A Survey of Diversity Oriented Optimization: Problems, Indicators, and Algorithms

Vitor Basto-Fernandes, Iryna Yevseyeva, André Deutz and Michael Emmerich

Abstract In this chapter it is discussed, how the concept of diversity plays a crucial role in contemporary (multi-objective) optimization algorithms. It is shown that diversity maintenance can have a different purpose, such as improving global convergence reliability or finding alternative solutions to a (multi-objective) optimization problem. Moreover, different algorithms are reviewed that put special emphasis on diversity maintenance, such as multicriteria evolutionary optimization algorithms, multimodal optimization, artificial immune systems, and techniques from set oriented numerics. Diversity maintenance enters in different search operators and is used for different reasons in these algorithms. Among them we highlight evolutionary, swarm-based, artificial immune system-based, and indicator-based approaches to diversity optimization. In order to understand indicator-based approaches, we will review some of the most common diversity indices that can be used to quantitatively assess diversity. Based on the discussion, 'diversity oriented optimization' is suggested as a term encompassing optimization techniques that adress diversity maintainance as a major ingredient of the search paradigm. To bring order into all these different approaches, an ontology on diversity oriented optimization is proposed. It provides a systematic overview of the various concepts, methods, and applications and it can be extended in future work.

V. Basto-Fernandes (✉)
Polytechnic Institute of Leiria, Instituto Universitário De Lisboa (ISCTE-IUL), University Institute of Lisbon (ISTAR-IUL), 1649-026 Lisboa, Portugal
e-mail: Vitor.Basto.Fernandes@iscte.pt

I. Yevseyeva
Faculty of Technology, School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, UK
e-mail: iryna.yevseyeva@dmu.ac.uk

A. Deutz · M. Emmerich
Leiden University, 2333 CA Leiden, The Netherlands
e-mail: a.h.deutz@liacs.leidenuniv.nl

M. Emmerich
e-mail: m.t.m.emmerich@liacs.leidenuniv.nl

M. Emmerich et al. (eds.), *EVOLVE – A Bridge Between Probability, Set Oriented Numerics and Evolutionary Computation VII*,
Studies in Computational Intelligence 662, DOI 10.1007/978-3-319-49325-1_1

1 Introduction and Motivation

The concept of *diversity* plays a crucial role in various optimization and search techniques. Diversity maintenance can help to find a globally optimal solution, but it might also be the goal of optimization to produce a diversified set. Strategies to maintain diversity are used in various methods, in particular in population-based metaheuristics and their variation and selection operators. Moreover, there exists a multitude of diversity measures, addressing different aspects of what common sense might tell us what diversity is.

In this chapter, we look at the concept of diversity across several different methods and try to define 'diversity oriented optimization' as an emerging topic in optimization methods. Towards this end, we propose an ontology that seeks to provide a systematic overview, and can be used by the algorithm community to identify essential similarities and differences between different methods. This can be useful to find related work across algorithmic sub-disciplines or to identify prevalent trends in the field.

Before giving more concrete definitions of diversity, a tentative definition of diversity could be given as follows: Diversity is a property of a multi-set the elements of which are all members of the same space, say $\mathbb{M}$. The space can be, for instance, the set of integer numbers, the set of real vectors of dimensions n, or, the set of all molecular structures. It is demanded that $\mathbb{M}$ is at least equipped with a dissimilarity measure $d : \mathbb{M} \times \mathbb{M} \to \mathbb{R}_0^+$. Intuitively, we would then say that a subset of $\mathbb{M}$ is more diverse than another subset of $\mathbb{M}$, if

1. it contains more different elements,
2. elements are more different with respect to each other,
3. and more evenly distributed over $\mathbb{M}$.

In the literature a diverse set of diversity measures has been suggested, emphasizing these or subsets of these three aspects.

Traditionally, formal definitions of diversity have been mainly investigated in biological statistics in order to measure population diversity, but recently there is a growing interest in other fields of science and economics, too. Examples are cultural sciences, innovation management, and financial portfolio theory. Last but not least, the concept of diversity is a concept of vital importance in contemporary optimization algorithms. In the following we will focus on this last mentioned topic. Thereby we will often refer to terminology and diversity measures developed in other fields of science.

In many optimization techniques it is not even made explicit which diversity measure is used. Rather it is claimed, that a certain operator or strategy is used to increase diversity or to maintain diversity, not being explicit what exactly is meant by diversity. This is however not the case in the so-called indicator-based optimization methods, which aim for improving a diversity measure that is defined *a priori*. In this chapter we will therefore first review methods that use a rather vague definition of diversity, and then introduce indicator-based methods that refer to exact definitions of diversity and review these definitions.

This work is structured in three parts:

- The first part is focused on different optimization problems and how the concept of diversity is important to solve or define these problems.
- The second part reviews optimization methods that emphasize the concept of diversity for various reasons. The section on indicator-based methods also discusses various measures of diversity.
- Finally, in the third part, an ontology that integrates the different theoretical concepts, methods, and applications is developed. Based on this ontology, commonalities and essential differences between methods and problem definitions are discussed.

Figure 1 presents an overall perspective of all dimensions of diversity oriented optimization addressed in this study.

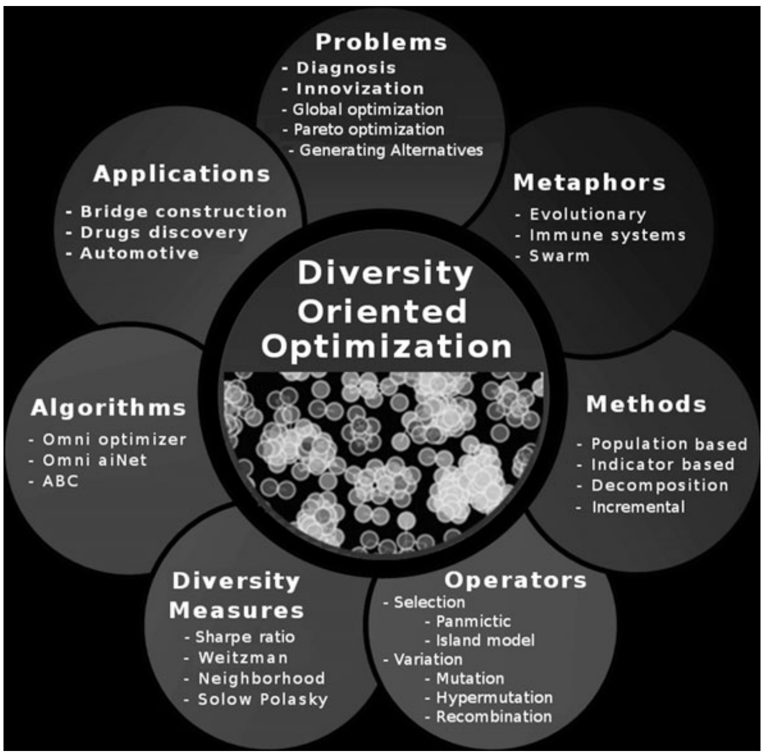

Fig. 1 Diversity oriented optimization. The lists serve as examples but are not exhaustive

2 Problem Domains in Diversity Oriented Optimization

This section reviews several problem domains, in which diversity is of importance to increase the performance of optimization methods or in which diversity is aimed for in the results of optimization methods.

