Emmerich et al. (Eds.): EVOLVE 2013, July 10-13, 2013
A Bridge between Probability, Set Oriented Optimization, and Evolutionary Computation
©LIACS, Leiden University, nl: ISBN 978-2-87971-118-8, ISSN 2222-9434

A Survey of Diversity-Oriented Optimization

Vitor Basto-Fernandes[a] and **Iryna Yevseyeva**[b] and **Michael Emmerich**[c]

[a]Informatics Engineering Department
Technology and Management School
Computer Science and Communications Research Centre
Polytechnic Institute of Leiria, 2411-901 Leiria, Portugal
vitor.fernandes@ipleiria.pt

[b]School of Computing Science
Centre for Cybercrime and Computer Security
Newcastle University, NE1 7RU Newcastle University upon Tyne, United Kingdom
iryna.yevseyeva@ncl.ac.uk

[c]Leiden Institute of Advanced Computer Science
Faculty of Science
Leiden University, 2333-CA Leiden, The Netherlands
emmerich@liacs.nl

1 Introduction

The concept of *diversity* plays a crucial role in many optimization approaches: On the one hand, diversity can be formulated as an essential goal, such as in level set approximation or multiobjective optimization where the aim is to find a diverse set of alternative feasible or, respectively, Pareto optimal solutions. On the other hand, diversity maintenance can play an important role in algorithms that ultimately search for single optimal solutions. Examples are dynamical optimization and global optimization of multimodal objective functions, where diversity maintenance during a population-based search process increases the robustness of optimization algorithms or heuristics.
While the motivations to study diversity and its maintenance can be various, some salient issues reoccur such as effective strategies for diversity maintenance, indicators used to measure diversity, and the study of dynamic processes on set-valued state spaces. Although there is a growing attention to methods that address diversity as a search objective, the research is so far spread out across various disciplines and research schools who have developed independently terminologies and classification schemes, making it difficult to find relations between different works.
Although there is an increasing interest in diversity-oriented search a broad survey of this topic is missing so far and to provide it will be the goal of this work. This survey is intended to develop an integrated view of diversity-oriented optimization algorithms. Rather than going into details of implementations and adding new methods, the aim is to develop a systematic classification scheme. To integrate the various layers of algorithm design and various terminologies and methods that were developed in different research schools/areas, an ontology will be developed with the intention to ease the classification of existing and future algorithms in this field and identify overlapping and related areas of research.
This work is structured in two parts: The first part provides a review of research in diversity-oriented optimization, looking at it from different angles. In the second part, an ontology that integrates these views is developed and existing results are discussed in the light of this ontology.

2 Types of Optimization Problems

There are several problems, in which the concept of diversity is considered in the formulation of the goal:

Generating Alternatives: In engineering problems, when an algorithm provides the decision maker a solution of a *single objective* optimization problem, it might not always correspond 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) [Zechman and Ranjithan, 2004] 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 really suitable. The famous chemist Linus Pauling summarized the process of discovery [Pauling, 1960] as '*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 [van der Horst et al., 2012].

Multiobjective Optimization: In *multiobjective* optimization it is a common practice to compute a diverse set of Pareto optimal solutions. As opposed to the so-called a-priori approach, 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 [Knowles, 2009], 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-off between different objectives and the availability of solutions satisfying certain goals. 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 [Shir et al., 2009], [Coelho and Von Zuben, 2011], [Ulrich et al., 2010b], and [Schütze et al., 2008].

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 decision space and even discover some design patterns, if some common properties are found for all (or subset) of high-performing solutions [Parmee and Bonham, 2000]. This problem, including the methods for post-processing and data mining diverse solution sets, has recently been termed *Innovization* [Deb, 2011]. In [Reehuis et al., 2011], 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 basis of the existing models used in the domain.

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 [Schönemann et al., 2004]. In [Deb and Srinivasan, 2008] the goal to find *all* global optima (multi-global optimization) is termed *multi-global optimization*, both in single- and multi-objective optimization. [Stoean et al., 2010] stress the need of active diversity maintenance to be considered in the context of those strategies. Moreover, algorithms that seek to gain knowledge about the topological structure of multimodal landscapes are recently developed [Preuss and Wessing, 2013].

