The Balance between Proximity and Diversity in Multi–Objective Evolutionary Algorithms

Peter A.N. Bosman, Dirk Thierens

Corresponding author is Peter A.N. Bosman. Institute of Information and Computing Sciences, Utrecht University P.O. Box 80.089, 3508 TB Utrecht, The Netherlands. Email: *Peter.Bosman@cs.uu.nl*. Telephone: +31 - 30 - 253 4097. Fax: +31 - 30 - 251 3791.

***Abstract*— Over the last decade, a variety of evolutionary algorithms (EAs) have been proposed for solving multi–objective optimization problems. Especially more recent multi–objective evolutionary algorithms (MOEAs) have been shown to be efficient and superior to earlier approaches. In the development of new MOEAs, the strive is to obtain increasingly better performing MOEAs. An important question however is whether we can expect such improvements to converge onto a specific efficient MOEA that behaves best on a large variety of problems. The best MOEAs to date behave similarly or are individually preferable with respect to different performance indicators. In this paper, we argue that the development of new MOEAs cannot converge onto a single new most efficient MOEA because the performance of MOEAs shows characteristics of multi–objective problems. While we will point out the most important aspects for designing competent MOEAs in this paper, we will also indicate the inherent multi–objective trade–off in multi–objective optimization between proximity and diversity preservation. We will discuss the impact of this trade–off on the concepts and design of exploration and exploitation operators. We also present a general framework for competent MOEAs and show how current state–of–the–art MOEAs can be obtained by making choices within this framework. Furthermore, we show an example of how we can separate non–domination selection pressure from diversity preservation selection pressure and discuss the impact of changing the ratio between these components.**

***Index Terms*— Multi–objective optimization, evolutionary algorithms, proximity, diversity, selection pressure, density estimation, exploitation, exploration**

I. INTRODUCTION

MULTI–OBJECTIVE optimization problems consist of m objectives $f_i(\boldsymbol{z})$, $i \in \boldsymbol{\mathcal{M}} = \{0, 1, \ldots, m-1\}$, that, without loss of generality, must all be minimized. However, there is no no expression of weight for any of the objectives, which means that the objectives cannot be combined in a single scalar objective to be minimized. As a result of this, sets of solutions exist such that each solution in this set is equally preferable. To formalize this notion, the following four concepts are of importance:

1) **Pareto dominance**. A solution $\boldsymbol{z}^0$ is said to (Pareto) *dominate* a solution $\boldsymbol{z}^1$ (denoted $\boldsymbol{z}^0 \succ \boldsymbol{z}^1$) if and only if $\left(\forall i \in \boldsymbol{\mathcal{M}} : f_i(\boldsymbol{z}^0) \leq f_i(\boldsymbol{z}^1)\right) \wedge \left(\exists i \in \boldsymbol{\mathcal{M}} : f_i(\boldsymbol{z}^0) < f_i(\boldsymbol{z}^1)\right)$
2) **Pareto optimal**. A solution $\boldsymbol{z}^0$ is said to be *Pareto optimal* if and only if $\neg\exists \boldsymbol{z}^1 : \boldsymbol{z}^1 \succ \boldsymbol{z}^0$
3) **Pareto optimal set**. The set $\boldsymbol{\mathcal{P}_S}$ of all Pareto optimal solutions: $\boldsymbol{\mathcal{P}_S} = \{\boldsymbol{z}^0 | \neg\exists \boldsymbol{z}^1 : \boldsymbol{z}^1 \succ \boldsymbol{z}^0\}$
4) **Pareto optimal front**. The set $\boldsymbol{\mathcal{P}_F}$ of all objective function values corresponding to the solutions in $\boldsymbol{\mathcal{P}_S}$: $\boldsymbol{\mathcal{P}_F} = \{\boldsymbol{f}(\boldsymbol{z}) = (f_0(\boldsymbol{z}), f_1(\boldsymbol{z}), \ldots, f_{m-1}(\boldsymbol{z})) | \boldsymbol{z} \in \boldsymbol{\mathcal{P}_S}\}$

The optimal solution for a multi–objective optimization problem is the Pareto optimal set $\boldsymbol{\mathcal{P}_S}$. The size of this set may however be infinite, in which case it is impossible to find this set using a finite number of solutions. In this case, a representative subset of $\boldsymbol{\mathcal{P}_S}$ is the desired result. The notion of searching a space by maintaining a finite population of solutions is characteristic of EAs, which makes them natural candidates for multi–objective optimization aiming to find a good approximation of the Pareto optimal front.

The current state–of–the–art MOEAs are capable of efficiently obtaining good approximations of the Pareto optimal front [1]. These current methods outperform earlier attempts. Different investigations regarding the performance of the algorithms have been published [2], [3], [4], [5], [6], [7]. However, comparing performances of MOEAs is not a trivial task since there is more than just a single goal that is of importance in finding a good approximation of the Pareto optimal front [3], [8], [9], [10], [11]. As a result, most of the currently best MOEAs do not outperform each other, but perform similarly or are preferable with respect to different performance indicators.

Simultaneously to the discovery of new MOEAs, research is also being devoted to investigating which components are the most important in designing competent MOEAs along with guidelines on the influence of certain operators on the performance of MOEAs [12], [7], [13], [14]. The combined research can be seen as an attempt at convergence towards a single framework that describes the components along with their settings for constructing the best possible performing MOEA. However, we will argue in this paper that this convergence is not possible because the performance of MOEAs shows characteristics of multi–objective problems.

The remainder of this paper is organized as follows. In Section II we discuss the goal in multi–objective optimization and indicate an important and inherent trade–off between proximity and diversity preservation. In Section III we describe how this trade–off has an impact on the concepts of exploitation and exploration in MOEAs. In Section IV we discuss the most important components for constructing competent MOEAs. Instances have to be chosen for these components to construct an actual MOEA. These choices however result in a bias that has a component regarding proximity with respect to the Pareto optimal front as well as a component regarding diversity preservation. Depending on the choices made, the bias towards each goal individually will be larger or smaller. We present our conclusions in Section V.

II. Multi–objective optimization goals

In this section we discuss the goal in solving multi–objective optimization problems. In Section II-A we indicate that although the optimum of a multi–objective optimization problem is well defined, there is more than one goal to take into account when evaluating approximations to the optimum. In Section II-B we discuss a few important performance indicators and in Section II-C we discuss the subtlety and the inherent trade–off in actually defining the goodness of an approximation.

A. Approximation sets, optimality and benchmarking

In this paper, we only consider the subset of all non–dominated solutions that is contained in the final population that results from running a MOEA. We call such a subset an *approximation set* and denote it by $\mathcal{S}$. The size of the approximation set depends on the settings used to run the MOEA with.

Regardless of the size of $\mathcal{P}_S$, we are interested in finding an approximation set of finite size that is a good approximate representation of $\mathcal{P}_S$. Ideally, we would like to obtain an approximation set that contains a selection of solutions from $\mathcal{P}_S$ such that the solutions in the approximation set are as diverse as possible. However, we do not have access to $\mathcal{P}_S$ on beforehand. Therefore, we want to get close to $\mathcal{P}_S$ but in such a way that the approximation set $\mathcal{S}$ that we find, is as diverse as possible, without compromising as much as possible the proximity of $\mathcal{S}$ with respect to $\mathcal{P}_S$. Regarding this diversity, it is important to note that it depends on the mapping between the parameter space and the objective space whether a good spread of the solutions in the parameter space is also a good spread of the solutions in the objective space. However, it is common practice to search for a good diversity of the solutions in the objective space along the approximation set [1]. The reason for this is that a decision–maker will ultimately have to pick a single solution. Therefore, it is often best to present a wide variety of trade–off solutions for the specified goals.

There is an inherent trade–off in the intuitive two–sided goal since it is desirable to obtain a diverse approximation set as well as it is desirable to obtain an approximation set that is close to the optimal one. However, this trade–off only exists if we assume that we are not able to find the optimal approximation set. The optimal approximation set is well defined if we assume a fixed size of the approximation set. The optimal approximation set is a selection of solutions from $\mathcal{P}_S$ such that the solutions in the approximation set are as diverse as possible. Since the distance to the Pareto optimal front for any solution in the optimal approximation set is 0 and we assume a fixed size of the approximation set, optimality can now be obtained by optimizing only a single objective, which is diversity. In general, there are two ways to benchmark EAs. Either we know the optimum and determine the resources such as population size and number of evaluations that are required on average to obtain the optimum in a predefined percentage of all runs, or we fix the number of evaluations on beforehand and determine the maximum score that the EAs obtain on average over all runs. The first way of benchmarking results in values for different EAs that can directly be compared to each other and be used to determine whether one EA is a more competent optimizer than is another. This is also the case for multi–objective optimization since the optimal approximation set is well defined. The second way of benchmarking represents a more practical situation, since we usually do not assume that an unlimited number of function evaluations is available. For single–objective optimization, the objective value can directly be used as the score in this type of benchmark. In multi–objective optimization this is not the case due to the trade–off in the two–sided goal in multi–objective optimization that we have pointed out. This trade–off will be reflected in the score that we use to compare the results of the EAs in the benchmark. It is this type of benchmarking of MOEAs and the resulting trade–off that we investigate in this paper.

B. Performance indicators

In this section, we discuss performance indicators. A performance indicator is a function that, given an approximation set $\mathcal{S}$, returns a real value that indicates how good $\mathcal{S}$ is with respect to a certain feature that is measured by the performance indicator. Performance indicators are commonly used to determine the performance of a MOEA and to compare this performance with other MOEAs if the number of evaluations is fixed on beforehand. However, there are some important limitations to the use of performance indicators. We first describe a few important performance indicators. Subsequently, we will discuss the limitations of performance indicators and point out the resulting implications for our investigation of the balance between proximity and diversity in MOEAs.