2.1 Generating Alternatives

In engineering problems, when an algorithm provides the decision maker with a solution of a *single objective* optimization problem, it might not always correspond to the decision maker's preferences. In order to satisfy a demanding decision maker, or several decision makers, the developers of Evolutionary Algorithm to Generate Alternatives (EAGAs) [47] suggested an algorithm that searches for several good (not necessarily optimal) but maximally different solutions.

In complex problem fields and search spaces, such as drug discovery, only by looking at a solution it can be judged by a domain expert, whether or not that solution is possibly suitable. The famous chemist Linus Pauling summarized the process of discovery [25] as follows: "*the best way to get a good idea is to get a lot of ideas*". Indeed, modern computational tools for drug discovery can be described as diversity-oriented search for generating a set of promising alternatives [42]. A similar view is taken in the research of Ulrich et al. [38], on finding diversified sets of architectural bridge designs and hardware configurations.

2.2 Multiobjective Optimization

In multi-objective optimization it is a common practice to compute a diverse set of Pareto optimal solutions. As opposed to the so-called *a priori* approach to multicriteria decision making, where different objectives are aggregated to a scalar utility value, the so-called *a posteriori* approach first computes all Pareto optimal solutions and presents this set to the decision maker.

As pointed out by Knowles [21], the Pareto optimal set can be viewed as the set of optimal solutions over all meaningful linear or non-linear scalarizing utility functions. The knowledge of the Pareto front provides the decision maker with information on the trade-offs between different objectives and the availability of solutions satisfying potential goal vectors. Typically, diversity is measured for the set of non-dominated objective function vectors, but more recently the importance of decision space diversity has been stressed in several publications [8, 29, 31, 40]. Here the idea is, that it can be important to compute different pre-images of the same preferred point on the Pareto front, given they exist. For instance, a decision maker might prefer some solutions to others for subjective, non-explicit reasons.

2.3 Innovization

By analyzing diversified sets of good (not necessary optimal) solutions, provided by a diversity-oriented search algorithm, a designer may learn some important properties of the decision space. The findings might even be generalizable and lead to the discovery of some design patterns. It could be that some properties are shared among high-performing solutions. Early research in this direction was conducted by Ian Parmee et al. [24] in the field of evolutionary design optimization using so-called cluster oriented genetic algorithms (COGAs).

Recently, the derivation of design principles from optimization results has been discussed in the context of multi-objective optimization. The question is here how a design changes when navigating along the Pareto front and whether solutions on the Pareto front are distinguished from dominated solutions. Modern data mining and data analysis techniques are used to discover patterns and rules. Deb et al. call this process of innovation by optimization '*innovization*', see e.g. [10–12] and it is believed that this methodology has a large potential to revolutionize the engineering design process in industry.

2.4 Novelty and Interestingness

In this context it is also important to consider resistance of decision makers to make large changes to existing solutions. This resistance is explained by the risk due to the lack of knowledge about a novel design. In [27], the concepts such as *interestingness* and *novelty* are defined in the context of innovization. It is emphasized that candidate solutions for new designs must be different to known solutions but still be understood well enough on the basis of existing models used in the domain.

In the interestingness measure derived by Reehuis et al. [27], diversity is not only seen as an asset but it also comes with a risk, due to the unknown properties of novel designs. It is stressed that a certain balance between novelty and sticking to known and well-tested designs must be maintained, in order to make solutions appear interesting to designers. Moreover, interestingness can only be measured relative to existing designs.

2.5 Finding Peaks in Multimodal Landscapes

When exploring *multimodal* landscapes for the global optimum, it might be beneficial to explore several attractor basins, or peaks, in parallel. Many optimization strategies focus, after a transient initial phase, only on a single attractor. This also holds for population-based algorithms due to several causes of diversity-loss [28]. Diversity maintenance techniques can be crucial to improve the probability of finding the global optimum. Among others, Stoean et al. [36] argue about the need of active diversity maintenance to be considered in the context of global optimization.

In [12] the need of finding *all* global optima (multi-global optimization) is termed *multi-global optimization*, both in single- and multi-objective optimization. Moreover, algorithms that seek to gain knowledge about the topological structure of multimodal landscapes are recently developed [26]. The goal of these methods is not only to discover global optimizers but also to find local optimizers and their connectedness.

2.6 Model-Based Diagnostics

A system of cause and effect elements can be modeled as a function from input variables (causes) to output variables (effects). When performing model-based fault diagnosis, it is important to find all possible causes to an observed effect. If there are different possible causes, this indicates that further data is needed to identify the true cause. This aspect has been discussed in [48]. Here a diversity-oriented evolutionary algorithm was developed for finding potential contaminant sources in water distribution networks.

2.7 Dynamic and Robust Optimization

Both, when the goal is to find robust optima or when the positions of optima change over time, the maintenance of diversity can be a crucial component of the algorithm. For instance, in [19], it was pointed out that the lack of diversity can lead to stagnation of population-based search in suboptimal regions in dynamically changing environments. The idea that diversity makes a group of individuals more robust to attacks by, for instance, viruses is a common explanation for the evolutionary benefits of diversity maintenance. When transferred to optimization, such considerations might play a role when it comes to selection of portfolios or teams, which will be discussed later in Sect. 5.

2.8 Summary

In this section various motivations for maintaining diversity in optimization were reviewed. Across these different motivations, it is widely common that diversity is seen as an asset. It is either useful to maintain diversity in order to find a better result, or the result itself should include different solutions in order to be more interesting for the decision maker.

Negative aspects of diversity are that a too large output set might confuse the decision maker or that solutions that are too different from existing designs might not be trusted. In both cases it would be a rather good strategy to use diversity maintenance in moderation, rather than abandoning it at all.

