

32 Niching in Evolutionary Algorithms

Ofer M. Shir

Department of Chemistry, Princeton University, NJ, USA

oshir@princeton.edu

1	<i>Introduction</i>	1036
2	<i>Background: From Speciation to Ecological Optima</i>	1037
3	<i>Population Diversity in Evolutionary Algorithms</i>	1041
4	<i>Evolutionary Algorithms Niching Techniques</i>	1044
5	<i>Experimental Methodology</i>	1055
6	<i>The Niche-Radius Problem</i>	1060
7	<i>Discussion and Outlook</i>	1064

Abstract

Niching techniques are the extension of standard evolutionary algorithms (EAs) to multimodal domains, in scenarios where the location of multiple optima is targeted and where EAs tend to lose population diversity and converge to a solitary basin of attraction. The development and investigation of EA niching methods have been carried out for several decades, primarily within the branches of genetic algorithms (GAs) and evolution strategies (ES). This research yielded altogether a long list of algorithmic approaches, some of which are bio-inspired by various concepts of organic speciation and ecological niches, while others are more computational-oriented. This chapter will lay the theoretical foundations for niching, from the perspectives of biology as well as optimization, provide a summary of the main contemporary niching techniques within EAs, and discuss the topic of experimental methodology for niching techniques. This will be accompanied by the discussion of specific case-studies, including the employment of the popular covariance matrix adaptation ES within a niching framework, the application to real-world problems, and the treatment of the so-called niche radius problem.

1 Introduction

Evolutionary Algorithms (EAs) have the tendency to lose diversity within their population of feasible solutions and to converge into a single solution (Bäck 1994, 1996; Mahfoud 1995a), even if the search landscape has multiple globally optimal solutions.

Niching methods, the extension of EAs to finding multiple optima in multimodal optimization within one population, address this issue by maintaining the diversity of certain properties within the population. Thus, they aim at obtaining parallel convergence into multiple attraction basins in the multimodal landscape within a single run.

The study of niching is challenging both from the theoretical point of view and from the practical point of view. The theoretical challenge is twofold – maintaining diversity within a population-based stochastic algorithm from the computational perspective, and also gaining an insight into *speciation* theory or *population genetics* from the evolutionary biology perspective. The practical aspect provides a real-world incentive for this problem – there is an increasing interest of the *applied optimization community* in providing decision makers with multiple solutions that ideally represent different conceptual designs, for single-criterion or multi-criterion search spaces (Avigad et al. 2004, 2005). The concept of “*going optimal*” is often extended nowadays into the aim of “*going multi optimal*”: **obtaining optimal results but also providing the decision maker with a variety of different options.** (The *Second Toyota Paradox* (Cristiano et al. 2001), which is often reviewed in management studies, promotes the consideration of multiple candidate solutions during the car production process: “Delaying decisions, communicating ambiguously, and pursuing an excessive number of prototypes, can produce better cars faster and cheaper”).

Among the three conventional streams of EAs (Bäck 1996) – *genetic algorithms* (GAs) (Goldberg 1989), *evolution strategies* (ES) (Beyer and Schwefel 2002), and evolutionary programming (EP) (Fogel 1966) – niching methods have been studied in the past four decades almost exclusively within the framework of GAs. However, niching methods have been mostly

a by-product of studying *population diversity*, and were hardly ever at the front of the evolutionary computation (EC) research as an independent subfield.

This chapter thus aims at achieving three main goals:

1. Laying the theoretical foundations for niching, both from the biological as well as the computational perspectives
2. Providing a summary of the main contemporary niching techniques within evolutionary algorithms
3. Discussing the topic of experimental methodology for niching techniques, and proposing a framework for it

We shall begin by providing the natural computing motivation for niching by means of the biological background for the evolutionary process of *speciation*, given in ♦ Sect. 2, where it will also be linked to the domain of function optimization. The important topic of *population diversity* within evolutionary algorithms is discussed in detail in ♦ Sect. 3, focusing on the two canonical streams of GAs and ES. This is followed by an overview of the existing niching techniques within EAs, in ♦ Sect. 4. A case study of a specific niching technique with the popular covariance matrix adaptation evolution strategy (CMA-ES) receives special attention by means of a detailed description in the end of that section. The topic of experimental methodology for research on niching is discussed in ♦ Sect. 5, where proposed synthetic test functions as well as recommended performance criteria are outlined. Niching with the CMA-ES is revisited in that section, as an experimental observation case study. ♦ Section 6 is dedicated to a crucial and challenging issue for niching methods, the so-called *niche radius problem*. Finally, ♦ Sect. 7 concludes this chapter and proposes directions for future research in this domain.

