq N$  be two sets of decision vectors. The function  $\mathcal{C}$  maps the ordered pair  $(X', X'')$  to the interval  $[0, 1]$ :

$$\mathcal{C}(X', X'') := \frac{|\{\mathbf{x}'' \in X''; \exists \mathbf{x}' \in X' : F(\mathbf{x}') \preceq F(\mathbf{x}'')\}|}{|X''|} \quad (10)$$

If  $\mathcal{C}(X', X'') = 1$ , the set  $X''$  is entirely covered by the solutions in  $X'$ . Conversely, if  $\mathcal{C}(X', X'') = 0$ , none of the solutions in  $X''$  are covered by the set  $X'$ . Since  $\mathcal{C}(X', X'')$  is not necessarily equal to  $1 - \mathcal{C}(X'', X')$ , both the possible ordered pairs have to be considered. The  $\mathcal{C}$  function, when applied to the non-dominated set resulting from two different EAs, shows whether one set dominates the other, although superiority is not quantified.

## 6.4 Results of Comparison

As in Zitzler et al. (2000), Figures 4-9 show an excerpt of the non-dominated fronts obtained by the EAs and the Pareto-optimal fronts (continuous curves). The points plotted are the non-dominated solutions extracted from the union set of the outcomes of the first five runs. The performance of GDEA is also compared to NSGA and SPEA according to the  $\mathcal{C}$  metric defined in Equation 10. Since the initial populations are not the same, 900  $\mathcal{C}$  values are computed here per ordered algorithm pair  $(A_1, A_2)$  and test function. The distribution of these values is shown using box plots in Figure 10. Each box includes 50% of the values, the upper and lower ends of the box being the upper and lower quartiles. A thick horizontal line shows the value corresponding to the median and the upper and lower ends of the vertical line behind the box represent the maximum and minimum value, respectively.

Figure 4: Test function  $\mathcal{T}_1$  (convex).Figure 5: Test function  $\mathcal{T}_2$  (non-convex).

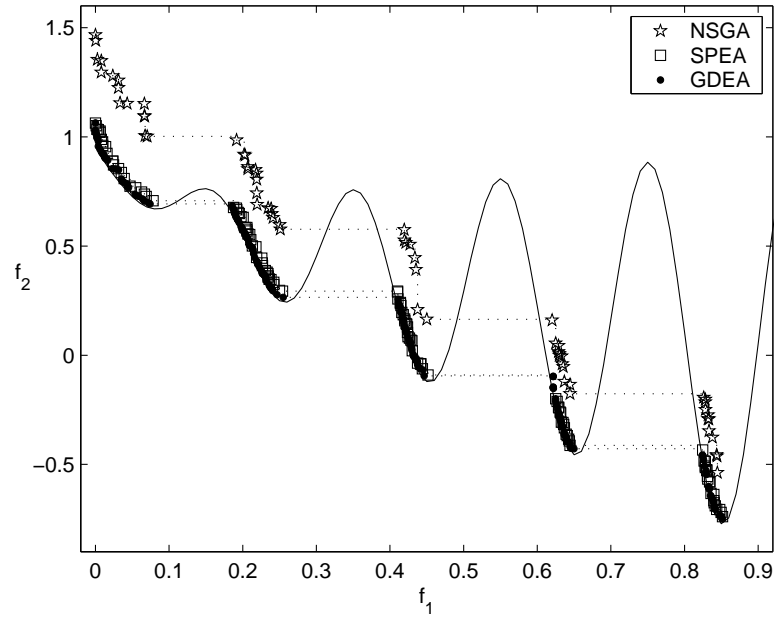


Figure 6: Test function  $\mathcal{T}_3$  (discrete).