3 Bio-Inspired Methods for Diversity Oriented Optimization

Next, it will be discussed how the problem of diversity maintenance is adressed in optimization algorithms.

Often, the mechanism has been inspired by nature and therefore the first part of the discussion of the methods will be devoted to nature inspired methods. In nature inspired methods terminology is often related to metaphors from biology. This, on the one hand makes it easy to see the analogy, on the other hand it can make it difficult to compare different bio-inspired methods with each other. Besides reviewing mechanisms, this section will also reveal commonalities between various strategies which might be slightly obscured by terminology.

3.1 Evolutionary Algorithms

In *Evolutionary Algorithms (EAs)*, a population (set) of individuals (solution candidates) evolves over time to a population with a better average objective function value among individuals. The process is driven by selection of promising parents, recombination and mutation.

Using a population of individuals by itself does not guarantee however that diversity is maintained. Even in situations where individuals share the same objective function value, that is optimization on a plateau function, diversity is quickly lost due to genetic drift [28]. Those individuals of types that are underrepresented in the populations tend to reproduce less frequently and the number of individuals of this type gets even smaller in subsequent generations. This effect can be quantified by Markov chain models (cf. [28]) and the time to extinction of an individuals tends to be proportional to the population size.

A common paradigm for maintaining diversity in EA populations is that of niching. For instance, *niching* [32] allows selecting only few solutions located within the same region of the objective or decision space, for parental or environmental selection. As an example of a global optimization strategy with diversity maintenance the omni-optimizer [13] will be discussed in more detail.

3.1.1 Omni-Optimizer

The omni-optimizer [13] was developed to allow both single and multi-objective optimization by means of a single generic evolutionary algorithm. It is based on the well-known generational NSGA-II algorithm for finding all Pareto optimal solutions for problems with multiple, usually conflicting objectives. However, omni-optimizer can also adapt automatically when simpler cases of multi-objective problems with a single optimum are detected, or when a single objective optimization problem should be solved either for finding multiple optima or a single optimum.

When compared to the original NSGA-II, omni-optimizer is improved with restricted mating selection of similar individuals competing in tournament. For tournament, two pairs of individuals are selected such that for each randomly selected individual its nearest neighbour in objective space becomes its competitor. Such restricted parent selection in tournaments limits competition to very similar individuals only, it also preserves possible multiple optima and speeds up convergence to them or to a single global optimum. A modified environmental selection is computed not only taking into account the phenotype (objective space evaluation as in NSGA-II), but also the genotype (decision space) of an individual. Moreover, pairwise comparison of individuals is based on a modified ε-domination evaluation, which neglects a small differences on objective function values between two individuals when deciding which one is the best. In addition, a more disruptive mutation operator is obtained by modifying the treatment of the original polynomial mutation at the boundaries of variables. The algorithm's initial population is based on Latin-hypercube random uniform sampling, but predefined sampling can also be used.

For the multi-objective case the winner of tournament is selected taking into account feasibility, constraints violation, dominance and crowding of each individual evaluated in objective space only. When selecting among two individuals the following criteria apply: (1) A feasible individual is preferred to an infeasible one; (2) A feasible non-dominated individual is preferred to a feasible dominated one; (3) Among two feasible non-dominated individuals the one with higher crowding distance is preferred (or randomly selected if crowding distance is the same); And (4) among two infeasible individuals the one with smallest constraint violation is preferred. When compared to NSGA-II, omni-optimizer evaluates dominance relation using ε-dominance. The advantage of this type of domination is its tolerance towards small differences of near non-dominated individuals. Keeping such individuals may be beneficial when diversity is an issue and when several rather than one solutions should be selected. A penalty for constraints violation is computed as a sum of violations of all equality and inequality constraints. In case of single objective optimization, the tournament selection degenerates to the above mentioned criteria (1) and (4), and feasible individuals with smallest objective function values are preferred.

For individuals represented by real-coded variables, SBX crossover (on half of variables, on average) and a modified polynomial mutation operators are used. When compared to the original polynomial mutation operator, which has the disadvantage of having no effect as soon as a variable reaches its boundary, the new polynomial mutation operator assigns non-zero probability of mutation even if one of the variables of an individual is on its boundary. For individuals represented by binary coded variables two-point crossover is applied together with bitwise mutation.

Although omni-optimizer follows a $(\mu + \lambda)$ schema, similar to NSGA-II, the environmental selection stage is modified. Omni-optimizer considers similar principles used for parents selection, such as feasibility, constraints violation, domination and crowding of individuals in objective space, but applies ε-domination similar to that in mating selection. Moreover, crowding distance is computed as an average between its two closest neighbours for both objective and decision spaces on all decision variables and objectives values, respectively. The biggest of the two values, either its normalized crowding distance in objective space or its normalized crowding distance in decision space, is taken as a crowding distance of an individual. By considering decision space, this crowding distance allows differentiating between two non-dominated individuals with the same/similar evaluation in objective space, but different structurally (in decision space).

The results of omni-optimizer testing on a number of single uni-optimal and multi-optimal test problems and multi-objective uni-optimal and multi-optimal test functions are reported in [13]. They are promising both in terms of quality and coverage of the Pareto front in multi-objective optimization problems.

However, omni-optimizer reveals poor results of crowding distance in three and more dimensions and a possible loss of optimal solutions when using ε-domination. Moreover, the evident advantage of omni-optimizer developed as a generalized solver to deal with a variety of problems does not exclude the fact that specific problem-oriented algorithms may be more efficient for specific problems. For instance, omni-optimizer was not designed to find local optima of multi-modal problems but only to find global one(s).

3.2 Artificial Immune Systems

Immune Systems have inspired various algorithms, among which are also algorithms in diversity oriented search. These so called Artificial Immune Systems (AIS) have been proposed by de Castro and Von Zuben [5]. In AISs, adaptation happens by cumulative variation and selection within cells (cloning and hypermutation): In the biological counterpart, immunoglobulin nucleotides are randomly inserted and deleted from recombined immunoglobulin gene segments. In AISs mutation operators (hypermutation and receptor editing) are applied to vectors that correspond to lymphocyte clones and might involve exchange or shifts of positions of individual representation data elements. These AIS operators are of major importance for cell diversification and affinity enhancement to antigens.