A Note on the Scope It should be stressed that there exist niching methods in other natural computing frameworks, such as in particle swarm optimization (PSO) or in ant colony optimization (ACO) (for general field reviews see Kennedy and Eberhart (2001) and Engelbrecht (2005) for PSO or Dorigo and Stützle (2004) and Blum (2005) for ACO). These frameworks exceed the scope of this chapter, which solely focuses on evolutionary algorithms. We refer the reader who wishes to learn about niching techniques in PSO to Brits et al. (2002) and Parrott and Li (2006), or to Angus (2006) in ACO.

A Note on Notation Upon consideration of optimization problems, this chapter will assume *minimization*, without loss of generality, unless specified otherwise.

2 Background: From Speciation to Ecological Optima

This section constitutes the introduction to niching, and thus covers a diverse set of interdisciplinary topics. It will review the *biological* elementary concepts that correspond to the core of niching methods, such as *population diversity* and *speciation*, while mainly relying on Freeman and Herron (2003). It will then make the linkage to computing, by shifting to the optimization arena and discussing the equivalent to ecological niches: *basins of attraction*.

A Note on Terminology A species is defined as the *smallest evolutionary independent unit*. The term *niche*, however, stems from ecology, and it has several different definitions. It is sometimes referred to as the collective environmental components which are favored by a specific species, but could also be considered as the ecosystem itself which hosts individuals of

various species. Most definitions would typically also consider the *hosting capacity* of the niche, which refers to the limited available resources for sustaining life in its domain. In the context of function optimization, *niche* is associated with a *peak*, or a *basin of attraction*, whereas a *species* corresponds to the subpopulation of individuals occupying that *niche*.

2.1 Preliminary: Organic Evolution and Genetic Drift

Organic evolution can be broken down into four defining fundamental mechanisms: *natural selection*, *mutation*, *migration* or *gene flow*, and *genetic drift*. The latter, which essentially refers to *sampling errors in finite populations*, was overlooked by Darwin, who had not been familiar with Mendelian genetics, and thus did not discuss this effect in his *Origin of Species* (Darwin 1999). We assume that the reader is familiar with the first three fundamental elements, that is, selection–mutation–migration, and would like to elaborate on the drift effect, as it is the least intuitive mechanism among the four aforementioned evolutionary forces. In short, *genetic drift* (Fisher 1922; Wright 1931; Kimura 1983) is a stochastic process in which the diversity is lost in finite populations. A distribution of genetic properties is transferred to the following generation in a limited manner, due to the finite number of generated offspring, or equivalently, the limited statistical sampling of the distribution. As a result, the distribution is likely to approach an *equilibrium distribution*, for example, fixation of specific alleles when subject to equal fitness. This is why *genetic drift* is often considered as a *neutral effect*. The smaller the population, the faster and stronger this effect occurs. An analogy is occasionally drawn between genetic drift to *Brownian motion* of particles in mechanics.

The *genetic drift* effect had been originally recognized by Fisher (1922) (referred to as *random survival*), and was explicitly mentioned by Wright when studying Mendelian populations (Wright 1931). It was, however, revisited and given a new interpretation in the *Neutral Theory of Molecular Evolution* of Kimura (1983). The *neutral theory* suggested that the *random genetic drift* effect is the main driving force within molecular evolution, rather than the *nonrandom natural selection* mechanism. *Natural selection* as well as *genetic drift* are considered nowadays, by the contemporary evolutionary biology community, as the combined driving force of organic evolution. Moreover, the importance of the *neutral theory* is essentially in its being a **null hypothesis model** for the *natural selection theory*, by definition.