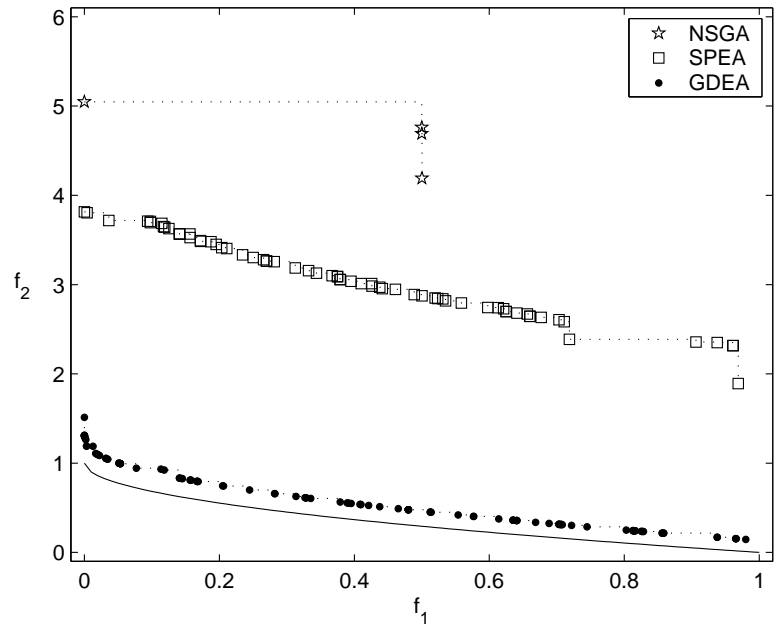
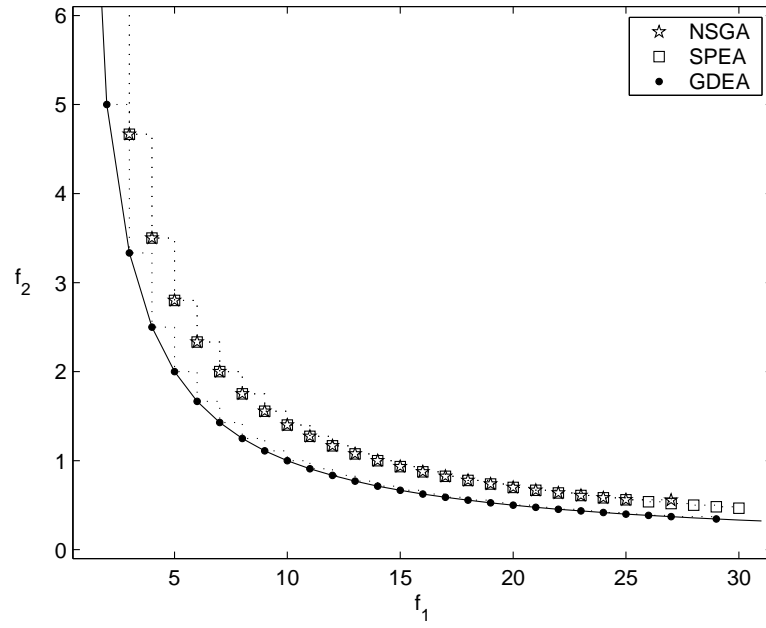
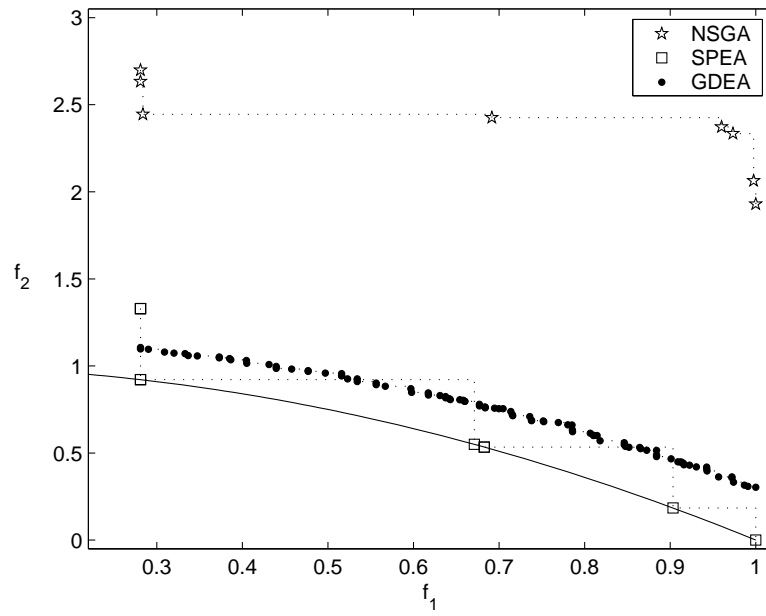