In AISs, diversity is preserved intrinsically by *clonal selection theory* [4] and *immune network theory* [18] principles. In AISs, a population evolves by cloning and mutation (hypermutation) processes, with genetic variability inversely proportional to affinity and concentration among individuals in the population. Self-adapting metrics, such as *affinity* among cells guarantee that concentration of similar cells decrease in the presence of better neighboring cells: the closer and higher the concentration of better cells, the higher their influence. Cells without better solutions in their neighborhood increase their concentration proportionally to their fitness.

3.2.1 Artificial Immune Systems *versus* Evolutionary approaches

Although coming from two different theoretical biology domains, AISs and evolutionary systems inspired by evolution theory, follow some similar fundamental principles: diversity, natural variation and selection are both present.

In both EAs and AISs, the population evolves by variation of selected cells/individuals. Genetic crossover and mutation are responsible for population diversity and fine-tuning. These variation mechanisms are present in both AISs and evolutionary approaches. In particular, in EAs, biological evolution happens by cumulative natural selection among individuals: crossover and mutation generates offspring from mixing parental genes.

When compared to other nature-inspired algorithms, *Artificial Immune Systems* (AIS)s automatically adjust the population size at each iteration depending on the problem needs and still preserves diversity of the shrinked or enlarged population. Moreover, they apply hypermutation, which can be seen as a concept that combines aspects of mutation and recombination.

Both approaches are successfully applied for diversity-oriented search, e.g. in multi-objective optimization, see e.g., [8, 31].

3.2.2 Omni-aiNet

Omni-aiNet [7] was developed to serve the same purpose as omni-optimizer: for solving single and multi-objective optimization. However, contrary to omni-optimizer, omni-aiNet is based not on evolutionary algorithms, but on artificial immune systems.

Omni-aiNet is based on opt-aiNet [9] and several AISs principles, such as *clonal selection principle* and *immune network theory*, and adapts some ways of solving a common optimization dilemma of driving search towards both exploration and exploitation of the objective space. The algorithm starts by initializing a randomly user-predefined number of individuals with a set of real-coded variables. The initial population enters the generational loop, at each iteration of which the current population goes through cloning, hypermutation, selection and gene duplication stages, until some stopping criterion is met. At the cloning stage for each individual of the current population user-predefined number of clones are created and mutated by polynomial mutation. The probability of mutation is selected inversely proportional to affinity of a clone with its antigen. This mutation parameter is defined empirically.

From the current population and the population of mutated clones the new population is selected similarly to the parental selection of omni-optimizer (based on ε-domination and constraints violation principles), except that instead of the crowding distance principle for selecting among two mutually non-dominating solutions, grid-based selection is used, similarly to the one introduced in the Pareto Archived Evolution Strategy (PAES) [22]. For each objective the interval between its minimal and maximal values is divided into a grid with user defined resolution. After partitioning the objective space into a grid of cells according to all objectives, only a user-predefined number of solutions closest to the centre of each cell is selected (here this number is equal to the number of grids).

3.3 Swarm Intelligence

Swarm Intelligence (SI) is another biological paradigm similar to AISs dynamic approach to diversity preservation: swarm self-adapts itself to the environment and/or communication between swarm members-agents when necessary [2]. For instance, in *Artificial Bee Colony* (ABC) algorithms, a special type of swarm bee-agents, called *scouts*, are activated on the last exploration stage of the algorithm to promote following diversification. After two exploitation-intensification phases, where the *employed* and *onlooker* bee-agents search for local optima, based on deterministic and probabilistic selection, respectively, *scout* bee-agents force abandoning of non-promising solutions and start exploring new solutions corresponding to new decision space regions. A similar multi-agent approach for multi-modal search, motivated by exploration strategies of *scouts*, is the *self-organizing scouts* (SOS) algorithm [3].

Biological paradigms are also addressed in spatial population structures as opposed to panmictic ones. Examples are cellular genetic algorithms [1] and spatial predator-prey algorithms in multi-objective optimization, which were investigated first in [23].

Besides biological paradigms, the mathematical programming community has developed several algorithms for diversity-oriented optimization that exploit mathematical structures of functions expressing diversity [46].

In Table 1 we present a summarized comparison of the bio-inspired metaphors that were described in previous sections.

Table 1 Bio-Inspired Computational Metaphors

	Metaphors		
	EA	AIS	SI
Rationale	*Natural selection*	*Clonal selection*	*Social cooperation*
Reproduction	*Recombination*	*Cloning*	*Specialization*
Local search	*Mutation*	*Hypermutation*	*Onlooker individuals*
Variation	*Recombination*	*Cumulative variation*	*Scout individuals*
Adaptation	*Natural selection*	*Immunology principles*	*Multi-agent*
Diversity	*Niching, Various*	*Variable pop. size*	*Locality*

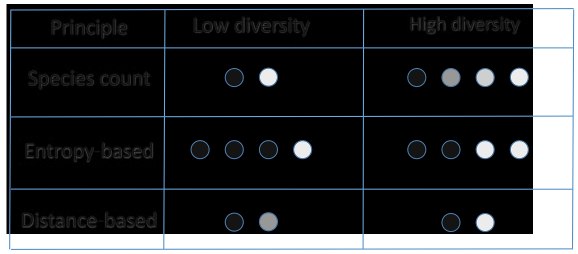

Fig. 2 Principles used in the definition of diversity. Individuals with different *gray* level are considered to be of different species. The more different the *gray* level, the more distant are the species to each other

4 Diversity Measures

When assessing the performance of diversity maintenance mechanisms, it is of crucial importance to apply reliable and well-understood diversity measures. Many of these measures stem from biostatistics, because in conservational biology it is a common problem to measure the diversity of species in a population. In diversity oriented optimization, with *species* we express the concept of a class of points that is essentially different from all other points in a population while being similar to each other. An overview of the underlying principles of diversity measures is provided in Fig. 2, where points with different colors symbolize different species. Diversity indicators might take into account simply the number of (essentially different) species in a population, or also measure the eveness of the distribution of species (entropy-based), or the dissimilarity of species with respect to each other. Next, we will discuss important diversity measures and their relation to each other in more detail.

4.1 Diversity Based on the Abundance of Species

The simplest diversity measure is *species richness*, which is just the total number of species in a population. This index has the drawback that it does not measure the relative abundance of a species. In other words, if a population with n species would almost entirely consist of individuals of the same species, it would have the same richness as a population with evenly distributed abundance of species. A couple of diversity indices presented next seek to circumvent this problem:

A classical diversity measure that takes into account the relative abundance of a species is the *Simpson index* [33]. For a given population P, the Simpson index $S(P)$ measures the probability that if we draw a random sample of an individual of a certain species without replacement, in a second experiment we draw an individual of a different species. If the distribution of species abundances is more even, this index has a higher value. Let n_i denote the number of individuals of species i, N denote the total number of individuals, and n the number of species. Then

$$S(P) = 1 - \frac{\sum_{i=1}^{n} n_i(n_i - 1)}{N(N-1)}.$$

A disadvantage of the Simpson index is, that it is difficult to interpret for large populations, because the probabilities tend to get very close to 0. It is difficult for humans to judge the difference between a probability of, e.g., 0.0099 and of, e.g., 0.00999.