1) Selected performance indicators: Since we are usually interested in the performance of a MOEA as measured in the objective space, we define the distance between two multi–objective solutions $\boldsymbol{z}^0$ and $\boldsymbol{z}^1$ to be the Euclidean distance between their objective values $\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{f}(\boldsymbol{y})$:

$$d(\boldsymbol{z}^0, \boldsymbol{z}^1) = \sqrt{\sum_{i=0}^{m-1} (f_i(\boldsymbol{z}^1) - f_i(\boldsymbol{z}^0))^2} \tag{1}$$

If we only want to measure diversity, we can use the **FS** (*Front Spread*) indicator. This performance indicator was first used by Zitzler [15]. The **FS** indicator indicates the size of the objective space covered by an approximation set. A larger **FS** indicator value is preferable. The **FS** indicator for an approximation set $\mathcal{S}$ is defined to be the maximum Euclidean distance inside the smallest m–dimensional bounding–box that contains $\mathcal{S}$. This distance can be computed using the maximum distance among the solutions in $\mathcal{S}$ in each dimension separately:

$$\mathbf{FS}(\mathcal{S}) = \sqrt{\sum_{i=0}^{m-1} \max_{(\boldsymbol{z}^0, \boldsymbol{z}^1) \in \mathcal{S} \times \mathcal{S}} \{(f_i(\boldsymbol{z}^0) - f_i(\boldsymbol{z}^1))^2\}} \tag{2}$$

In combination with the **FS** indicator, it is also important to know how many points are available in the set of non–dominated solutions, because a larger set of trade–off points is more desirable. This quantity is called the **FO** (*Front Occupation*) indicator and was first used by Van Veldhuizen [16]. A larger **FO** indicator value is preferable.

$$\mathbf{FO}(\mathcal{S}) = |\mathcal{S}| \tag{3}$$

The ultimate goal is to cover the Pareto optimal front. An intuitive way to define the distance between an approximation set $\mathcal{S}$ and the Pareto optimal front is to average the minimum distance between a solution and the Pareto optimal front over each solution in $\mathcal{S}$. We refer to this distance as the distance from a set of non–dominated solutions to the Pareto optimal front and it serves as a proximity indicator, which we denote by $\boldsymbol{D}_{\mathcal{S}\to\boldsymbol{\mathcal{P}_F}}$. This performance indicator was first used by Van Veldhuizen [16]. A smaller value for this performance indicator is preferable.

$$\boldsymbol{D}_{\mathcal{S}\to\boldsymbol{\mathcal{P}_F}}(\mathcal{S}) = \frac{1}{|\mathcal{S}|}\sum_{\boldsymbol{z}^0\in\mathcal{S}} \min_{\boldsymbol{z}^1\in\boldsymbol{\mathcal{P}_S}}\{d(\boldsymbol{z}^0,\boldsymbol{z}^1)\} \tag{4}$$

An approximation set with a good $\boldsymbol{D}_{\mathcal{S}\to\boldsymbol{\mathcal{P}_F}}$ indicator value does not imply that a good diverse representation of the Pareto optimal set has been obtained, since the indicator only reflects how far away the obtained points are from the Pareto optimal front on average. An approximation set consisting of only a single solution can already have a low value for this indicator. To include the goal of diversity, the reverse of the $\boldsymbol{D}_{\mathcal{S}\to\boldsymbol{\mathcal{P}_F}}$ indicator is a better guideline for evaluating MOEAs. In the reverse distance indicator, we compute for each solution in the Pareto optimal set the distance to the closest solution in an approximation set $\mathcal{S}$ and take the average as the indicator value. We denote this indicator by $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ and refer to it as the distance from the Pareto optimal front to an approximation set. A smaller value for this performance indicator is preferable. In the definition of this indicator, we must realize that the Pareto optimal front may be continuous. For an exact definition, we therefore have to use a line integration over the entire Pareto front. For a 2–dimensional multi–objective problem we obtain the following expression:

$$\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}(\mathcal{S}) = \int_{\boldsymbol{\mathcal{P}_F}} \min_{\boldsymbol{z}^0\in\mathcal{S}}\{d(\boldsymbol{z}^0,\boldsymbol{z}^1)\}d\boldsymbol{f}(\boldsymbol{z}^1) \tag{5}$$

In most practical test applications, it is easier to compute a uniformly sampled set of many solutions along the Pareto optimal front and to use this discretized representation of $\boldsymbol{\mathcal{P}_F}$ instead. A discretized version of the Pareto optimal front is also available if a discrete multi–objective optimization problem is being solved. In the discrete case, the $\boldsymbol{D}_{\mathcal{S}\to\boldsymbol{\mathcal{P}_F}}$ indicator is defined by:

$$\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}(\mathcal{S}) = \frac{1}{|\boldsymbol{\mathcal{P}_S}|}\sum_{\boldsymbol{z}^1\in\boldsymbol{\mathcal{P}_S}} \min_{\boldsymbol{z}^0\in\mathcal{S}}\{d(\boldsymbol{z}^0,\boldsymbol{z}^1)\} \tag{6}$$

An illustration of the $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator is presented in Figure 1. The $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator represents both the proximity and the diversity goal in multi–objective optimization. The $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator for an approximation set $\mathcal{S}$ is zero if and only if all points in $\boldsymbol{\mathcal{P}_F}$ are contained in $\mathcal{S}$ as well. Furthermore, a single solution from the Pareto optimal set will lead to the same $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator as a more diverse set of solutions that has objective values that are slightly further away from the Pareto optimal front. Moreover, a similarly diverse approximation set of solutions that is closer to the Pareto optimal front, will have a lower $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator value. However, an approximation set of solutions that is extremely diverse but far away from the Pareto optimal front, such as the non–dominated solutions of a randomly generated set of solutions, has a bad $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator value. This underlines the important point that diversity is *not* equally important as is proximity because a larger diversity is often not hard to come by. What is important is the diversity *along* the objectives of a set of non–dominated solutions that is as close as possible to the Pareto optimal front.

A performance indicator that is closely related to the $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator, is the hypervolume indicator by Knowles and Corne [9]. In the hypervolume indicator, a point in the objective space is picked such that it is dominated by all points in the approximation sets that need to be evaluated. The indicator value is then equal to the hypervolume of the multi–dimensional region enclosed by the approximation set and the picked reference point. This value is an indicator of the region in the objective space that is dominated by the approximation set. The main difference between the hypervolume indicator and the $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator is that for the hypervolume indicator a reference point has to be chosen. Different reference points lead to different indicator values. Moreover, different reference points can lead to indicator values that indicate a preference for different approximation sets. Since in the $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator the true Pareto optimal front is used, the $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator does not suffer from this drawback. Of course, a major drawback of the $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator is that in a real application the true Pareto optimal front is not known on beforehand. In that case, the Pareto front of all approximation sets could be used as a substitute for the actual Pareto optimal front.

2) The relation between performance indicators and the comparison of MOEAs: If we want to use performance indicators to investigate the performance of a MOEA and compare it with other MOEAs, there are some important limitations to consider that have been proven by Zitzler *et al.* [11]. These limitations are related to the extent to which performance indicators are capable of truly indicating whether one approximation set $\mathcal{S}^0$ is better than $\mathcal{S}^1$ in a certain sense. To this end, the concept of domination has to be extended to approximation sets. Zitzler *et al.* [11] consider the following dominance relations for approximation sets:

1) **Strict Pareto dominance**. An approximation set $\mathcal{S}^0$ is said to *strictly (Pareto) dominate* an approximation set $\mathcal{S}^1$ (denoted $\mathcal{S}^0 \succ\succ \mathcal{S}^1$) if and only if $\forall \boldsymbol{z}^1 \in \mathcal{S}^1 : \left(\exists \boldsymbol{z}^0 \in \mathcal{S}^0 : \left(\forall i \in \boldsymbol{\mathcal{M}} : f_i(\boldsymbol{z}^0) < f_i(\boldsymbol{z}^1)\right)\right)$
2) **Pareto dominance**. An approximation set $\mathcal{S}^0$ is said to *(Pareto) dominate* an approximation set $\mathcal{S}^1$ (de-

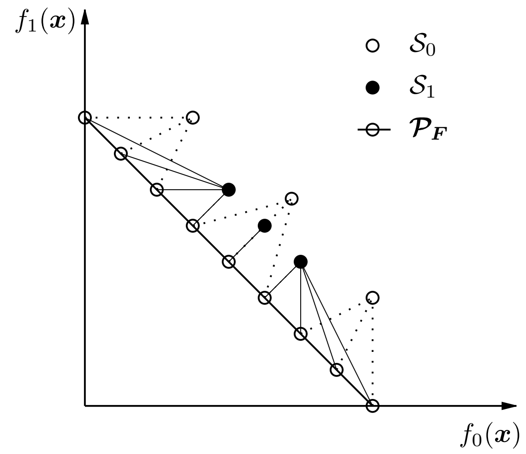

Fig. 1. The approximation set $\mathcal{S}_1$ is closer to the (discretized) Pareto optimal front but has less diversity, while approximation set $\mathcal{S}_0$ is further away from the front but has greater diversity: both sets have approximately the same $\boldsymbol{D}_{\mathcal{P}_F \to \mathcal{S}}$ indicator value though.

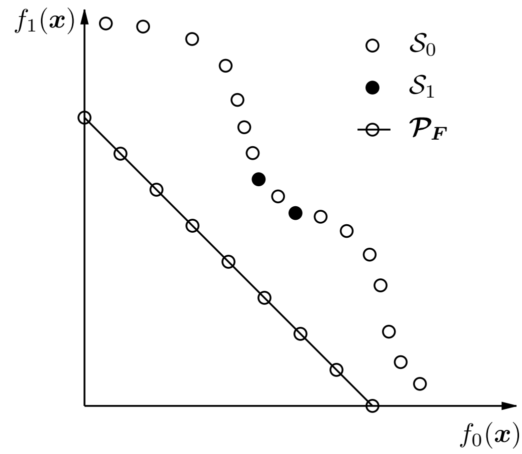

Fig. 2. Although approximation sets $\mathcal{S}_0$ and $\mathcal{S}_1$ are very different and approximation set $\mathcal{S}_0$ has many more non–dominated solutions and a better diversity than $\mathcal{S}_1$, using the dominance criteria by Zitzler *et al.* [11] $\mathcal{S}_0$ and $\mathcal{S}_1$ can only be classified as incomparable.

noted $\mathcal{S}^0 \succ \mathcal{S}^1$) if and only if $\forall \boldsymbol{z}^1 \in \mathcal{S}^1 : \left(\exists \boldsymbol{z}^0 \in \mathcal{S}^0 : \boldsymbol{z}^0 \succ\!\succ \boldsymbol{z}^1\right)$

3) **Better**. An approximation set $\mathcal{S}^0$ is said to be *better* than an approximation set $\mathcal{S}^1$ (denoted $\mathcal{S}^0 \rhd \mathcal{S}^1$) if and only if $\mathcal{S}^0 \neq \mathcal{S}^1 \wedge \left(\forall \boldsymbol{z}^1 \in \mathcal{S}^1 : \left(\exists \boldsymbol{z}^0 \in \mathcal{S}^0 : \left(\forall i \in \boldsymbol{\mathcal{M}} : f_i(\boldsymbol{z}^0) \leq f_i(\boldsymbol{z}^1)\right)\right)\right)$
4) **Weak Pareto dominance**. An approximation set $\mathcal{S}^0$ is said to *weakly (Pareto) dominate* an approximation set $\mathcal{S}^1$ (denoted $\mathcal{S}^0 \succeq \mathcal{S}^1$) if and only if $\forall \boldsymbol{z}^1 \in \mathcal{S}^1 : \left(\exists \boldsymbol{z}^0 \in \mathcal{S}^0 : \left(\forall i \in \boldsymbol{\mathcal{M}} : f_i(\boldsymbol{z}^0) \leq f_i(\boldsymbol{z}^1)\right)\right)$
5) **Incomparable**. An approximation set $\mathcal{S}^0$ is said to be *incomparable* to an approximation set $\mathcal{S}^1$ (denoted $\mathcal{S}^0 \parallel \mathcal{S}^1$) if and only if $\neg\left(\mathcal{S}^0 \succeq \mathcal{S}^1\right) \wedge \neg\left(\mathcal{S}^1 \succeq \mathcal{S}^0\right)$

It was shown by Zitzler *et al.* [11] that for any finite combination of performance indicators such as the ones presented in the previous section, there is no function of these performance indicators that specifies for any two approximation sets $\mathcal{S}^0$ and $\mathcal{S}^1$ whether $\mathcal{S}^0 \rhd \mathcal{S}^1$ holds. Thus, using the terminology and definitions by Zitzler *et al.* [11], we may not draw any conclusions regarding whether one approximation set is better than another approximation set on the basis of performance indicators such as the ones we have described so far.