Figure 7: Test function  $\mathcal{T}_4$  (multi-modal).

Figure 8: Test function  $\mathcal{T}_5$  (deceptive).Figure 9: Test function  $\mathcal{T}_6$  (non-uniform).

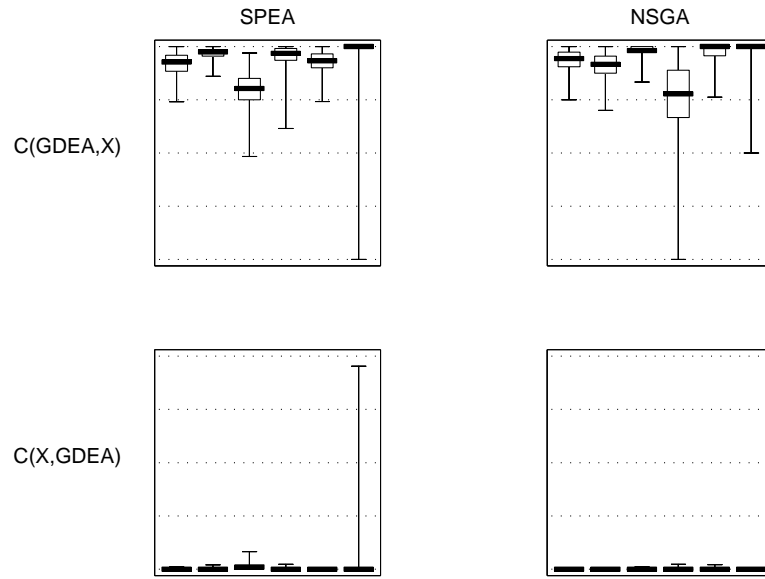


Figure 10: Box plots based on the  $\mathcal{C}$  metric. Each square contains six box plots representing the distribution of  $\mathcal{C}$  values for the six test functions, from  $\mathcal{T}_1$  (leftmost) to  $\mathcal{T}_6$  (rightmost). The scale is 0 at the bottom and 1 at the top in all squares.

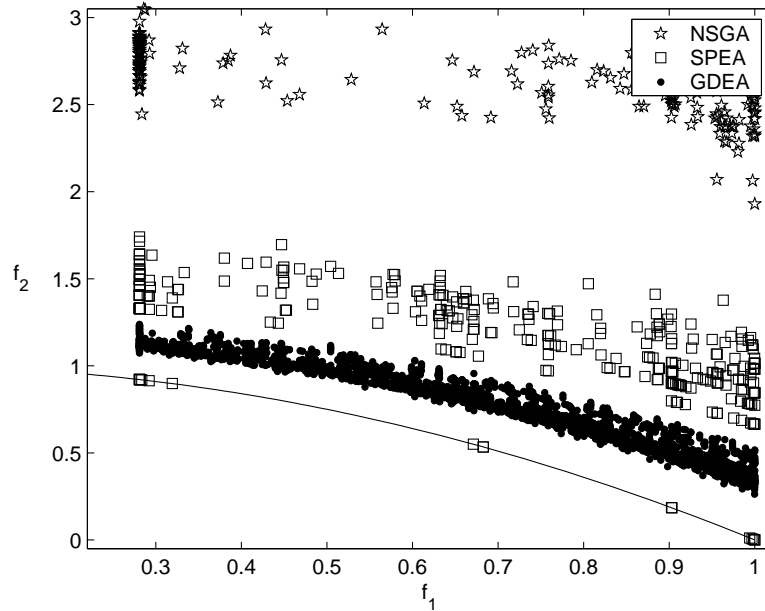


Figure 11: Test function  $\mathcal{T}_6$  (non-uniform). The outcomes of all the 30 simulation runs are plotted.