A similar measure is the Shannon entropy:

$$S(P) = -\sum_{i=1}^{n} \frac{n_i}{N} \cdot \log \frac{n_i}{N}.$$

The Shannon entropy reaches its maximum, if the species are equally abundant and grows with the number of species. Still the growth is limited by the slow growth of the logarithm.

The *Rich Gini Simpson quadratic index* (RGS) [17] obtains values from 0 to $n - 1$. The maximum is obtained for an evenly distributed population. It, thus, gives a good idea about species diversity, and can be compared for populations of different sizes. It is computed as:

$$RGS(P) = n \sum_{i=1}^{n} \frac{n_i}{N}(1 - n_i/N) = n\left(1 - \sum_{i=1}^{n} \left(\frac{n_i}{N}\right)^2\right).$$

The evenness of a distribution of species abundances could also be measured independently of the population size by the *Gini index*, which originated from measuring welfare of an economy [6]. It is given by the average absolute deviation from the mean and tends to zero as the population tends to be evenly distributed over the species.

$$G(P) = \frac{1}{n-1}\left(n + 1 - 2\left(\frac{\sum_{i=1}^{n}(n+1-i)n_i}{\sum_{i=1}^{n} n_i}\right)\right).$$

4.2 Diversity Based on Distances

In the aforementioned methods, the distance between species is not considered. For instance Weitzman [43] suggested to consider sets with the same number of species but bigger dissimilarity between species to be more diverse. In addition, he demanded that a diversity measure should grow if the number of species increases. Based on these and a number of additional properties, he suggested a diversity measure that we will term *Weitzman Diversity* $W(P)$. It is defined as follows: Given a population P and a distance matrix d_{ij} for members i and j, $i = 1, \ldots, N$, $j = 1, \ldots, N$: Let $d(i, Q)$ be the distance between i and the closest element (aka neighbor) in Q, for some non-empty $Q \subset P$. Moreover, let $P \setminus i$ define the population with the i-th individual removed. Then, the Weitzman diversity is recursively defined via

$$W(P) := \max_{i \in \{1,\ldots,N\}} \left(W(P \setminus i) + d(i, P \setminus i)\right).$$

The Weitzman diversity has interesting theoretical properties, but it is costly to compute it in practice for large populations. The running time of the fastest known algorithm scales with $O(2^N)$.

A simplification of the Weitzman distance would be to compute only the first iteration of the recursion, which is known as the MAX-MIN diversity [15]:

$$M(P) := \max_{i \in \{1,\ldots,N\}} (d(i, P \setminus i)).$$

The MAX-MIN diversity is used as a straightforward way to measure diversity of subsets in operations research. It was shown that the problem of finding a k-size subset from P of maximal MAX-MIN diversity is NP hard. The MAX-MIN diversity can be used to compare populations that have the same size. Its maximum often yields evenly distributed populations, because every point seeks to maximize the distance to its nearest neighbor. However, in relative comparisons it might be misleading. Consider for instance a population $P_1 = \{1, 2, 8, 9\}$ and a population $P_2 = \{1, 2, 3, 4\}$. Then $M(P_1) = M(P_2)$, if we consider the absolute deviation as a distance measure. However, clearly P_1 is more widespread than P_2.

A proposal for a distance measure that shares most properties of the Weitzman diversity, but can be computed faster, is the *Solow Polasky diversity* [35]. It requires, however, a parameter $\theta > 0$ that needs to be chosen by the user. The definition of the Solow Polasky diversity $SP(P)$ is as follows: Let $c_{ij} = \exp(-\theta d_{ij})$ denote a correlation between point i and point j. If two points are of the same species the correlation is one. Let $\mathbf{M} = \mathbf{C}^{-1}$, assuming that $\mathbf{C}$ is of full rank. Then

$$SP(P) = \sum_{i=1}^{N} \sum_{j=1}^{N} m_{ij}.$$

It is easy to show that $SP(P)$ tends to N, if the distance between all species tends to be very large. Moreover, $SP(P)$ tends to one, if species are very similar with respect to each other. The parameter θ determines how fast the population tends to N when the distances increase.

In the literature also other distance-based diversity measures occur. For instance, the variance $V(P)$ of a population is given by $\mathbf{1}^T \Sigma \mathbf{1}$, where Σ is a problem specific covariance matrix. The entries of the covariance matrix can also depend on the dissimilarity. The variance of a population is often used in portfolio theory in order to measure the variance of the return (or risk) of a portfolio [44].

Another way to measure diversity is to compute arithmetic or geometric averages of distances to nearest neighbors. This measure can be computed efficiently and maximizing it will also lead to evenly spaced populations without duplicates. These so-called gap measures were discussed in [14] and used in various contexts in optimization. They, however, can only be used to compare populations of the same size.

It might be somewhat tempting to use the sum of pairwise distances as a diversity measure. However, when maximizing such measures, often clustering at the boundary is obtained. For instance, the population $\{1, 1, 4, 4\}$ has a higher sum of pairwise distances than $\{1, 2, 3, 4\}$. The first population has a sum of pairwise distances of 24 while the second population has a sum of pairwise distances of only 20.

In the next section we will discuss so-called indicator-based algorithms that are directly oriented towards maximizing these diversity measures.

5 Indicator-Based Optimization Methods

A recent trend is to stress quality performance measures in the design of an algorithm. In the context of metaheuristics, methods that directly seek to maximize some quality indicator of a set are called indicator-based optimization algorithms [49]. Originating from methods that seek to find approximations to Pareto fronts, indicator-based optimization algorithms are recently also used in other domains.

Quality performance measures targeted to optimization algorithms need to cope with the fact that, the outcome is not constituted by a single solution, but by a set of solutions (e.g., in multimodal and multi-objective optimization problems).