Model-based Diagnostics: Systems with cause and effect can be modeled as a function from input variables (causes) to output variables (effects). When performing model-based fault diagnosis, it can be of crucial importance to find all possible causes to a measured effect. If there are different possible causes, this indicates that further data is needed to identify the true cause. This aspect has been stressed in [Zechman and Ranjithan, 2009] for finding contaminant sources in water distribution networks, using evolutionary diversity-oriented optimization.

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 [Jin and Branke, 2005], it was pointed out that lack of diversity can lead to stagnation of population-based search in suboptimal regions in dynamically changing environments.

3 Biological Paradigms

When designing optimization algorithms within different biological population-based paradigms, maintaining diversity is necessary to prevent population from premature convergence, e.g. to local optima. Interestingly, each of the biological paradigms borrowed by artificial intelligence researchers from nature, approaches diversity in different ways. Terminology of algorithms and their components is often related to metaphors from the biological field, by which the paradigm is inspired.
In *Evolutionary Algorithms (EAs)*, the population evolves by selecting promising parents, recombining them and mutating for obtaining offspring, without any implicit diversity preservation mechanism. A common paradigm for maintaining diversity in EA populations is that of niching. For instance, *niching* [Shir, 2012] allows selecting only few solutions located within the same chunk of the objective or decision space, for parental or environmental selection.
In *Artificial Immune Systems (AISs)* based optimization algorithms, diversity is preserved intrinsically by *clonal selection theory* [Burnet, 1978] and *immune network theory* [Jerne, 1974] 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, like *affinity* among cells (individuals), guarantees that concentration of similar individuals decreases in the presence of better neighboring cells (the closer and higher the concentration of better cells, the higher the influence), and individuals without better solutions in their neighborhood increase concentration proportionally to their fitness.
In both EAs and AISs, the population evolves by variation of selected individuals. While in EAs, biological evolution happens by cumulative natural selection among individuals (parents recombination and mutation), in AISs adaptation happens by cumulative variation and selection within individuals (cloning and hypermutation). Among other applications, AISs are successfully applied for diversity oriented search, for instance to increase decision space diversity in multioobjective optimization [Coelho and Von Zuben, 2011].
The approach of *Swarm Intelligence (SI)* biological paradigm to diversity is similar to that of AISs, following self-adaptation of swarm to the environment and/or communication between swarm members-agents when necessary [Beekman et al., 2008]. For instance, in *Artificial Bee Colony (ABC)* algorithms, special type of swarm bee-agents, called *scouts*, are activated on the last exploration stage of the algorithm to promote search 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 [Branke et al., 2000].
Biological paradigms are also addressed in spatial population structures as opposed to panmictic ones. Examples are cellular GA [Alba and Dorronsoro, 2008] and spatial predator-prey algorithms in multiobjective optimization [Laumanns et al., 1998]. Besides biological paradigms, the mathematical programming community has developed several algorithms for diversity-oriented optimization that exploit mathematical structures of functions expressing diversity [Zadorojniy et al., 2012].

4 Diversity Indicators and Indicator-based Algorithms

A recent trend is to stress the quality performance measure in the design of an algorithm. In fact, the definition of diversity and how to combine diversity and optimality are by themselves topics of much debate and research. 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 multiobjective optimization problems). 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 [Jost, 2006] and [Weitzman, 1992]. A diversity index measuring the number of different points and also the spread of solutions is recommended.
Different diversity indicators were proposed in the context of diversity-oriented search. Ulrich et al. [Ulrich and Thiele, 2011] suggested to chose indicators from bio-diversity. In this field the Weitzmann diversity [Weitzman, 1992] and the Solow Polasky indicator [Solow et al., 1993] are common indicators. While the Weitzmann indicator [Weitzman, 1992] is motivated by phylogenetic trees with maximum parsimony and has exponential time complexity, the latter 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 is favored by Ulrich et al. [Ulrich and Thiele, 2011]. Even faster are indicators based on simple statistics on gaps between nearest neighbors [Emmerich et al., 2012], [Preuss and Wessing, 2013], although they can only provide comparisons among populations of the same size.
Each of the indicators allows comparing algorithms with respect to one of several properties, among which are quality of individuals sets, 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 [Zitzler and Künzli, 2004]. 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. [Emmerich et al., 2012], or as an integral part of an indicator, see e.g. [Ulrich et al., 2010b]. Ulrich et al. [Ulrich et al., 2010a] emphasized the importance of decision space diversity by suggesting diversity-optimizing single objective (NOAH) and multiobjective (DIOP) algorithms, see [Ulrich and Thiele, 2011] and [Ulrich et al., 2010a], respectively, as well as an algorithm that integrates diversity within the hypervolume indicator, see DIVA algorithm in [Ulrich et al., 2010b].
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 [Emmerich et al., 2012], and Hausdorff distance-based indicator was suggested for level set approximation within Evolutionary Level Set Approximation (ELSA) algorithm. Indicators for multimodal optimization are discussed in [Preuss and Wessing, 2013].