Although the result by Zitzler *et al.* [11] is very important, its implications only apply to cases in which it is clear from a domination point of view that one approximation set is better than another approximation set. For instance, if $\mathcal{S}^0 \succ\!\succ \mathcal{S}^1$ holds, then $\mathcal{S}^0$ is truly preferable over $\mathcal{S}^1$. However, there are some important further aspects to consider that relate to the comparison of competent MOEAs. Even if $\mathcal{S}^0 \parallel \mathcal{S}^1$ holds, we could still prefer $\mathcal{S}^0$ over $\mathcal{S}^1$. Consider for instance the example in Figure 2. Following the definitions for comparing approximation sets by Zitzler *et al.* [11], $\mathcal{S}^0 \parallel \mathcal{S}^1$ holds. However, $\mathcal{S}^0$ has many more non–dominated solutions and a much larger diversity than does $\mathcal{S}^1$. Even if $\mathcal{S}^1$ had only a single solution placed somewhere on the line between the current two solutions in $\mathcal{S}^1$, the two approximation sets would still be incomparable. Still, it is fair to say here that approximation set $\mathcal{S}^0$ is preferable.

The class of incomparable approximation sets is very large and it can be argued that this class includes sets that may clearly be called preferable over other sets in the same class. It can furthermore be argued that this class is filled with pairs of approximation sets such that one approximation set of this pair is not clearly preferable over the other approximation set. This is for instance often the case if the two approximation sets intersect in the objective space and have a comparable diversity and size. Another example of pairs of approximation sets that are not easily said to be preferable over each other is given in Figure 3. This example represents the arguable statement that the class of incomparable approximation sets contains a large number of approximation sets that represent a true trade–off between the goals of proximity and diversity.

The existence of trade–off approximation sets in the class of incomparable approximation sets is very important when comparing MOEAs. As the efficiency of newly designed MOEAs increases, results such as the one in Figure 3 will become ever more likely to occur. Clearly, if two algorithms have the same emphasis on diversity preservation as they have on getting as close as possible to the Pareto optimal front, these algorithms will end up with approximation sets that are incomparable, unless one algorithm is truly less competent than the other, in which case testing the results of the algorithms using the categorization by Zitzler *et al.* [11] will point out which algorithm is superior. However, if one algorithm places more emphasis on diversity preservation and the other algorithm places more emphasis on getting get as close as possible to the Pareto optimal front, results such as the ones in Figures 2 and 3 are likely to occur. The categorization by Zitzler *et al.* [11] will in both cases point out that the algorithms are incomparable although it can be argued very plausibly in the case of Figure 2 that one result is less preferable than the other. In this case, performance indicators such as the ones that we have described can offer additional information. In the case of Figure 2 we will find that although the $\boldsymbol{D}_{\mathcal{S} \to \mathcal{P}_F}$ and $\boldsymbol{D}_{\mathcal{P}_F \to \mathcal{S}}$

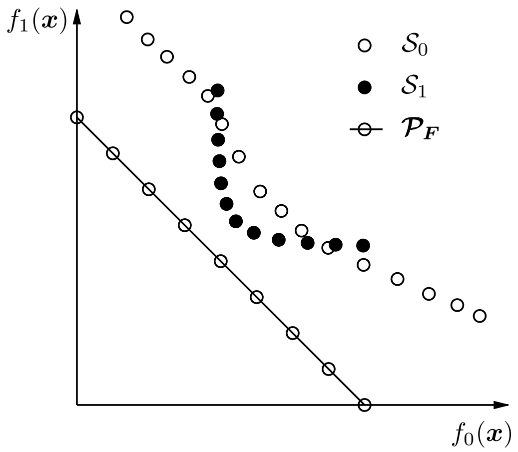

Fig. 3. Whereas approximation set $\mathcal{S}_0$ is very diverse and has an overall good approximation of the Pareto optimal front, approximation set $\mathcal{S}_1$ is less diverse but has a better proximity with respect to the Pareto optimal front. These two approximation sets represent a trade–off in that, without a preference for diversity or proximity with respect to the Pareto optimal front, neither approximation set can be called preferable. Using the dominance criteria by Zitzler *et al.* [11] $\mathcal{S}_0$ and $\mathcal{S}_1$ are classified as incomparable.

indicators are relatively similar, the **FS** and **FO** indicators will be significantly better for $\mathcal{S}_0$ than for $\mathcal{S}_1$. This will lead us to conclude that $\mathcal{S}_0$ is indeed preferable. In the case of Figure 3 we will find that the $\boldsymbol{D}_{\mathcal{P}_F \to \mathcal{S}}$ and **FO** indicators are relatively similar, but the $\boldsymbol{D}_{\mathcal{S} \to \mathcal{P}_F}$ indicator is better for $\mathcal{S}_1$ whereas the **FS** indicator is significantly better for $\mathcal{S}_0$. This will lead us to decide that neither approximation set is preferable.

Concluding, there is a good chance that two MOEAs are classified as being incomparable with respect to the definitions of Zitzler *et al.* [11] unless a truly significant competence difference between the MOEAs exists. Moreover, if the results of two MOEAs are classified as being incomparable, the one MOEA may still be called more preferable than the other depending on the balance between proximity and diversity. On the other hand, if the two MOEAs are both competent, the trade–off that lies in the balance between proximity and diversity can cause the results of the two algorithms to be quite different or to be quite similar and yet we cannot clearly say that one MOEA is more preferable than the other.

C. Multi–objective trade–off between the goals

Unless an unlimited number of evaluations is allowed, it will depend on towards which goal we bias our MOEA whether we will arrive at a lower $\boldsymbol{D}_{\mathcal{S} \to \mathcal{P}_F}$ indicator value or at a higher **FS** indicator value. Depending on the importance that we associate with diversity along the resulting approximation set, a larger or smaller bias will be needed towards diversity preservation. In this sense, there is a Pareto optimal set of MOEAs such that, depending on the emphasis on approaching the Pareto optimal front or on preserving diversity, all MOEAs in this set are incomparable and moreover no MOEA in this set is more preferable than any other MOEA in this set. The inherent trade–off in performance is reflected by the $\boldsymbol{D}_{\mathcal{P}_F \to \mathcal{S}}$ indicator value that we use in this paper as a plausible joint indicator of the two goals, since a less diverse approximation set that is closer to the Pareto optimal front will result in the same indicator value as a more diverse approximation set that is slightly further away from the Pareto optimal front.

III. Balancing proximity and diversity in exploitation and exploration

The fact that the multi–objective optimization goal is two–sided, has a direct implication on the notions of *exploitation* and *exploration* as commonly used in EA terminology. To avoid confusion, we will give an exact definition of what we consider to be exploitation and exploration:

1) **Exploitation** indicates the parts of an EA that are concerned with the selection of a set of parent solutions from the current population and the construction of a new population given the current population, the selected set of parent solutions and the set of offspring solutions. This definition of exploitation thus includes traditional selection, but also all replacement schemes such as crowding.
2) **Exploration** indicates the part of an EA that is concerned with the generation of new offspring solutions from a given set of parent solutions. Since we are only interested in how a new set of solutions is generated if we supply a set of solutions, our definition of exploration includes the way in which mating is performed. The actual operator that constructs a new solution using a set of mated parents is called the *variation* operator.

In this section, we indicate the implications that the two–sided goal in multi–objective optimization has on the classical exploitation and exploration concepts in EAs. These two phases can be split into two subprocedures that aid the search for proximity as well as for diversity amongst non–dominated solutions. This is an important issue that should be considered when constructing new MOEAs. In the following subsections we discuss the splitting of the exploitation and exploration phases in more detail. Furthermore, we also indicate the importance of elitism in multi–objective optimization, and discuss its contribution to exploitation and exploration.

A. Exploitation of proximity

In a practical application, we do not have access to a performance indicator such as the $\boldsymbol{D}_{\mathcal{S} \to \mathcal{P}_F}$ indicator that can give us an idea of how close we are to the Pareto optimal front. To ensure selection pressure towards the Pareto optimal front in the absence of such an indicator, the best we can do is to find solutions that are not dominated by any other solution. A selection operator that selects non–dominated solutions in combination with effective exploration operators will effectively drive the search towards the Pareto optimal front.

A straightforward way to obtain selection pressure towards non–dominated solutions is to count for each solution in the population the number of times it is dominated by another solution in the population, which is called the *domination count* of a solution [7], [17]. The rationale behind the domination count approach is that ultimately we would like no solution to be dominated by any other solution, so the less times a solution is dominated, the better. A lower domination count is preferable. Using this value we can apply truncation selection or tournament selection to obtain solid pressure towards non–dominated solutions.

Another approach to ensuring a preference for solutions that are dominated as little as possible, is to assign a preference to different domination *ranks* [18], [19]. The solutions that are in the j–th rank are those solutions that are non–dominated if the solutions of all ranks $i < j$ are disregarded. Note that the best domination rank contains all solutions that are non–dominated in the complete population. A lower rank is preferable. Using this value we can again apply for instance either truncation selection or tournament selection. Similar to the domination count approach, this approach effectively prefers solutions that are closer to the set of non–dominated solutions.

B. Exploitation of diversity

To ensure that diversity is preserved, the selection procedure must be provided with a component that prefers a diverse selection of solutions. However, since the goal is to preserve diversity *along* an approximation set that is as close as possible to the Pareto optimal front, rather than to preserve diversity in general, the exploitation of diversity should not precede the exploitation of proximity.

In most multi–objective selection schemes, diversity is used as a second comparison key in the exploitation phase. This prohibits tuning the amount of diversity exploitation that can be done compared to the amount of proximity exploitation. An example is the approach taken in the NSGA–II in which solutions are selected based on their non–domination rank using tournament selection [19]. If the ranks of two solutions are equal, the solution that has the largest total distance between its two neighbors summed over each objective, is preferred. This gives a preference to non–crowded solutions.