In general, the experimental results show that GDEA is able to converge towards the true Pareto-optimal front and to develop a widely and well distributed non-dominated set of solutions. The comparison with the other two best-performing MOEAs according to the  $\mathcal{C}$  metric proves that the performance of GDEA is somewhat superior. The non-dominated solutions obtained using GDEA cover, on average, more than the 75 percent of those obtained using SPEA or NSGA. On the other hand, the non-dominated solutions obtained using SPEA or NSGA hardly cover those obtained using GDEA.

Considering the specific features of the six test functions, GDEA shows similar performances both on convex and non-convex Pareto-optimal fronts. NSGA and SPEA seem instead to have more difficulties with non-convexity, since the non-dominated solutions found on test function  $\mathcal{T}_1$  are nearest to the Pareto-optimal front than those found on  $\mathcal{T}_2$ . Discreteness (test function  $\mathcal{T}_3$ ) is the case in which, on average, GDEA covers less solutions by SPEA (about 80%) and SPEA covers only a small percentage of the solutions by GDEA. This is probably due to the problem of determining the points of discontinuity of the Pareto-optimal front. The performance of GDEA is particularly remarkable in the cases of multi-modality and deception, which are considered by Zitzler et al. (2000) as the hardest problem features. GDEA gets nearer to the Pareto-optimal front of test function  $\mathcal{T}_4$  than NSGA and SPEA and it evolves to the Pareto-optimal front of  $\mathcal{T}_5$ . The poor coverage of NSGA by GDEA (about 75%) is due to the fact that an extremum of the Pareto-optimal front is among the solutions of one NSGA run. Finally, GDEA outperforms NSGA and SPEA in the case of biased search space (test function  $\mathcal{T}_6$ ) and is also able to evolve a well-distributed non-dominated set. These facts are not obvious in Figure 9, but, if all the non-dominated fronts achieved are plotted (Figure 11), it appears that GDEA covers all the solutions by SPEA apart from those obtained in a few lucky runs. On the other hand, the median of the distribution of both  $\mathcal{C}(\text{GDEA}, \text{SPEA})$  and  $\mathcal{C}(\text{GDEA}, \text{NSGA})$  is near one.

## 7 Conclusions

This paper has presented a new diversity-preserving mechanism (GeDEM) that is based on the idea that genetic diversity must show up as a real objective in the evaluation phase. Fitness assignment is performed by ranking the solutions according to the Pareto ranks scored with respect to the objectives of the MOOP and a distance-based measure of genetic diversity, creating a two-criteria optimization problem in which the objectives are the goals of the search process itself. A new MOEA was also provided (GDEA) that is designed around GeDEM and features a  $(\mu + \lambda)$  selection scheme as an elitist strategy.

Although this evolutionary approach to multi-objective optimization is not supported by sophisticated heuristics and makes “blind” use of the diversity information, the experimental results on a well-established suite of test functions proved that it is very effective at converging towards the Pareto-optimal set and at distributing the population regularly along it. The performance comparison showed that GDEA is at the same level of the best state-of-the-art MOEAs. This algorithm has also been battle-tested on real-world engineering problems in the fields of energy engineering (Benini et al., 2002; Toffolo and Lazzaletto, 2002) and turbomachinery aerodynamics (Benini and Toffolo, 2002a; Benini and Toffolo, 2002b), achieving excellent results.