5 Diversity-Specific Applications

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 [Ulrich, 2012], truss bridge design and optimization [Ulrich and Thiele, 2011], drug discovery [van der Horst et al., 2012], quantum control [Shir et al., 2008], and space mission design [Schütze et al., 2008].

Figure 1: Diversity Oriented Optimization Taxonomy.

6 Taxonomy and Ontology

In artificial intelligence, ontologies are used to formally represent a knowledge domain [Gruber, 1993]. 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 with Figure 1, and Figure 2, respectively, and have been developed with the Protégé ontology editor (http://protege.stanford.edu/).

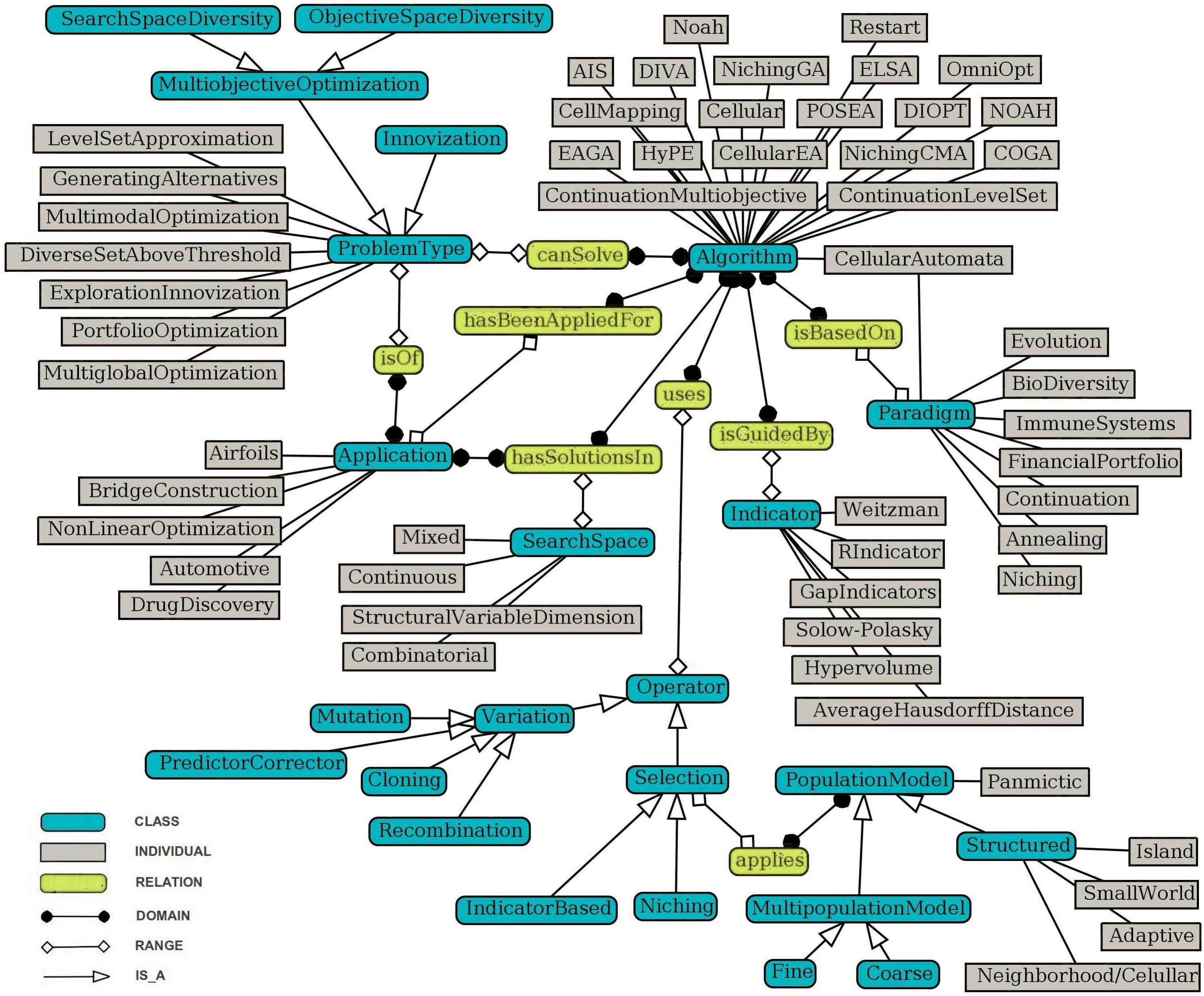

Figure 2: Diversity Oriented Optimization Ontology.

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 shown 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.

References

[Alba and Dorronsoro, 2008] Alba, E. and Dorronsoro, B. (2008). *Cellular genetic algorithms*, volume 42. Springer.

[Beekman et al., 2008] Beekman, M., Sword, G. A., and Simpson, S. J. (2008). Biological foundations of swarm intelligence. In Blum, C. and Merkle, D., editors, *Swarm Intelligence*, Natural Computing Series, pages 3–41. Springer-Verlag, Berlin, Heidelberg.

[Branke et al., 2000] Branke, J., Kaußler, T., Smidt, C., and Schmeck, H. (2000). A multi-population approach to dynamic optimization problems. In *Evolutionary Design and Manufacture*, pages 299–307. Springer.

[Burnet, 1978] Burnet, F. (1978). Clonal selection and after. *Theoretical Immunology*, 63:85.

[Coelho and Von Zuben, 2011] Coelho, G. P. and Von Zuben, F. J. (2011). A concentration-based artificial immune network for multi-objective optimization. In *Evolutionary Multi-Criterion Optimization*, pages 343–357. Springer.

[Deb, 2011] Deb, K. (2011). Innovization: discovering innovative solution principles through optimization. *Springer, Sept*, 2012:9–10.

[Deb and Srinivasan, 2008] Deb, K. and Srinivasan, A. (2008). Innovization: Discovery of innovative design principles through multiobjective evolutionary optimization. Multiobjective Problem Solving from Nature Natural Computing Series, pages 243–262.