The explicit exploitation of diversity may serve more than just the purpose of ensuring that a diverse subset is selected from a certain set of non–dominated solutions. If we only apply selection pressure to finding the non–dominated solutions and enable diversity preservation only to find a good spread of solutions in the approximation set, we increase the probability that we only find a subset of a discontinuous Pareto optimal front. Diversity exploitation will most likely be too late in helping out to find the other parts of the discontinuous Pareto optimal front as well. Therefore, we may need to spend more attention on diversity preservation during optimization and perhaps even increase the amount of diversity preservation. Another reason why we may need to increase the exploitation of diversity preservation is that a variation operator is used that can find many more non–dominated solutions, which could cause a MOEA to converge prematurely onto subregions of a Pareto optimal front or onto locally optimal sets of non–dominated solutions, unless the population size is increased. However, given a fixed number of evaluations, this can be a significant drawback in approaching the Pareto optimal front. This problem can be alleviated by placing more emphasis on diversity exploitation and by consequently reducing the effort in the exploitation of proximity. By doing so, the variation operator is presented with a more diverse set of solutions from which a more diverse set of offspring will result. Furthermore, solutions that are close to each other will now have a smaller joint chance that they will both be selected, which improves the ability to approach the Pareto optimal front since premature convergence is less likely.

C. Exploration of proximity

Although it is important to have a competent selection mechanism that is capable of selecting a diverse set of solutions close to the set of non–dominated solutions, it is also important to have an exploration mechanism that is capable of producing new non–dominated solutions for the optimization process to proceed towards the Pareto optimal front. However, based on a proper selection of solutions, competent exploration operators should be able to generate new solutions in which good features of the selected solutions are combined so as to obtain better solutions. Essentially, this does not differ much from the necessity to generate better solutions by combining information from parent solutions in single–objective EAs.

As an alternative to classical recombination and mutation, more involved operators exist that are capable of analyzing the structure of the problem based on the selected solutions. This problem structure can subsequently be used to generate better solutions with a larger probability than can be done using classical operators. Such operators exploit observable regularities of a certain form and attempt to respect these regularities as much as possible when constructing new solutions. Examples of such competent operators are the ones used in the mGA [20], the fmGA [21], the GEMGA [22], the LLGA [23] and the BBF–GA [24].

Another interesting and relatively new field of EAs that attempts to model the regularities of the problem structure, uses probabilistic models to describe the probability distribution of the selected samples. By drawing new samples from the estimated probability distribution, a more global statistical inductive type of iterated search is obtained. Algorithms that use such techniques have obtained an increasing amount of attention over the last few years, obtaining promising results on a large variety of problems [25], [26], [27], [28], [29], [30], [31], [32], [7]. It has been indicated that the use of such approaches can be beneficial in multi–objective optimization as well [7], [33].

D. Exploration of diversity

Similar to the necessity of proximity exploration, exploration of diversity is equally important. If we are not able to construct a diverse set of good solutions, there will be hardly any diversity to be preserved at all. Just as is the case for the construction of new non–dominated solutions, the construction of a diverse set of new solutions depends on the competence of the exploration operators and the solutions that were offered to them. Ideally, an exploration phase results in new non–dominated solutions that are spread across a wide range in the objective space. Such behavior can be stimulated by clustering the selected solutions based on their objective values and by using a simpler exploration operator in each cluster separately. In the case of the probabilistic operators, such an approach constructs a mixture model. By clustering the objective space, the exploration of diversity along the set of non–dominated solutions has been shown to be effectively stimulated [7], [33].

Note that, although it is crucial to have good exploration in both the proximity sense as well as in the diversity sense, the actual implementation thereof is relatively independent from the necessity of having a robustly tunable trade–off between proximity exploitation and diversity preserving exploitation. Clearly, better exploration operators will lead to better results, but if the two selection components are not properly established and combined, the resulting MOEA is likely not to be in the Pareto optimal set of best approaches for multi–objective optimization.

E. Exploitation and exploration by use of elitism

In the use of elitism, the best solutions of the current generation are copied into the next generation. Alternatively, an external archive of a predefined maximum size n_a may be used that contains only non–dominated solutions. This is actually a similar approach to using elitism in a population, because this archive can be seen as the first few population members in a population for which the size is at least n_p and at most $n_p + n_a$, where n_p is the size of the population in an archive–based approach and n_a is the size of the external archive.

Note that the notion of elitist solutions is also subject to the twofold goal in multi–objective optimization. On the one hand, we can choose to only maintain the non–dominated solutions, as is usually done. On the other hand, since diversity is also important, it is a valid choice to let the elitist set be equal to the set of solutions that was selected for exploration. This set may very well have been chosen to contain more diversity than is contained when merely selecting the non–dominated solutions. This approach allows elitism to incorporate solutions based on their added value to diversity as well. Clearly, care must be taken that most of this diversity contributes to diversity *along* the front of non–dominated solutions, but otherwise such twofold elitism corresponds directly to the twofold multi–objective goal. Moreover, seen in this way, elitism directly contributes to exploitation as it determines which solutions are certainly selected to survive a generation.

Elitism plays an important role in multi–objective optimization since many solutions exist that are all equally preferable. It is important to have access to many of them during optimization to advance the complete set of non–dominated solutions further. An ideal variation operator is capable of generating solutions that are better in the proximity sense across the entire current set of non–dominated solutions as well as possibly outside it to extend the diversity of the set of non–dominated solutions even further. However, obtaining new and diverse non–dominated solutions is hard, especially as the set of non–dominated solutions approaches the Pareto optimal front. If a non–dominated solution gets lost in a certain generation, it may take quite some effort before a new non–dominated solution in its vicinity is generated again. For this reason, elitism is commonly accepted [4], [13] to be a very important tool for improving the results obtained by any MOEA. Seen in this way, elitism also helps in exploration since it allows to preserve good solutions which are hard to generate for the exploration operator.

IV. A GENERAL FRAMEWORK FOR MULTI–OBJECTIVE EVOLUTIONARY ALGORITHMS BASED ON THE MOST IMPORTANT COMPONENTS FOR BALANCING PROXIMITY AND DIVERSITY

In Section IV-A we briefly discuss a few of the most prominent multi–objective evolutionary algorithms and point out how proximity and diversity exploitation are balanced. With only a single exception, none of these algorithms are capable of tuning the amount of diversity exploitation versus proximity exploitation.

In Section IV-B we present a general framework that contains the most important components for building competent MOEAs and point out how the trade–off goal between proximity and diversity is addressed by making different choices in this framework.

In Section IV-C, we give an example instance of the general framework presented in the previous section. In this example instance, we use a single control parameter to define the ratio between proximity exploitation and diversity exploitation. For diversity exploitation, we use a heuristic based on nearest neighbor information.

A. Existing multi–objective evolutionary algorithms

One of the most important aspects that caused the pioneering MOEAs such as the VEGA (*Vector Evaluated Genetic Algorithm*) by Schaffer [34], the approach by Fonseca and Fleming [17], the NPGA (*Niched Pareto Genetic Algorithm*) [35] and the NSGA (*Non–dominated Sorting Genetic Algorithm*) by Srinivas and Deb [36], to perform inferior to more recent MOEAs, is the absence of elitism.

The current state–of–the–art in multi–objective evolutionary optimization is represented by a Pareto set of different MOEAs, which include the NSGA–II by Deb *et al.* [19], the SPEA by Zitzler and Thiele [3], the SPEA–II by Zitzler *et al.* [37], the PAES (*Pareto Archived Evolution Strategy*) by Knowles and Corne [38], the M–PAES (*Memetic PAES*) by Knowles and Corne [39] and the $\mathbb{M}\mathbb{I}\mathbb{D}\mathbb{E}\mathbb{A}$ (*Multi–objective Mixture–based Iterated Density Estimation Evolutionary Algorithm*) by Thierens and Bosman [33], [7]. These MOEAs differ both in exploitation as well as in exploration. Although their multi–objective frameworks are defined apart from the actual exploration operators that can be used to construct a specific MOEA, specific exploration operators have often been associated with them in the literature. In NSGA–II and SPEA, binary encodings and standard crossover operators have often been applied. For NSGA–II, real–valued variables and the SBX operator have also been used. In PAES and M–PAES, the evolution strategy is used and in the $\mathbb{M}\mathbb{I}\mathbb{D}\mathbb{E}\mathbb{A}$, learning and sampling from probabilistic (mixture) models is mainly used. Each of these exploration mechanisms can be used in each of the MOEAs, resulting in more similar approaches. It is therefore more interesting to focus on the differences between the selection and elitism strategies in these MOEAs.

In the NSGA–II, the solutions in the population are sorted using rank–based non–domination, after which all ranks are included in a preselection up to a size of $\frac{1}{2}n$. Using a diversity selection approach, the last rank that will cause more than $\frac{1}{2}n$ solutions to be included in the preselection is filtered to ensure a preselection of the right size. This preselection is used to apply tournament selection and recombination so as to generate $\frac{1}{2}n$ new solutions. The tournament selection operator compares two individuals first on their domination rank and secondly on how crowded they are as explained in Section III-B. Since the goal of diversity preservation is thus always secondary, the amount of diversity preservation cannot be tuned.

In the SPEA, the non–dominated solutions found so far are stored externally from the population. If the number of non–dominated solutions exceeds the size of this external storage, clustering is performed on the external storage in the objective space and some solutions are discarded from each cluster. Crossover and mutation are applied to solutions that are selected from both the population as well as the external storage. Selection is performed using tournament selection with a tournament size of 2. The most characteristic and profound item in SPEA is the way fitness is assigned before selection. Each solution in the external storage is assigned a strength proportional to the number of solutions it dominates in the population. Each solution in the population is assigned one plus the sum of the strengths of the solutions in the external storage that dominate it. The additional value of one is required to ensure that the externally stored solutions are always better (a lower strength is preferable). This mechanism prefers individuals near the set of non–dominated solutions and distributes them at the same time. Again, this is a choice in how much effort is devoted to diversity preservation and how much is devoted to the selection of non–dominated solutions. There is no means of tuning their ratio.

In the PAES, a population of non–dominated individuals is maintained. At any time, only a single solution is adapted. If the adaptation has led to an improvement in non–domination, it is included into the population and the dominated solutions are deleted. If the adaptation has led to a non–dominated solution that increases diversity, the new solution is either added to the population or replaces a current solution, depending on whether or not the maximum population size has already been reached. The improvement in diversity is measured using a grid. Grid locations with a lower niche count are preferred. Similar to the NSGA–II, acceptance is first based on non–domination, after which a certain diversity measure is used in case the domination rank is identical.

In the M–PAES, an external storage is used similar to SPEA. Local search is applied to each solution in the population, after which an acceptance criterion similar to the one used in PAES is used in combination with a selection of solutions from the external storage, to determine whether the locally searched solution should be recorded into the external storage. Recombination is applied to random combinations of solutions chosen from both the population as well as the external storage. Again, a grid structure is used to choose between two non–dominating solutions.

In none of these approaches, the exploitation of diversity can be tuned. The only way to spend more effort on diversity, is to use exploration operators that stimulate the generation of a wider spread set of non–dominated solutions, such as the clustering approaches in the $\mathbb{MIDEA}$. The only exception in the current state–of–the–art MOEAs, is the most recent variant of the $\mathbb{MIDEA}$ approach [7] in which a preselection is first made from the population, based solely on non–domination. The size of the preselection is larger than the final selection size. From the preselection, the final selection is made based solely on diversity. By increasing the ratio between the preselection size and the final selection size, the balance between effort spent on proximity exploitation and diversity exploitation can be tuned.