2.2 Organic Diversity

Diversity among individuals or populations in nature can be attributed to different evolutionary processes which occur at different levels. We distinguish here between variations that are observed within a single species to a *speciation* process, during which a new species arises, and both of them are reviewed shortly.

2.2.1 Variations Within a Species

Diversity of organisms within a single species stems from a variance at the genotypic level, referred to as *genetic diversity*, or from the existence of a spectrum of phenotypic realizations to a specific genotype. These effects are quantified and are usually associated with *genotypic variance* and *phenotypic variance*, respectively. Several hypotheses explaining *genetic diversity*

have been proposed within the discipline of *population genetics*, including the *neutral evolution theory*. It should be noted that genetic diversity is typically considered to be advantageous for survival, as it may allow better adaptation of the population to environmental changes, such as climate variations, diseases, etc.

Phenotypic variance is measured on a continuous spectrum, also known as quantitative variation. Roughly speaking, the main sources of quantitative variations (Freeman and Herron 2003; Scheiner and Goodnight 1984) are outlined here:

1. Genes have *multiple loci*, and hence are mapped onto a large set of phenotypes.
2. *Environmental effects* have direct influence on natural selection; fitness is time-dependent, and thus phenotypic variations in the outcome of selection are expected to occur.
3. *Phenotypic plasticity* is the amount in which the *genotypic expression* vary in different environments. (Bradshaw (1965) gave the following qualitative definition to phenotypic plasticity: “The amount by which the expressions of individual characteristics of a genotype are changed by different environments is a measure of the plasticity of these characters.”) And it is a direct source of variation at the phenotypic level.
4. The plastic response of the genotype to the environment, that is, the joint effect of genetic and environmental elements, also affects the selection of a specific phenotype, and thus can lead to variations. This effect is known as *genotype–environment interaction* (“G-by-E”).

Thus, *quantitative variations* are mainly caused by genotypic and phenotypic realizations and their interaction with the environment. The ratio between *genetic variance* to total *phenotypic variance* is defined as *heritability* (Wright 1931).

2.2.2 Speciation

Speciation, on the other hand, is the process during which a new species arises. In this case, statistical disassociation, which is the trigger to speciation, originates from gradually decreasing physical linkage.

The essence of the speciation process is **lack of gene flow**, where physical isolation often plays the role of the barrier to gene flow. Lack of gene flow is only one of the necessary conditions for speciation. Another necessary condition for speciation to occur is that the reduction of gene flow will be followed by a phase of **genetic divergence**, by means of *mutation*, *selection*, or *drift*. Finally, the completion or elimination of divergence can be assessed via the so-called *secondary contact* phase: interbreeding between the parental populations would possibly fail (offspring is less fit), succeed (offspring is fitter), or have a neutral outcome (offspring has the same fitness). This would correspond, respectively, to increasing, decreasing, or stabilizing the differentiation between the two arising species. Note that the speciation can occur *de facto*, without the actual secondary contact taking place; the latter is for observational assessment purposes.

In organic evolution, four different levels of *speciation* are considered, corresponding to four levels of physical linkage between the subpopulations:

1. **Allopatric speciation** The split in the population occurs only due to complete geographical separation, for example, migration or mountain building. It results in two geographically isolated populations.

2. **Peripatric speciation** Species arise in small populations, which are not geographically separated but rather isolated in practice; the effect occurs mainly due to the *genetic drift* effect.
3. **Parapatric speciation** The geographical separation is limited, with a physical overlap between the two zones where the populations split from each other.
4. **Sympatric speciation** The two diverging populations coexist in the same zone, and thus the speciation is strictly non-geographical. This is observed in nature in parasite populations, that are located in the same zone, but associated with different plant or animal hosts (McPherson et al. 1988).

These four modes of speciation correspond to four levels of geographically decreasing linkages. Generally speaking, *statistical association* of genetic components in nature, such as *loci*, typically results from *physical linkage*. In this case, we claim that statistical disassociation, which is the trigger to speciation, originates from gradually decreasing physical linkage.