Each of the indicators allows comparing algorithms with respect to one of several properties, among which are the quality of sets of individuals, diversity and distance to the optimal set (assuming it is known). At the same time, researchers noticed that optimizing indicators themselves is a good strategy for population evolution in the framework of an algorithm, and suggested several indicator-based algorithms, a trend that started in evolutionary multi-objective algorithm research [49]. Due to the importance of diversity for set-based optimization, recently developed indicator-based algorithms tend to include diversity, either as a separate indicator, see e.g. [14], or as an integral part of an indicator, see e.g. [40].

Although there is no consensus on how to best capture and use the concept of diversity in optimization, some robust definitions of diversity and measures are pointed out in [20, 43]. A diversity index measuring the number of different potential solutions and also the spread of solutions is recommended.

Different diversity indicators were proposed in the context of diversity-oriented search. Ulrich et al. [41] suggested to chose indicators from bio-diversity. In this field the Weitzmann diversity [43] and the Solow Polasky indicator [34] are common diversity measures. While the Weitzmann indicator [43] is motivated by phylogenetic trees with maximum parsimony and has exponential time complexity, the Solow Polasky indicator is motivated by a utilitarian model of species conservation and its computation can be accomplished in polynomial time. Due to its higher efficiency the latter indicator is favored by Ulrich et al. [41]. Even faster are indicators based on simple statistics on gaps between nearest neighbors [14, 26], although they can only provide comparisons among populations of the same size.

Ulrich et al. [39] emphasized the importance of decision space diversity by suggesting diversity-optimizing single objective (NOAH) and multi-objective (DIOP) algorithms, see [39, 41], respectively, as well as an algorithm that integrates diversity within the hypervolume indicator, see DIVA algorithm in [40].

When searching for *level set* approximations (e.g. for approximating an implicitly defined manifold), set-proximity indicators are required, which often strongly correlate with diversity indicators. Several diversity-based indicators were compared in [14], and the Hausdorff distance-based indicator was suggested for level set approximation within the Evolutionary Level Set Approximation (ELSA) algorithm. Indicators for multimodal optimization are discussed in [26].

An alternative approach to balance convergence and diversity in evolutionary algorithms was proposed in [44]. It stems from portfolio-optimization. It is well-known in strategic decision making and financial management, and it is based on the idea of composing portfolios of assets, which have a high potential of high return in future and are of lowest possible risk of failing (for the same reasons). The latter is related to how assets differ from each other. Diversity of assets selected in the same portfolio is shown to reduce potential risk for the portfolio of assets as a whole. By analogy the population of individuals in evolutionary algorithms can be selected as a portfolio of diverse, highly performing individuals. It can be formulated as a bi-objective optimization problem with two objectives: maximizing potential return of a portfolio as a whole and minimizing risk (or, in other words, maximizing diversity). These two objectives can also be combined into a single indicator, e.g. Sharpe ratio. This indicator shows how return compensates the risk taken. Traditionally, similarity of assets can be measured by covariance of their individual returns, which could be evaluated in a probabilistic sense taking into account their current performance. Then individuals with low (preferably negative) covariance are selected in the portfolio. Such selection can be done at both the parental and the environmental selection phases of evolutionary algorithms, and the latter can even be combined with preserving solutions in the archive as in the Portfolio Optimization Selection Evolutionary Algorithm (POSEA) [44]. Testing the approach on some benchmarks shows its efficiency when compared to the state-of-the-art indicator-based evolutionary algorithms and its potential for many objective optimization application due to the fact that its complexity is independent of the number of objectives.

6 Taxonomy and Ontology

In artificial intelligence, ontologies are used to formally represent a knowledge domain [16]. All instances, attributes and classes in the universum of discourse are represented as well as relations between these concepts and instances. As opposed to simple taxonomies or hierarchical descriptions of a field, ontologies allow to model in parallel different concepts and relate them with rules. An advantage of representing the survey of the domain of diversity-oriented optimization by means of an ontology is that besides the formal representation of classes and relations between them, also predicate logic rules can be specified that will help the user to classify algorithms correctly and find related work. Moreover, graphical representations of the ontology allow for a quick assessment of the research activities within this field and how they are related with each other.

Preliminary proposals for this taxonomy and ontology are provided in Figs. 3 and 4, respectively, and have been developed with the Protégé ontology editor [37].

Most relevant classes, concepts and relations presented graphically in the ontology were contextualized and described in detail in the previous sections.

Fig. 3 Diversity oriented optimization taxonomy

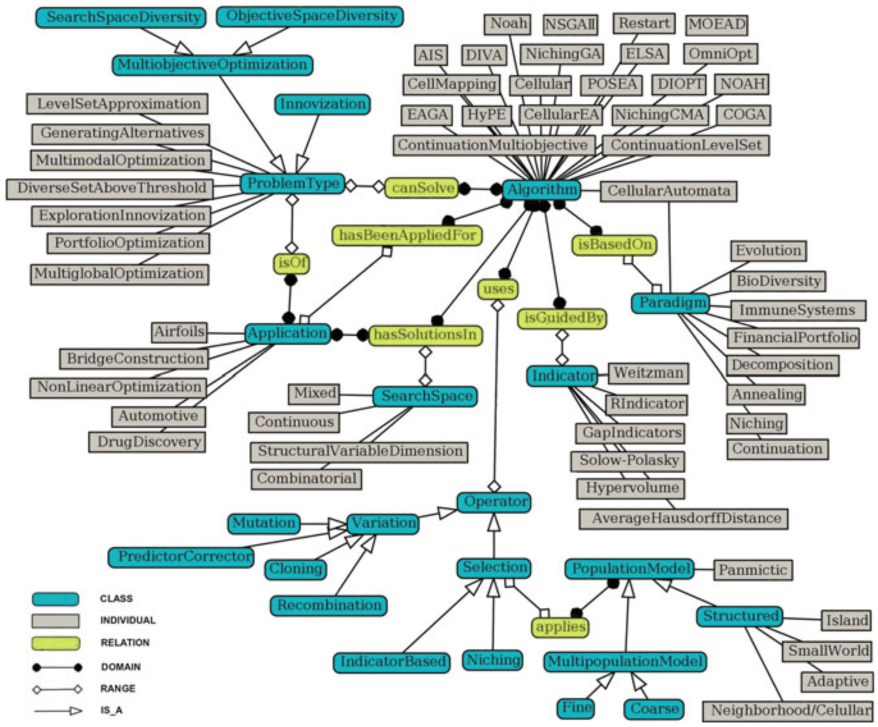

Fig. 4 Diversity oriented optimization ontology

7 Conclusion and Future Research

The overview proposed in this work aimed at capturing the development directions of diversity-oriented optimization algorithms. To represent the domain of diversity-oriented search in a systematic way, a diversity-oriented optimization taxonomy was established following an ontology discussion. In the nearest future, we aim at extending this ontology with various concepts, definitions and relations of diversity indicators, algorithms, and integrate diversity-related theoretical results into the survey.