[Emmerich et al., 2012] Emmerich, M., Deutz, A., and Kruisselbrink, J. (2012). On quality indicators for black-box level set approximation. volume 447 of *Studies in Computational Intelligence*, pages 157–185. Springer-Verlag, Berlin, Heidelberg.

[Gruber, 1993] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2):199–220.

[Jerne, 1974] Jerne, N. K. (1974). Towards a network theory of the immune system. *Ann. Immunol. (Inst. Pasteur)*, 125C:373–389.

[Jin and Branke, 2005] Jin, Y. and Branke, J. (2005). Evolutionary optimization in uncertain environments-a survey. *Evolutionary Computation, IEEE Transactions on*, 9(3):303–317.

[Jost, 2006] Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2):363–375.

[Knowles, 2009] Knowles, J. (2009). Closed-loop evolutionary multiobjective optimization. *IEEE Computational Intelligence Magazine*, 4(3):77–91.

[Laumanns et al., 1998] Laumanns, M., Rudolph, G., and Schwefel, H.-P. (1998). A spatial predator-prey approach to multi-objective optimization: A preliminary study. In *Parallel Problem Solving from Nature - PPSN V*, pages 241–249. Springer.

[Parmee and Bonham, 2000] Parmee, I. C. and Bonham, C. (2000). Towards the support of innovative conceptual design through interactive designer/evolutionary computing strategies. *AI EDAM*, 14(01):3–16.

[Pauling, 1960] Pauling, L. (1960). *The nature of the chemical bond and the structure of molecules and crystals: an introduction to modern structural chemistry*, volume 18. Ithaca, NY, Cornell University Press.

[Preuss and Wessing, 2013] Preuss, M. and Wessing, S. (2013). Measuring multimodal optimization solution sets with a view to multiobjective techniques. In *EVOLVE-A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation IV*, pages 123–137. Springer.

[Reehuis et al., 2011] Reehuis, E., Kruisselbrink, J., Olhofer, M., Graening, L., Sendhoff, B., and Bäck, T. (2011). Model-guided evolution strategies for dynamically balancing exploration and exploitation. In et al., J. H., editor, *Proceedings of the 10th International Conference on Artificial Evolution, EA 2011*, pages 306 – 317.