B. A general multi–objective algorithmic framework

Based on the current state–of–the–art MOEAs, a few items can be outlined that are of major importance when constructing competent MOEAs:

- Selection of better solutions should be done based both on non–domination as well as on diversity preservation, although non–domination is the most important since it is diversity *along* the objectives for a set of non–dominated solutions that we are interested in.
- Elitism should be used by saving the best solutions of the previous generation either using a fixed population size or a non–fixed population size (external archive approach).
- Diversity selection should be applied at least when too many non–dominated solutions are in the elitist population or in the external archive, so as to ensure diversity preservation along the set of non–dominated solutions.

Similar considerations have led to the definition of the UMMEA (*Unified Model for Multi–objective Evolutionary Algorithms*) by Laumanns *et al.* [14]. Although the UMMEA framework is important, the underlying message is different from the one in this paper. On the one hand, important considerations and choices, such as the use of elitism, lead to better MOEAs in general. To this end, a general framework such as the UMMEA framework is crucial for designing of new efficient MOEAs. On the other hand, the main point in this paper is that by making certain choices in such a framework, different MOEAs can be constructed that are incomparably good when aiming to satisfy both the goal of proximity as well as diversity. Furthermore, unless diversity exploitation and proximity exploitation are separated such that the ratio between them can be controlled, we have no means to explore even a subrange of the possible instances of the general framework. In the next section, we shall present one possible instance in which we do have control over this ratio and illustrate some results on two multi–objective optimization problems. Before we do so, we first present

1 $\mathcal{P} \leftarrow$ Generate a set of solutions randomly
2 Repeat until termination
2.1 $\mathcal{S} \leftarrow$ Select solutions from $\mathcal{P}$ based on non–dominance and diversity
2.2 $\mathcal{E} \leftarrow$ Select elitist solutions from $\mathcal{P}$
2.3 $\mathcal{O} \leftarrow$ Construct new solutions by applying an exploration operator to $\mathcal{S}$
2.4 $\mathcal{P} \leftarrow (\mathcal{E}, \mathcal{O})$

Fig. 4. A general framework for multi–objective evolutionary algorithms containing the most prominent components.

a general framework in which the current state–of–the–art MOEAs can be placed.

In Figure 4, a general framework is presented based on the considerations in this paper. First of all, it should be noted that we do not enforce a fixed population size, which allows for the modelling of external archives by using a subset of the population. First, a set of solutions should be selected. As explained, both proximity as well as diversity play an important role due to the composite goal in multi–objective optimization. Second, an elitist set is selected. This set does not have to be identical to the set used for exploration. In the archive–based approaches for instance, only the non–dominated solutions are explicitly saved, whereas the selected solutions may be chosen from both the archive as well as the remainder of the population. Exploration is applied to the set of selected solutions, after which the new population is constructed by combining the elitist solutions with the newly generated solutions. Furthermore, MOEAs that have been shown to guarantee global convergence in the limit, require that we can distinguish between old solutions (elitist ones) and the offspring [40], [41], [42]. To this end, it should be noted that the population, the elitist collection and the offspring collection cannot be sets since this would not allow for multiple occurrences of the same solution. Although allowing these collections to be multisets would solve this problem, it does not allow to distinguish the elitist solutions from the offspring once the new population is constructed. Therefore, we point out that these collections are actually vectors of solutions such that the elitist solutions in the new population can be found by inspecting the first $|\mathcal{E}|$ solutions of $\mathcal{P}$. This thus allows the MOEAs that guarantee global convergence also to be modelled by our general framework.

The current state–of–the–art MOEAs all fit the general framework in Figure 4. For NSGA–II for instance, the population is of size n, $\mathcal{S}$ is determined by first making a preselection of size $\frac{1}{2}n$ using truncation selection on the domination ranks, after which tournament selection based on domination ranks and diversity is used on the preselection to obtain the final selection. The elitist set equals the preselection set. For SPEA, the initial population is of size $n = n_p$. The archive equals the first n_a solutions in the population. On beforehand, $n_a = 0$. The solutions in $\mathcal{P}$ are assigned a fitness value as defined in the SPEA selection procedure, after which the actual selection takes place. The elitist set is exactly the first n_a solutions in the population. However, if n_a is larger than the maximum number of solutions that are allowed in the archive, a selection is made using the cluster–based pruning method. Exploration operators are used to generate a new population of size n_p which is joined with the elitist set.

The other state–of–the–art MOEAs can be fit into the general framework in a similar manner. These algorithms are the result of making choices for the components in the general framework that explicitly points out the trade–off to be made in selection and elitism between proximity and diversity. As a result, some algorithms are really better than others, as has been indicated in an empirical study [4]. However, with respect to the multi–objective goal of proximity versus diversity, the current state–of–the–art of these algorithms are mostly non–dominating and incomparable. As an example, a balance parameter can be set in the $\mathbb{MIDEA}$ framework that represents the amount of effort spent on diversity preservation. In the experiments that were performed using the $\mathbb{MIDEA}$, this parameter was set in such a way that a larger effort was made towards diversity preservation [7] than is usually the case in MOEAs. The resulting MOEA seems particularly well suited for diversity preservation along the set of non–dominated solutions. However, it does not always score equally good in getting close to the Pareto optimal front. An illustration of the influence of making such choices between diversity preservation and non–domination selection is given in the next section as an example of making choices within of the general framework in Figure 4.

C. Balancing proximity and diversity exploitation: An example instance of the general framework

In this section we present an example instance of the general framework for MOEAs in which we are able to shift the balance between proximity and diversity in the resulting approximation set through a single parameter and use it to experimentally indicate the multi–objective trade–off. In Section IV-C.1 we first describe our example instance. Next, in Section IV-C.2 we describe the multi–objective optimization problems that we have used to perform our experiments with. Finally, in Section IV-C.3 we perform a few experiments that illustrate the existence of the trade–off between proximity and diversity by varying the bias of our example EA towards proximity and diversity.

1) Our example instance: If we would have access to the Pareto optimal set, a good and robust heuristic based on nearest neighbor information could be used to find a representative and diverse subset of a predefined size [7]. First, an individual with a maximum value for an arbitrary objective is deleted from $\boldsymbol{\mathcal{P}_S}$ and added to the selected set $\mathcal{S}$. Ties are broken by iteratively considering other objectives. For all solutions in $\boldsymbol{\mathcal{P}_S}$, the nearest neighbor distance is computed to the single solution in $\mathcal{S}$. Different types of distance metrics can be used here, such as for instance the Euclidean distance scaled to the sample range in each objective. The solution in $\boldsymbol{\mathcal{P}_S}$ with the *largest* distance is then deleted from $\boldsymbol{\mathcal{P}_S}$ and added to $\mathcal{S}$. The distances in $\boldsymbol{\mathcal{P}_S}$ are updated by investigating whether the distance to the newly added point in $\mathcal{S}$ is smaller than the currently stored distance. These last two steps are repeated until the desired number of solutions are in the final selection. This diversity selection operator has a running time complexity of $\mathcal{O}(|\boldsymbol{\mathcal{P}_S}|^2)$.

Unfortunately, we are of course not in the luxurious position of having access to $\boldsymbol{\mathcal{P}_S}$ in a real–life situation. However, this diversity preserving selection operator can be combined with a non–domination selection operator in such a way that we have control over how much diversity preservation is done. One approach is to first make a preselection $\mathcal{S}^P$ from the population using for instance truncation selection on the domination count. Subsequently, the final selection is made by selection solutions from the preselection using the nearest neighbor heuristic. The size of the preselection is $\lfloor \delta\tau n \rfloor$, where n is the population size and $\lfloor \tau n \rfloor$ is the finally desired selection size ($\tau \in [\frac{1}{n}; 1[, \delta \in [1; \frac{1}{\tau}]$). Pseudo–code is given in Figure 5. If δ is increased, a more diverse selection is possible since more solutions are available. However, the probability that a non–dominated solution is *not* selected, also increases as δ is increased. The reason for this is that although the non–dominated solutions are included in the preselection, there is no guarantee that they will be included in the final selection if we only select solutions based on their diversity. As a result, the δ parameter is a control parameter that determines the amount of diversity that may be preserved during multi–objective evolutionary optimization. Although the $\boldsymbol{D}_{\mathcal{S}\to\boldsymbol{\mathcal{P}_F}}$ indicator will most likely increase as δ is increased, since the non–domination pressure reduces, the preserved diversity will also most likely increase. As a result, the $\boldsymbol{D}_{\boldsymbol{\mathcal{P}_F}\to\mathcal{S}}$ indicator is probably similar a range of values for δ, which reflects the trade–off between proximity and diversity.

It is important to note that if the solution with the largest domination count to end up in $\mathcal{S}^P$ by truncation selection has a domination count of 0, *all* individuals with a domination count of 0 should be selected instead, resulting in $|\mathcal{S}^P| \geq \lfloor \delta\tau n \rfloor$. This ensures that once the search starts to converge onto a certain set of non–dominated solutions, we enforce diversity over all of the available solutions in the set of non–dominated solutions. If we do not do this, we are likely to quickly loose the ability to properly preserve diversity along the set of non–dominated solutions since not all regions of the set of non–dominated solutions may be properly represented in the final selection. Pseudo–code for this example instance is given in Figure 5.

It is interesting to note that most state–of–the–art MOEAs are somewhat similar to our example approach for $\delta = 1$. Compared to NSGA–II for instance, the set of selected solutions equals the best non–dominated solutions, especially if rank–based non–dominated selection is used. Diversity filtering is applied if the number of non–dominated solutions becomes larger than $\lfloor \tau n \rfloor$, which also happens in the NSGA–II. The elitist solutions are the same as the selected solutions, which is also similar to NSGA–II. The main difference is the additional selection step performed by NSGA–II when generating new offspring.

2) Multi–objective optimization test problems: The problems that we have used to illustrate the behavior of our example instance on, are given in Figure 6.

Problem ZDT_4 was introduced by Zitzler *et al.* [4] and is defined by a function of real–valued variables. It is very hard to obtain the optimal front $f_1(\boldsymbol{y}) = 1 - \sqrt{y_0}$ in ZDT_4 since there are many local fronts.

The multi–objective knapsack problem was first used to test MOEAs on by Zitzler and Thiele [3]. We are given m knapsacks with a specified capacity and n items. Each item can have a different weight and profit in every knapsack. Selecting item i in a solution implies placing it in every knapsack. For a solution to be feasible, the capacity of each knapsack may not be exceeded.