In some application domains, the need for generating diverse solution sets has been particularly stressed and diversity-oriented optimization was successfully applied, for instance, in discrete design optimization in the car industry [38], truss bridge design and optimization [41], drug discovery [42, 45], quantum control [30], and space mission design [29]. It is expected that this is just the beginning and diversity oriented optimization will become increasingly important to design high performce and user friendly designs and search tools.

References

1. Alba, E., Dorronsoro, B.: Cellular Genetic Algorithms, Operations Research/Computer Science Interfaces, vol. 42. Springer, Heidelberg (2008)
2. Beekman, M., Sword, G.A., Simpson, S.J.: Biological foundations of swarm intelligence. In: Blum, C., Merkle, D. (eds.) Swarm Intelligence, Natural Computing Series, pp. 3–41. Springer, Heidelberg (2008)
3. Branke, J., Kaußler, T., Smidt, C., Schmeck, H.: A multi-population approach to dynamic optimization problems. In: Parmee, I. (ed.) Proceedings of the 4th International Conference on Adaptive Computing in Design and Manufacture (ACDM'2000, Plymouth, UK, April 26–28, 2000), pp. 299–307. Springer, Heidelberg (2000)
4. Burnet, F.: Clonal selection and after. In: Bell, G.I., Perelson, A.S., Pimbley Jr., G.H. (eds.) Theoretical Immunology, pp. 63–85. Marcel Dekker Inc., New York (1978)
5. de Castro, L., Von Zuben, F.: Learning and optimization using the clonal selection principle. Evol. Comput. IEEE Trans. **6**(3), 239–251 (2002)
6. Ceriani, L., Verme, P.: The origins of the Gini index: Extracts from Variabilità e Mutabilità (1912) by Corrado Gini. J. Econ. Inequal. **10**(3), 421–443 (2012)
7. Coelho, G.P., von Zuben, F.J.: Omni-aiNet: An immune-inspired approach for omni optimization. In: Bersini, H., Carneiro, J. (eds.) Proceeding of the 5th International Conference on Artificial Immune Systems (ICARIS, Oeiras, Portugal, September 4–6, 2006). Lecture Notes in Computer Science, vol. 4163, pp. 294–308. Springer, Heidelberg (2006)
8. Coelho, G.P., Von Zuben, F.J.: A concentration-based artificial immune network for multi-objective optimization. In: Takahashi, R.H.C., et al. (eds.) Proceedings of the 6th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2011, Ouro Preto, Brazil, April 5–8, 2011). Lecture Notes in Computer Science, vol. 6576, pp. 343–357. Springer, Heidelberg (2011)
9. de Castro, L., Timmis, J.: An artificial immune network for multimodal function optimization. In: Proceedings of the 2002 Congress on Evolutionary Computation, (CEC 2002, Honolulu, Hawaii, USA, May 12–17, 2002), vol. 1, pp. 699–704. IEEE (2002)
10. Deb, K.: Innovization: Discovery of innovative solution principles using multi-objective optimization. In: Purshouse, R., et al. (eds.) Proceedings of the 7th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2013, Sheffield, UK, March 19–22, 2013), pp. 4–5. Springer, Heidelberg (2013)
11. Deb, K., Srinivasan, A.: Innovization: Innovating design principles through optimization. In: Cattolico, M. (ed.) Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation (GECCO'06, Seattle, WA, USA, July 08–12, 2006), pp. 1629–1636. ACM, New York (2006)
12. Deb, K., Srinivasan, A.: Innovization: Discovery of innovative design principles through multiobjective evolutionary optimization. In: Knowles, J., et al. (eds.) Multiobjective Problem Solving from Nature, Natural Computing Series, pp. 243–262. Springer, Heidelberg (2008)
13. Deb, K., Tiwari, S.: Omni-optimizer: A generic evolutionary algorithm for single and multi-objective optimization. Eur. J. Oper. Res. **185**(3), 1062–1087 (2008)
14. Emmerich, M.T., Deutz, A.H., Kruisselbrink, J.: On quality indicators for black-box level set approximation. In: Tantar, E., et al. (eds.) EVOLVE - A bridge between Probability, Set Oriented Numerics and Evolutionary Computation, Studies in Computational Intelligence, vol. 447, pp. 157–185. Springer, Heidelberg (2012)
15. Ghosh, J.B.: Computational aspects of the maximum diversity problem. Oper. Res. Lett. **19**(4), 175–181 (1996)
16. Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. **5**(2), 199–220 (1993)
17. Guiasu, R.C., Guiasu, S.: The Rich-Gini-Simpson quadratic index of biodiversity. Nat. Sci. **2**(10), 1130–1137 (2010)
18. Jerne, N.K.: Towards a network theory of the immune system. Ann. Immunol. **125C**, 373–389 (1974)