[Schönemann et al., 2004] Schönemann, L., Emmerich, M., and Preuß, M. (2004). On the extinction of evolutionary algorithm subpopulations on multimodal landscapes. *Informatica (Slowenien)*, 28(4):345–351.

[Schütze et al., 2008] Schütze, O., Vasile, M., and Coello, C. A. C. (2008). Approximate solutions in space mission design. In Rudolph, G., Jansen, T., Lucas, S. M., Poloni, C., and Beume, N., editors, *PPSN*, volume 5199 of *Lecture Notes in Computer Science*, pages 805–814. Springer.

[Shir, 2012] Shir, O. (2012). Niching in evolutionary algorithms. In Rozenberg, G., Bäck, T., and Kok, J., editors, *Handbook of Natural Computing*, pages 1035–1069. Springer Berlin Heidelberg.

[Shir et al., 2008] Shir, O., Beltrani, V., Bäck, T., Rabitz, H., and Vrakking, M. (2008). On the diversity of multiple optimal controls for quantum systems. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 41(7):074021.

[Shir et al., 2009] Shir, O., Preuss, M., Naujoks, B., and Emmerich, M. (2009). Enhancing decision space diversity in evolutionary multiobjective algorithms. volume 5467 of *Studies in Computational Intelligence*, pages 95–109. Springer-Verlag, Berlin, Heidelberg.

[Solow et al., 1993] Solow, A., Polasky, S., and Broadus, J. (1993). On the measurement of biological diversity. *Journal of Environmental Economics and Management*, 24(1):60–68.

[Stoean et al., 2010] Stoean, C., Preuss, M., Stoean, R., and Dumitrescu, D. (2010). Multimodal optimization by means of a topological species conservation algorithm. *IEEE Transactions on Evolutionary Computation*, 14(6):842–864.

[Ulrich, 2012] Ulrich, T. (2012). *Exploring Structural Diversity in Evolutionary Algorithms*. PhD thesis, TIK Institut für Technische Informatik und Kommunikationsnetze.

[Ulrich et al., 2010a] Ulrich, T., Bader, J., and Thiele, L. (2010a). Defining and optimizing indicator-based diversity measures in multiobjective search. In *Proceedings of the 11th international conference on Parallel problem solving from nature: Part I*, PPSN'10, pages 707–717, Berlin, Heidelberg. Springer-Verlag.

[Ulrich et al., 2010b] Ulrich, T., Bader, J., and Zitzler, E. (2010b). Integrating decision space diversity into hypervolume-based multiobjective search. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, GECCO '10, pages 455–462, New York, NY, USA. ACM.

[Ulrich and Thiele, 2011] Ulrich, T. and Thiele, L. (2011). Maximizing population diversity in single-objective optimization. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, GECCO '11, pages 641–648, New York, NY, USA. ACM.

[van der Horst et al., 2012] van der Horst, E., Marqués-Gallego, P., Mulder-Krieger, T., van Veldhoven, J., Kruisselbrink, J., Aleman, A., Emmerich, M. T., Brussee, J., Bender, A., and IJzerman, A. P. (2012). Multi-objective evolutionary design of adenosine receptor ligands. *Journal of Chemical Information and Modeling*, 52(7):1713–1721.

[Weitzman, 1992] Weitzman, M. L. (1992). On diversity. *The Quarterly Journal of Economics*, 107(2):363–405.

[Zadorojniy et al., 2012] Zadorojniy, A., Masin, M., Greenberg, L., Shir, O. M., and Zeidner, L. (2012). Algorithms for finding maximum diversity of design variables in multi-objective optimization. *Procedia Computer Science*, 8:171–176.

[Zechman and Ranjithan, 2004] Zechman, E. and Ranjithan, S. (2004). An evolutionary algorithm to generate alternatives (EAGA) for engineering optimization problems. *Engineering Optimization*, 36(5):539–553.

[Zechman and Ranjithan, 2009] Zechman, E. and Ranjithan, S. (2009). Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system. *Journal of Water Resources Planning and Management*, 135(5):334–343.

[Zitzler and Künzli, 2004] Zitzler, E. and Künzli, S. (2004). Indicator-based selection in multiobjective search. In *Parallel Problem Solving from Nature-PPSN VIII*, pages 832–842. Springer.