In the set covering problem, we are given l locations at which we can place some service at a specified cost. Furthermore, associated with each location is a set of regions that is a subset of $\{0, 1, \ldots r-1\}$ that can be serviced from that

```
1 P ← Generate a set of n solutions randomly
2 Repeat until termination
    2.1 Select solutions:
        (a) S^P ← Select the best ⌊δτn⌋ solutions
                   from P based on non–domination
        (b) S ← Select ⌊τn⌋ solutions from S^P using
                 the nearest neighbor heuristic
    2.2 E ← S
    2.3 O ← Construct n − ⌊τn⌋ new solutions by
             applying an exploration operator
             to S
    2.4 P ← (E, O)
```

Fig. 5. Combining selection based solely on non–domination and selection based solely on diversity preservation using a single trade–off control parameter δ in a multi–objective EA framework.

ZDT_4	
Minimize	$(f_0(\boldsymbol{y}), f_1(\boldsymbol{y}))$
Where	• $f_0(\boldsymbol{y}) = y_0$ • $f_1(\boldsymbol{y}) = \gamma\left(1 - \sqrt{\frac{f_0(\boldsymbol{y})}{\gamma}}\right)$ • $\gamma = 1 + 10(l–1) + \sum_{i=1}^{l-1}\left(y_i^2 - 10\cos(4\pi y_i)\right)$ • $y_0 \in [0, 1]$ • $y_i \in [-5, 5] \quad (1 \leq i < l)$
Knapsack	
Maximize	$(f_0(\boldsymbol{x}), f_1(\boldsymbol{x}), \ldots, f_{m-1}(\boldsymbol{x}))$
Where	• $\forall i \in \boldsymbol{\mathcal{M}} : f_i(\boldsymbol{x}) = \sum_{j=0}^{l-1} P_{ij}x_j$
Such That	• $\forall i \in \boldsymbol{\mathcal{M}} : \sum_{j=0}^{l-1} W_{ij}x_j \leq c_i$
Set Covering	
Minimize	$(f_0(\boldsymbol{x}), f_1(\boldsymbol{x}), \ldots, f_{m-1}(\boldsymbol{x}))$
Where	• $\forall i \in \boldsymbol{\mathcal{M}} : f_i(\boldsymbol{x}) = \sum_{j=0}^{l-1} C_{ij}x_j$
Such That	• $\forall i \in \boldsymbol{\mathcal{M}} : \forall 0 \leq j < r : \sum_{k=0}^{l-1}(A_i)_{jk}x_k \geq 1$

Fig. 6. Multi–objective optimization illustration problems.

location. The goal is to select locations such that *all* regions are serviced against minimal costs. In the multi–objective variant of set covering, m services are placed at a location. Each service however covers its own set of regions when placed at a certain location and has its own cost associated with a certain location. A binary solution indicates at which locations the services are placed.

We used $l = 10$ variables for the *ZDT*$_4$ problem and $l = 100$ variables for the knapsack and set covering problems. We allowed a maximum of $20 \cdot 10^3$ evaluations in any single run in all our experiments. As a result of imposing the restriction of a maximum of evaluations, a value for the population size n exists for each MOEA such that the MOEA will perform best. For too large population sizes, the search will move towards a random search and for too small population sizes, there is not enough information to adequately and competently generate new good solutions. We therefore increased the population size in steps of 25 to find the best results. To select the best population size, we used the result with the lowest $\boldsymbol{D}_{\mathcal{P}_F \rightarrow \mathcal{S}}$ indicator value.

For the knapsack problem, we generated an instance by generating random weights in $[1; 10]$ and random profits in $[1; 10]$. The capacity of a knapsack was set at half of the total weight of all the items, weighted according to that knapsack objective. For set covering, the costs were generated at random in $[1; 10]$. We used 250 regions to be serviced. We set the problem difficulty through the region–location adjacency relation. Each location was made adjacent to 70 randomly selected regions.

The binary problems have constraints. To deal with them, we can use a repair mechanism to transform infeasible solutions into feasible solutions. Another approach is introduced by the notion of constraint–domination introduced by Deb *et al.* [43]. This notion allows to deal with constrained multi–objective problems according to a very general scheme. A solution $\boldsymbol{x}$ is said to *constraint–dominate* solution $\boldsymbol{y}$ if any of the following is true:

1) Solution $\boldsymbol{x}$ is *feasible* and solution $\boldsymbol{y}$ is *infeasible*
2) Solutions $\boldsymbol{x}$ and $\boldsymbol{y}$ are both *infeasible*, but $\boldsymbol{x}$ has a smaller overall constraint violation
3) Solutions $\boldsymbol{x}$ and $\boldsymbol{y}$ are both *feasible* and $\boldsymbol{x} \succ \boldsymbol{y}$

In the above definition, the overall constraint violation is the amount by which a constraint is violated, summed over all constraints. We have used this principle for set covering. For knapsack we have used a repair mechanism that was proposed in earlier MOEA research [3]. If a solution violates a constraint, the repair mechanism iteratively removes items until all constraints are satisfied. The order in which the items are investigated, is determined by the maximum profit/weight ratio. The items with the lowest profit/weight ratio are removed first.

3) Experimental illustrations of the trade–off between proximity and diversity: For each problem, we used one–point crossover with a probability of 0.8 in combination with bit flipping mutation with a probability of 0.01. For the real–valued *ZDT*$_4$ problem, we encoded every variable with 30 bits. We applied our example instance of the general framework using both domination counting and domination ranking to determine the preselection. We have varied the value of δ from 1 to 3 in steps of 0.25 and have kept τ fixed at 0.3. Figure 7 shows the the resulting values on each problem for the four different performance indicators from Section II-B obtained with the population size that resulted in the best $\boldsymbol{D}_{\mathcal{P}_F \rightarrow \mathcal{S}}$ indicator value, averaged over 10 runs.

The results for the two different non–domination preselection approaches do not differ much. The behavior with respect to the different performance indicators on each problem is similar. In the remainder of our illustrations we shall therefore only use the domination count approach.

We already argued that for $\delta = 1$, we have an approach that is quite similar to NSGA–II, which is a representative of the current state–of–the–art MOEAs. The results obtained for $\delta = 1$ are indeed comparable to those obtained by the NSGA–II on the same test problems, the results of which can be found elsewhere [7]. For small values of δ, the ability to find solutions close to the Pareto optimal front worsens as δ is increased and thus more effort is spent on diversity preservation. This can be seen in the figure for the $\boldsymbol{D}_{\mathcal{S} \rightarrow \mathcal{P}_F}$ indicator value. Furthermore, the number of solutions on the front rapidly drops to lower values since in our example instance elitism does not always maintain the non–dominated solutions. Still, the added effort spent on diversity does pay off in a certain way, since the diversity as measured by the front spread indicator increases as δ is increased. The most interesting results can be seen in the figure that displays the $\boldsymbol{D}_{\mathcal{P}_F \rightarrow \mathcal{S}}$ indicator value. For quite a large range of values for δ, the indicator value does not worsen, but sometimes becomes even better. Within this range, the trade–off between diversity preservation along the set of non–dominated solutions and the proximity of non–dominated solutions with respect to the Pareto optimal front is the most interesting. With respect to the performance indicator used, there is a certain optimal value. However, this performance indicator only reflects a certain balance between the two goals. Since the average distance to the front only worsens and the front spread only increases, most different settings for the algorithm do not outperform each other if we have no preference for these two goals. Outperformance can only be detected if δ becomes very large. In that case, the front spread increases slightly but the average distance of each point in the resulting approximation set to the Pareto optimal front increases very much. These observations regarding outperformance are confirmed by the results in Figure 8. The results in this figure show for the use of one–point crossover the most frequently occurring relation from the categorization of Zitzler *et al.* [11] when comparing the approximation sets of a MOEA using one value for δ with another value for δ. Indeed, in almost all cases, the approximation sets are most frequently categorized as incomparable. Only for $\delta = 3.0$ there are some cases in which we can speak of true outperformance. Moreover, within this large set of incomparable MOEAs, the distance to the Pareto front worsens monotonically as δ is increased, but the front spread improves monotonically as δ is increased. As we argued earlier in Section II-B.2, the additional information based on

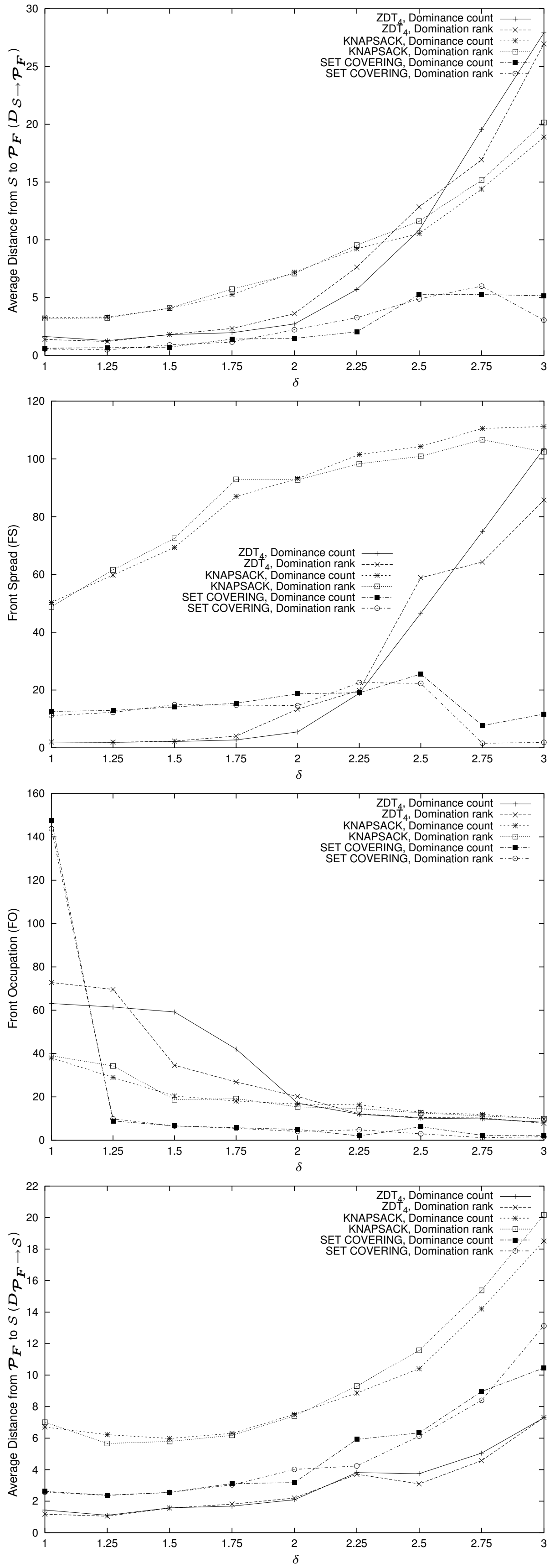

Fig. 7. Results for the ZDT_4, the knapsack and the set covering problems using one–point crossover in the example instance of the general elitist framework. The results measured in four different performance indicators are shown as a function of δ.