## References

- Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York, New York.
- Bäck, T., Fogel, D. and Michalewicz, Z. (1997). *Handbook of Evolutionary Computation*. Oxford University Press, Oxford, England.
- Benini, E., Toffolo, A. and Lazzaretto, A. (2002). Evolutionary algorithms for multi-objective design optimization of combined-cycle power plants. In Giannakoglou K. et al., editors, *Evolutionary Methods for Design, Optimisation and Control*, CIMNE, Barcelona, Spain.
- Benini, E. and Toffolo A. (2002a). Development of high-performance airfoils for axial flow compressors using evolutionary computation. *AIAA Journal for Propulsion and Power*, 18(3):544-554.
- Benini, E. and Toffolo A. (2002b). Optimal design of horizontal-axis wind turbines using blade-element theory and evolutionary computation. *ASME Journal of Solar Energy Engineering*, 124(4):357-363.
- Coello Coello, C. A., Van Veldhuizen, D. A. and Lamont, G. B. (2002). *Evolutionary Algorithms for Solving Multi-objective Problems*. Kluwer Academic Publishers, New York, NY.
- Deb, K. (1999). Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation*, 7(3):205-230.
- Deb, K. (2001). *Multi-objective Optimization using Evolutionary Algorithms*. John Wiley, Chichester, UK.
- Fonseca, C. M. and Fleming, P. J. (1993). Genetic algorithms for multiobjective optimization: formulation, discussion and generalization. In Forrest, S., editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 416-423, Morgan Kaufmann, San Mateo, California.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley, Reading, Massachusetts.
- Goldberg, D. E. and Richardson, J. (1987). Genetic algorithms with sharing for multimodal function optimization. In Grefenstette, J. J., editor, *Proceedings of the Second International Conference on Genetic Algorithms*, pages 41-49, Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Hajela, P. and Lin C.-Y., (1992). Genetic search strategies in multicriterion optimal design. *Structural Optimization*, 4:99-107.
- Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, Michigan.
- Horn, J. and Nafploitis, N. (1993). Multiobjective optimization using the niched pareto genetic algorithm. IlliGAL Technical Report 93005, Illinois Genetic Algorithm Laboratory, University of Illinois, Urbana, Illinois.
- Horn, J., Nafploitis, N. and Goldberg, D. E. (1994). A niched pareto genetic algorithm for multi-objective optimization. In Michalewicz, Z., editor, *Proceedings of the First IEEE International Conference on Evolutionary Computation*, pages 82-87, IEEE Press, Piscataway, New Jersey.
- Mahfoud, S. W. (1995). *Niching Methods for Genetic Algorithms*. Ph.D. Thesis, University of Illinois, Urbana, Illinois.
- Schaffer, J. D. (1984). *Multiple Objective Optimization with Vector Evaluated Genetic Algorithms*. Unpublished Ph.D. Thesis, Vanderbilt university, Nashville, Tennessee.



- Schaffer, J. D. (1985). Multiple objective optimization with vector evaluated genetic algorithms. in Grefenstette, J. J., editor, *Proceedings of an International Conference on Genetic Algorithms and Their Applications*, pages 93-100, sponsored by Texas Instruments and U.S. Navy Center for Applied Research in Artificial Intelligence (NCARAI).
- Schwefel, H.-P. (1995). *Evolution and Optimum Seeking*. John Wiley, New York, New York.
- Srinivas, N. and Deb, K. (1994). Multi-objective function optimization using non-dominated sorting genetic algorithms. *Evolutionary Computation*, 2(3):221-248.
- Toffolo, A. and Lazzaretto, A. (2002). Evolutionary algorithms for multi-objective energetic and economic design optimization of energy systems. *Energy*, 27:549-567.
- Van Veldhuizen, D. A. (1999). *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. Ph.D. Thesis, AFIT/DS/ENG/99-01, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.
- Van Veldhuizen, D. A. and Lamont, G. B. (2000). Multiobjective evolutionary algorithms: analyzing the state-of-the-art. *Evolutionary Computation*, 8(2):125-147.
- Zitzler, E., Deb, K. and Thiele, L. (2000). Comparison of multiobjective evolutionary algorithms: empirical results. *Evolutionary Computation*, 8(2):173-195.
- Zitzler, E. and Thiele, L. (1999). Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257-271.