19. Jin, Y., Branke, J.: Evolutionary optimization in uncertain environments-a survey. IEEE Trans. Evol. Comput. **9**(3), 303–317 (2005)
20. Jost, L.: Entropy and diversity. OIKOS **113**(2), 363–375 (2006)
21. Knowles, J.: Closed-loop evolutionary multiobjective optimization. IEEE Comput. Intell. Mag. **4**(3), 77–91 (2009)
22. Knowles, J., Corne, D.: The Pareto archived evolution strategy: a new baseline algorithm for Pareto multiobjective optimisation. In: P.J. Angeline, et al. (eds.) Proceedings of the 1999 Congress on Evolutionary Computation (CEC 99, Washington, USA, July 6–9, 1999), vol. 1, pp. 98–105. IEEE, New Jersey (1999)
23. Laumanns, M., Rudolph, G., Schwefel, H.P.: A spatial predator-prey approach to multi-objective optimization: A preliminary study. In: Eiben, A.E., et al. (eds.) Proceedings of the 5th International Conference on Parallel Problem Solving from Nature (PPSN V, Amsterdam, The Netherlands, September 27–30, 1998). Lecture Notes in Computer Science, vol. 1498, pp. 241–249. Springer, Heidelberg (1998)
24. Parmee, I.C., Bonham, C.R.: Towards the support of innovative conceptual design through interactive designer/evolutionary computing strategies. AI EDAM **14**(1), 3–16 (2000)
25. Pauling, L.: The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry, vol. 18, 3d edn. Cornell University Press, Ithaca (1960)
26. Preuß, M., Wessing, S.: Measuring multimodal optimization solution sets with a view to multiobjective techniques. In: Emmerich, M.T., et al. (eds.) Proceedings of the 4th International Conference: EVOLVE-A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation (EVOLVE 2013, Leiden, The Netherlands, July 10–13, 2013), Advances in Intelligent Systems and Computing, vol. 227, pp. 123–137. Springer, Heidelberg (2013)
27. Reehuis, E., Kruisselbrink, J., Olhofer, M., Graening, L., Sendhoff, B., Bäck, T.: Model-guided evolution strategies for dynamically balancing exploration and exploitation. In: Hao, J., et al. (eds.) Proceedings of the 10th International Conference on Artificial Evolution, (EA 2011, Angers, France, October 24–26, 2011), pp. 306–317. Springer, Heidelberg (2011)
28. Schönemann, L., Emmerich, M.T., Preuß, M.: On the extinction of evolutionary algorithm subpopulations on multimodal landscapes. Informatica (Slowenien) **28**(4), 345–351 (2004)
29. Schütze, O., Vasile, M.: Coello Coello, C.A.: Approximate solutions in space mission design. In: Proceedings of the 10th International Conference on Parallel Problem Solving from Nature (PPSN X. Dortmund, Germany, September 13–17, 2008). Lecture Notes in Computer Science, vol. 5199, pp. 805–814. Springer, Berlin (2008)
30. Shir, O., Beltrani, V., Bäck, T., Rabitz, H., Vrakking, M.: On the diversity of multiple optimal controls for quantum systems. J. Phys. B At. Mol. Opt. Phys. **41**(7), (2008)
31. Shir, O., Preuß, M., Naujoks, B., Emmerich, M.: Enhancing decision space diversity in evolutionary multiobjective algorithms. Evolutionary Multi-Criterion Optimization. Studies in Computational Intelligence, pp. 95–109. Springer, Heidelberg (2009)
32. Shir, O.M.: Niching in evolutionary algorithms. In: Rozenberg, G., Bäck, T., Kok, J.N. (eds.) Handbook of Natural Computing: Theory, Experiments, and Applications, pp. 1035–1069. Springer, Heidelberg (2012)
33. Simpson, E.H.: Measurement of diversity. Nature **163**(4148), 688 (1949)
34. Solow, A., Polasky, S., Broadus, J.: On the measurement of biological diversity. J. Environ. Econ. Manag. **24**(1), 60–68 (1993)
35. Solow, A.R., Polasky, S.: Measuring biological diversity. Environ. Ecol. Stat. **1**(2), 95–107 (1994)
36. Stoean, C., Preuß, M., Stoean, R., Dumitrescu, D.: Multimodal optimization by means of a topological species conservation algorithm. IEEE Trans. Evol. Comput. **14**(6), 842–864 (2010)
37. Tudorache, T., Nyulas, C., Noy, N.F., Musen, M.A.: WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web. Semant. web **4**(1), 89–99 (2013)
38. Ulrich, T.: Exploring structural diversity in evolutionary algorithms. Ph.D. thesis, ETH Zurich, TIK Institut für Technische Informatik und Kommunikationsnetze (2012)

39. Ulrich, T., Bader, J., Thiele, L.: Defining and optimizing indicator-based diversity measures in multiobjective search. In: Schaefer, R., et al. (eds.) Proceedings of the 11th International Conference on Parallel Problem Solving from Nature: Part I (PPSN XI, Krakow, Poland, September 11–15, 2010), pp. 707–717. Springer, Heidelberg (2010)
40. Ulrich, T., Bader, J., Zitzler, E.: Integrating decision space diversity into hypervolume-based multiobjective search. In: Pelikan, M., Branke, J. (eds.) Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation (GECCO'10, Portland, USA, July 07–11, 2010), pp. 455–462. ACM, New York (2010)
41. Ulrich, T., Thiele, L.: Maximizing population diversity in single-objective optimization. In: Krasnogor, N., Lanzi, P.L. (eds.) Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation (GECCO '11, Dublin, Ireland, July 12–16, 2011), pp. 641–648. ACM, New York (2011)
42. van der Horst, E., Marqués-Gallego, P., Mulder-Krieger, T., van Veldhoven, J., Kruisselbrink, J., Aleman, A., Emmerich, M.T., Brussee, J., Bender, A.: IJzerman, A.P.: Multi-objective evolutionary design of adenosine receptor ligands. J. Chem. Inf. Model. **52**(7), 1713–1721 (2012)
43. Weitzman, M.L.: On diversity. Q. J. Econ. **107**(2), 363–405 (1992)
44. Yevseyeva, I., Guerreiro, A.P., Emmerich, M.T., Fonseca, C.M.: A portfolio optimization approach to selection in multiobjective evolutionary algorithms. In: Bartz-Beielstein, T., et al. (eds.) Proceedings of the 13th International Conference on Parallel Problem Solving from Nature (PPSN XIII, Ljubljana, Slovenia, September 13–17, 2014). Lecture Notes in Computer Science, vol. 8672, pp. 672–681. Springer, Heidelberg (2014)
45. Yevseyeva, I., Lenselink, E.B., de Vries, A., Ijzerman, A.P., Deutz, A.H., Emmerich, M.T.: Multiobjective portfolio optimization for drug discovery using deterministic and stochastic methods. In: M.J. Geiger (ed.) Abstracts of the 23d International Conference on Multicriteria Decision Making (MCDM 2015 - Bridging Disciplines, Hamburg, Germany, August 2–7 (2015)
46. Zadorojniy, A., Masin, M., Greenberg, L., Shir, O.M., Zeidner, L.: Algorithms for finding maximum diversity of design variables in multi-objective optimization. Procedia Comput. Sci. **8**, 171–176 (2012)
47. Zechman, E., Ranjithan, S.: An evolutionary algorithm to generate alternatives (EAGA) for engineering optimization problems. Eng. Optim. **36**(5), 539–553 (2004)
48. Zechman, E., Ranjithan, S.: Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system. J. Water Res. Planning Manag. **135**(5), 334–343 (2009)
49. Zitzler, E., Künzli, S.: Indicator-based selection in multiobjective search. In: Yao, X. (ed.) Proceedings of the 8th International Conference on Parallel Problem Solving from Nature (PPSN VIII, Birmingham, UK, September 18–22, 2004), pp. 832–842. Springer-Verlag, Berlin, Heidelberg (2004)