Knapsack, one–point crossover

Knapsack, one–point crossover

δ	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
1.00	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
1.25	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
1.50	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
1.75	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\succ\succ$
2.00	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\succ\succ$
2.25	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
2.50	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
2.75	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
3.00	$\parallel$	$\parallel$	$\parallel$	$\prec\prec$	$\prec\prec$	$\parallel$	$\parallel$	$\parallel$	$\parallel$

Fig. 8. Comparing all combinations of the results of using one–point crossover for all tested values of δ on the multi–objective knapsack problem. The entries in the table represent row–versus–column relations.

the performance indicators leads us to conclude that there truly is no preference for any of the MOEAs that are classified as being incomparable. However, the classification in Figure 8 becomes less certain as the value of δ is increased. The reason for this is that not all combinations of approximation sets over the different runs are classified as being incomparable. As δ is increased, the frequency of the classification of being incomparable gets closer to the classification of being better or even being strictly dominating. If these frequencies become very close, we might already find one MOEA preferable over another. Intuitively, all of this is reflected by the large, relatively flat part in the $D_{\mathcal{P}_F\rightarrow S}$ performance indicator. We truly have no preference for any of the MOEAs that correspond to this flat part in the graph. However, as this indicator value increases, a preference starts to be formed towards MOEAs with a lower value for the $D_{\mathcal{P}_F\rightarrow S}$ indicator. Indeed, this already happens for smaller values of δ than $\delta = 3$. This corresponds to the observation of the decrease of the frequency of the incomparable classification. In general, making modest choices on whether we spend more effort on diversity preservation or on proximity leads to MOEAs that have a performance such that we truly do not prefer one MOEA over another.

In Figure 9 additional results are shown for the knapsack problem using different variation methods than just the one–point crossover recombination operator together with bit–wise mutation. Each time a selection of solutions was made, a probability distribution was estimated over the selected solutions in the parameter space. Using the estimated probability distribution, we drew new samples that serve as the offspring solutions. The first type of probability distribution that we estimate, is the univariately factorized probability distribution or univariate factorization for short in which each random variable is assumed to be independent of each other random variable. The second type of probability distribution that we estimate, is the tree–structured Bayesian factorization. Such a factorization can be estimated optimally using the optimal dependency tree algorithm by Chow and Liu [44]. For more details on the use of probability distributions as a variation operator, we refer the interested reader to specialized literature [25], [26], [27], [28], [29], [30], [31], [32], [7]. By clustering the selected solutions in the objective space before estimating a probability distribution in each cluster, a special mixture probability distribution is constructed that stimulates a diverse exploration. We have used the leader clustering algorithm in the objective space such that four clusters were constructed on average. If the number of clusters becomes too large, the requirements for the population size increases to facilitate proper factorization selection in each cluster. We do not suggest that the number of clusters we use is optimal, but it will serve to indicate the effectiveness of parallel exploration as well as diversity preservation.

A similar behavior is observed for the different variation operators as observed for only one–point crossover using different selection strategies on all three problems. However, one interesting additional phenomenon can be seen in the graphs in Figure 9. The results obtained for the approaches that use clustering in the objective space, have an *intrinsically* better performance with respect to the front spread performance indicator than any other method. For all values of δ, both the front spread as well as the distance to the Pareto optimal front are better if clustering is used to construct a mixture of univariate factorizations instead of the univariate factorization. This is confirmed only in part by the classification results in Figure 10. Indeed, for a larger variety of values for δ, the univariate factorization is outperformed by the mixture of univariate factorizations. However, based on the results provided by the performance indicators, one would expect the figure to show that for all values of δ, the MOEA that uses the univariate factorization is outperformed by the MOEA that uses the mixture of univariate factorizations. However, the majority of the comparisons result in the classification of being incomparable. Whereas in the case when we were comparing one–point crossover with itself, the classifications of being incomparable were a result of approximation sets that are in most cases not preferable over one another such as in Figure 3, in this case the classifications of being incomparable are a result of cases such as the one in Figure 2. The performance indicators now show that we can truly speak of a preference for using the mixture of univariate factorizations over the univariate factorization for the multi–objective knapsack problem since they are all in favor of the mixture of univariate factorizations. The added use of clustering seems to lead to more advanced MOEAs than when clustering is not used to stimulate parallel exploration. This example serves to show that although there is an intrinsic trade–off in the choices that are to be made in the general framework, this does not imply that we cannot make some general choices that lead to intrinsically better MOEAs. On the other hand, within for instance the use of a mixture model, by changing the value of δ, again different choices for spending more or less effort on diversity preservation are made. Similar arguments and comparison classifications can be made to show again that the performance for the different goals in multi–objective optimization when using clustering in the objective space are mostly incomparable, which indicates that the trade–off between diversity preservation and proximity is still present, even if intrinsically better variation operators are used.

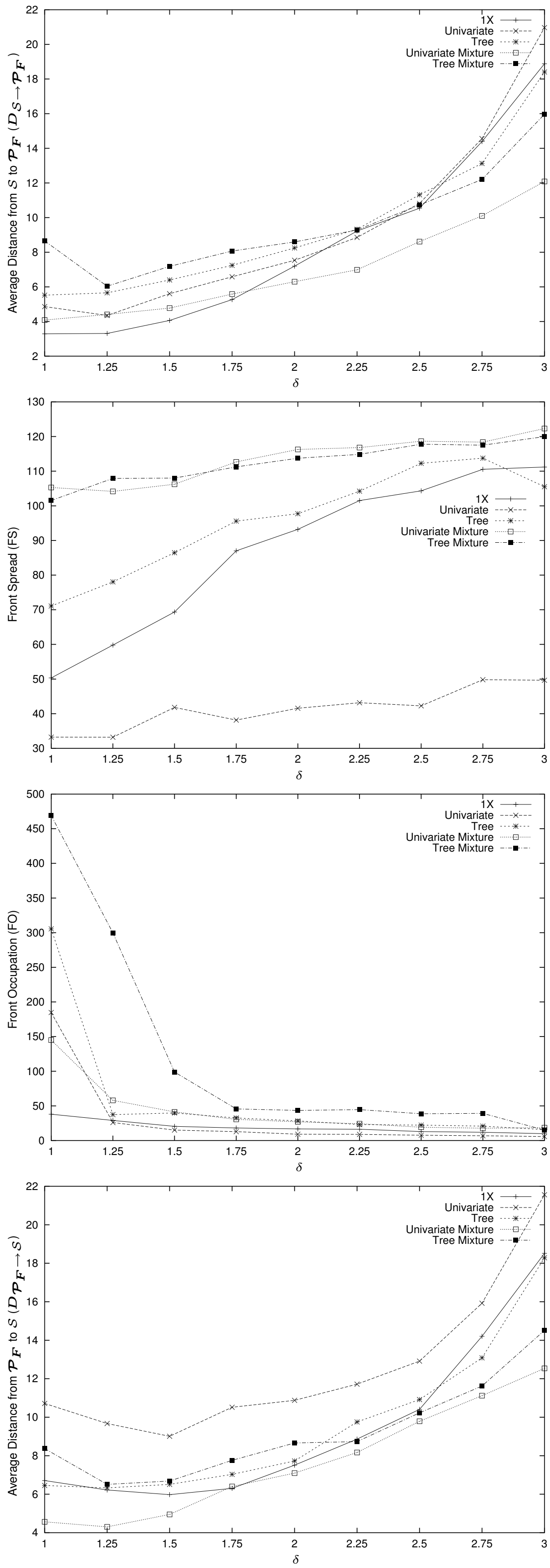

Fig. 9. Additional results for the knapsack problem using different variation operators in the example instance of the general elitist framework. Selection is based on the domination count. The results measured in four different evaluation metrics are shown.

Knapsack, mixture of univariate factorizations

Knapsack, univariate factorization

δ	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
1.00	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
1.25	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
1.50	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
1.75	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
2.00	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
2.25	$\parallel$	$\prec\prec$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
2.50	$\prec\prec$	$\prec\prec$	$\prec\prec$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
2.75	$\prec\prec$	$\prec\prec$	$\prec\prec$	$\prec\prec$	$\prec\prec$	$\parallel$	$\parallel$	$\parallel$	$\parallel$
3.00	$\prec\prec$	$\prec\prec$	$\prec\prec$	$\prec\prec$	$\prec\prec$	$\prec\prec$	$\prec\prec$	$\prec\prec$	$\parallel$

Fig. 10. Comparing all combinations of the results of using the univariate factorization for all tested values of δ with the result of using the mixture of univariate factorizations for all tested values of δ on the multi–objective knapsack problem. The entries in the table represent row–versus–column relations.

There is one more interesting thing to be observed in Figure 9. The use of tree–structured Bayesian factorizations is a more involved method than using univariate factorizations, since the latter approach is quite similar to using uniform crossover with a crossover probability of 1.0 and no mutation. However, the use of the univariate factorization is still capable of producing solutions that are closer to the Pareto optimal front. One of the reasons for this is that the approach based on estimating tree–structured Bayesian factorizations is capable of generating solutions at locations in which the less involved variation operators are not capable of generating new solutions. This can be seen in the figure for the front occupation, since the tree–structured Bayesian factorization approach is capable of generating more non–dominated solutions than the less involved variation operators. Also, a larger front–spread is obtained using the tree–structured Bayesian factorization because of a more effective variation of the selected solutions. As a result of this, a better approximation of the Pareto optimal front is more likely to be obtained, but more evaluations may be required because a larger set of solutions may be found than is possible using the less involved operators. This is another important aspect to consider when evaluating MOEAs.

Before we end this empirical section of the paper, we note that in our example instance of the general framework, the δ parameter is used to control the combination of non–domination selection and diversity selection. However, this parameter also controls the combination of proximity and diversity that is offered to the exploration operator. It would also be interesting to see the results if we have one δ parameter for determining the ratio between proximity and diversity for selection and elitism and a different δ parameter for determining the ratio between proximity and diversity for the selection that is offered to the exploration part of the MOEA. In this case, finer grained control and insights would be obtained for determining the influence of added effort on diversity for exploitation and separately for exploration. In our example instance, these influences are linked together since the selected solutions are also used for elitism, which may quickly cause many non–dominated solutions to be lost as δ is increased.

V. Conclusions

In this paper, we have argued that the quest for finding the components that result in the best EAs for multi–objective optimization is not likely to converge to a single, specific MOEA. The intrinsic trade–off between the goals of proximity and diversity preservation plays a prominent role in the exploitation and exploration phases of any MOEA. By making choices on how to effectively attend to both goals, very effective MOEAs may be constructed. When shifting these choices more towards proximity or more towards diversity preservation, a different performance will be obtained that is not inferior with respect to plausible performance indicators that measure proximity with respect to the Pareto optimal front and diversity along the finally obtained set of non–dominated solutions.

Although the existence of the trade–off and consequently the existence of such choices to be made is important, it does not imply that the current state–of–the–art MOEAs cannot be improved any further. It only argues that such choices will always remain. This trade–off should therefore always be kept in mind when designing new MOEAs and when comparing the experimental results of different MOEAs. Specifically, if we are able to separate the effort spent on diversity from the effort spent on obtaining non–dominated solutions such that their ratio can be controlled, a MOEA can be constructed that is more of a meta–type. Depending on the demands of the final decision maker, such a MOEA is capable of dealing with the trade–off goals in multi–objective optimization by adjusting the ratio.

References

[1] C. A. Coello Coello, "A comprehensive survey of evolutionary-based multiobjective optimization techniques," *Knowledge and Information Systems. An International Journal*, vol. 1, no. 3, pp. 269–308, 1999.

[2] S. Obayashi, S. Takahashi, and Y. Takeguchi, "Niching and elitist models for MOGAs," in *Parallel Problem Solving from Nature – PPSN V*, A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, Eds. 1998, pp. 260–269, Springer-Verlag.

[3] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.

[4] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evolutionary Computation*, vol. 8, no. 2, pp. 173–195, 2000.

[5] J. B. Zydallis, D. A. van Veldhuizen, and G. B. Lamont, "A statistical comparison of multiobjective evolutionary algorithms including the MOMGA–II," in *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization – EMO 2001*, E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, and D. Corne, Eds. 2001, pp. 226–240, Springer-Verlag.

[6] T. Hanne, "Global multiobjective optimization with evolutionary algorithms: Selection mechanisms and mutation control," in *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization – EMO 2001*, E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, and D. Corne, Eds. 2001, pp. 197–212, Springer-Verlag.

[7] P. A. N. Bosman and D. Thierens, "Multi–objective optimization with diversity preserving mixture–based iterated density estimation evolutionary algorithms," *International Journal of Approximate Reasoning*, vol. 31, pp. 259–289, 2002.

[8] V. G. da Fonseca, C. M. Fonseca, and A. O. Hall, "Inferential performance assessment of stochastic optimisers and the attainment function," in *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization – EMO 2001*, E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, and D. Corne, Eds. 2001, pp. 213–225, Springer-Verlag.

[9] J. Knowles and D. Corne, "On metrics for comparing non–dominated sets," in *Proceedings of the 2002 Congress on Evolutionary Computation CEC 2002*. 2002, pp. 666–674, IEEE Press.

[10] M. P. Hansen and A. Jaszkiewicz, "Evaluating the quality of approximations of the non–dominated set," Technical report, Institute of Mathematical Modeling, Technical University of Denmark IMM–REP–1998–7, 1998.

[11] E. Zitzler, M. Laumanns, L. Thiele, C. M. Fonseca, and V. Grunert da Fonseca, "Why quality assessment of multiobjective optimizers is difficult," in *Proceedings of the GECCO–2002 Genetic and Evolutionary Computation Conference*, W. B. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke, and N. Jonoska, Eds. 2002, pp. 666–674, Morgan Kaufmann.

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

[13] M. Laumanns, E. Zitzler, and L. Thiele, "On the effects of archiving, elitism, and density based selection in evolutionary multi–objective optimization," in *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization – EMO 2001*, E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, and D. Corne, Eds. 2001, pp. 181–197, Springer-Verlag.

[14] M. Laumanns, E. Zitzler, and L. Thiele, "A unified model for multi–objective evolutionary algorithms with elitism," in *Proceedings of the 2000 Congress on Evolutionary Computation*. 2000, pp. 46–53, IEEE Press.

[15] E. Zitzler, "Evolutionary algorithms for multiobjective optimization: Methods and applications," PhD thesis, Swiss Federal Institute of Technology (ETH) Zurich, Switzerland., 1999.

[16] D. A. Van Veldhuizen, "Multiobjective evolutionary algorithms: Classifications, analyses, and new innovations.," PhD thesis, Graduate School of Engineering of the Air Force Institute of Technology., 1999.

[17] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multiobjective optimization," *Evolutionary Computation*, vol. 3, no. 1, pp. 1–16, 1995.

[18] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison–Wesley, Reading, 1989.

[19] K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II," in *Parallel Problem Solving from Nature – PPSN VI*, M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H.-P. Schwefel, Eds. 2000, pp. 849–858, Springer.

[20] D. E. Goldberg, B. Korb, and K. Deb, "Messy genetic algorithms: Motivation, analysis and first results," *Complex Systems*, vol. 10, pp. 385–408, 1989.

[21] D. E. Goldberg, K. Deb, H. Kargupta, and G. Harik, "Rapid, accurate optimization of difficult problems using fast messy genetic algorithms," in *Proceedings of the Fifth International Conference on Genetic Algorithms*, S. Forrest, Ed. 1993, pp. 56–64, Morgan Kauffman.

[22] S. Bandyopadhyay, H. Kargupta, and G. Wang, "Revisiting the GEMGA: Scalable evolutionary optimization through linkage learning," in *Proceedings of the 1998 IEEE International Conference on Evolutionary Computation*. 1998, pp. 603–608, IEEE Press.

[23] F. G. Lobo, K. Deb, D. E. Goldberg, G. R. Harik, and L. Wang, "Compressed introns in a linkage learning genetic algorithm," in *Genetic Programming 1998: Proceedings of the Third Annual Conference*, W. Banzhaf, K. Chellapilla, K. Deb, M. Dorigo, D. B. Fogel, M. H. Garzon, D. E. Goldberg, H. Iba, and R. Riolo, Eds. 1998, pp. 551–558, Morgan Kauffman.

[24] C. H. M. van Kemenade, "Building block filtering and mixing," in *Proceedings of the 1998 IEEE International Conference on Evolutionary Computation*. 1998, pp. 505–510, IEEE Press.

[25] P. Larrañaga and J. A. Lozano, *Estimation of Distribution Algorithms. A New Tool for evolutionary computation.*, Kluwer Academic Publishers, 2001.

[26] R. Etxeberria and P. Larrañaga, "Global optimization using Bayesian networks," in *Proceedings of the Second Symposium on Artificial Intelligence CIMAF–1999*, A. A. O. Rodriguez M. R. S. Ortiz and R. S. Hermida, Eds. 1999, pp. 332–339, Institute of Cybernetics, Mathematics and Physics.

[27] G. Harik, "Linkage learning via probabilistic modeling in the ECGA," IlliGAL Technical Report 99010, 1999.

[28] H. Mühlenbein and T. Mahnig, "FDA – a scalable evolutionary algorithm for the optimization of additively decomposed functions," *Evolutionary Computation*, vol. 7, pp. 353–376, 1999.

[29] M. Pelikan, D. E. Goldberg, and E. Cantú-Paz, "BOA: The Bayesian optimization algorithm," in *Proceedings of the GECCO–1999 Genetic and Evolutionary Computation Conference*, W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela, and R. E. Smith, Eds. 1999, pp. 525–532, Morgan Kauffman.

[30] M. Pelikan, D. E. Goldberg, and K. Sastry, "Bayesian optimization algorithm, decision graphs and occam's razor," in *Proceedings of the GECCO–2001 Genetic and Evolutionary Computation Conference*, L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon, and E. Burke, Eds. 2001, pp. 519–526, Morgan Kaufmann.

[31] A. Ochoa, H. Mühlenbein, and M. Soto, "A factorized distribution algorithm using single connected Bayesian networks," in *Parallel Problem Solving from Nature – PPSN VI*, M. Schoenauer *et al.*, Ed. 2000, pp. 787–796, Springer.

[32] P. A. N. Bosman and D. Thierens, "Advancing continuous IDEAs with mixture distributions and factorization selection metrics," in *Proceedings of the Optimization by Building and Using Probabilistic Models OBUPM Workshop at the Genetic and Evolutionary Computation Conference GECCO–2001*, M. Pelikan and K. Sastry, Eds. 2001, pp. 208–212, Morgan Kaufmann Publishers.

[33] D. Thierens and P. A. N. Bosman, "Multi–objective mixture–based iterated density estimation evolutionary algorithms," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon, and E. Burke, Eds., San Francisco, California, 2001, pp. 663–670, Morgan Kaufmann.

[34] J. D. Schaffer, "Multiple objective optimization with vector evaluated genetic algorithms," in *Genetic Algorithms and their Applications: Proceedings of the First International Conference on Genetic Algorithms*, J. J. Grefenstette, Ed. 1985, pp. 93–100, Lawrence Erlbaum.

[35] J. Horn, N. Nafpliotis, and D. E. Goldberg, "A niched pareto genetic algorithm for multiobjective optimization," in *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, J. J. Grefenstette, Ed. 1994, vol. 1, pp. 82–87, IEEE Press.

[36] N. Srinivas and K. Deb, "Multiobjective optimization using nondominated sorting in genetic algorithms," *Evolutionary Computation*, vol. 2, no. 3, pp. 221–248, 1994.

[37] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization," in *EUROGEN 2001 - Evolutionary Methods for Design, Optimisation and Control with Applications to Industrial Problems*, 2001.

[38] J. D. Knowles and D. W. Corne, "The pareto archived evolution strategy: A new baseline algorithm for multiobjective optimisation," in *Proceedings of the 1999 Congress on Evolutionary Computation*. 1999, pp. 98–105, IEEE Press.

[39] J. D. Knowles and D. W. Corne, "M–PAES: A memetic algorithm for multiobjective optimization," in *Proceedings of the 2000 Congress on Evolutionary Computation*. 2000, pp. 325–332, IEEE Press.

[40] T. Hanne, "On the convergence of multiobjective evolutionary algorithms," *European Journal of Operational Research*, vol. 117, no. 3, pp. 553–564, 1999.

[41] G. Rudolph and A. Agapie, "Convergence properties of some multi-objective evolutionary algorithms," in *Proceedings of the 2000 Congress on Evolutionary Computation CEC00*. 2000, pp. 1010–1016, IEEE Press.

[42] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Archiving with guaranteed convergence and diversity in multi–objective optimization," in *Proceedings of the GECCO–2002 Genetic and Evolutionary Computation Conference*, W. B. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke, and N. Jonoska, Eds. 2002, pp. 439–447, Morgan Kaufmann.

[43] K. Deb, A. Pratap, and T. Meyarivan, "Constrained test problems for multi–objective evolutionary optimization," in *First International Conference on Evolutionary Multi-Criterion Optimization*, E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, and D. Corne, Eds. 2001, pp. 284–298, Springer Verlag.

[44] C. K. Chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, pp. 462–467, 1968.

Peter A.N. Bosman is currently finishing his Ph.D. thesis in computer science on the topic of iterated density estimation evolutionary algorithms at Utrecht University in the Netherlands. His main research interests are in the field of computational intelligence and specifically include evolutionary computation and statistical learning.

Dirk Thierens received his Ph.D. in computer science on the topic of the analysis and design of genetic algorithms in 1995 at the University of Leuven in Belgium. From 1991 to 1994 he was a visiting researcher at the Illinois Genetic Algorithms Laboratory of the University of Illinois at Urbana–Champaign. He is currently an assistant professor at the Institute of Information and Computing Sciences at Utrecht University. His main research interests are in the field of computational intelligence, specifically evolutionary computation.