In summary, speciation typically occurs throughout three steps:

1. Geographic isolation or reduction of gene flow
2. Genetic divergence (mutation, selection, drift)
3. Secondary contact (observation/assessment)

2.3 “Ecological Optima”: Basins of Attraction

We devote this section to the definition of basins of attraction. The task of defining a *generic basin of attraction* seems to be one of the most difficult problems in the field of *global optimization*, and there have only been few attempts to treat it theoretically. (Intuitively, and strictly metaphorically speaking, we may think of a *region of attraction* of \mathbf{x}_L as the region, where if water is poured, it will reach \mathbf{x}_L . Accordingly, we may then think of the basin of \mathbf{x}_L as the maximal region that will be covered when the cavity at \mathbf{x}_L is filled to the lowest part of its rim (Törn and Zilinskas 1987).)

Rigorously, it is possible to define the basin by means of a *local optimizer*. In particular, consider a gradient descent algorithm starting from \mathbf{x}_0 , which is characterized by the following dynamics:

$$\frac{d\mathbf{x}(t)}{dt} = -\nabla f(\mathbf{x}(t)) \quad (1)$$

with the initial condition $\mathbf{x}(0) = \mathbf{x}_0$. Now, consider the set of points for which the limit exists:

$$\mathcal{V} = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}(0) = \mathbf{x} \wedge \mathbf{x}(t)|_{t \geq 0} \text{ satisfies Eq. 1} \wedge \lim_{t \rightarrow \infty} \mathbf{x}(t) \text{ exists} \right\} \quad (2)$$

Definition 1 The *region of attraction* $A(\mathbf{x}_L)$ of a local minimum, \mathbf{x}_L , is

$$A(\mathbf{x}_L) = \left\{ \mathbf{x} \in \mathcal{V} \mid \mathbf{x}(0) = \mathbf{x} \wedge \mathbf{x}(t)|_{t \geq 0} \text{ satisfies Eq. 1} \wedge \lim_{t \rightarrow \infty} \mathbf{x}(t) = \mathbf{x}_L \right\} \quad (3)$$

The *basin* of \mathbf{x}_L is the **maximal level set** that is fully contained in $A(\mathbf{x}_L)$.

In the case of several disconnected local minima with the same function value, it is possible to define the region of attraction as the union of the nonoverlapping connected sets.

2.4 Classification of Optima: The Practical Perspective

On a related note to the theoretical definition of the basin, the practical perspective for the *classification of optima shapes*, also referred to as global topology, is worth mentioning. This topic is strongly related to the emerging subfield of *robustness study* (see, e.g., Tsui 1992), which aims at attaining optima of high quality with large basins (i.e., low partial derivative values in the proximity of the peak). Moreover, yet visited from a different direction, another approach was introduced recently by Lunacek and Whitley for classifying different classes of multimodal landscapes with respect to algorithmic performance (Lunacek and Whitley 2006). The latter defines the *dispersion metric* of a landscape as the degree to which the local optima are globally clustered near one another. Landscapes with low dispersion have their best local optima clustered together in a single *funnel*. (We deliberately avoid the definition of a funnel as it is rather vague. We refer the reader to Doye et al. (2004).) This classification of low dispersion and high dispersion may be associated with the algorithmic trade-off between exploration of the landscape and exploitation of local structures. Upon considering landscapes with multiple funnels, a recent study (Lunacek et al. 2008) investigated the impact of a global structure of two uneven funnels on the evolutionary search. It concluded that EAs tend to converge into the larger funnel, even if it is of suboptimal quality, and thus put their effectiveness in global exploration in multi-funnel landscapes into question.

3 Population Diversity in Evolutionary Algorithms

The term *population diversity* is commonly used in the context of Evolutionary Algorithms, but it rarely refers to a rigorous definition. Essentially, it is associated both with *genetic diversity* as well as with *speciation* – the two different concepts from organic evolution that were discussed in the previous section. This is simply due to the fact that the differences between the two concepts do not have any practical effect on the evolutionary search nor on the goal of maintaining diversity among the evolving candidate solutions. In the well-known trade-off between *exploration* and *exploitation* of the landscape during a search, *maintaining population diversity* is a driving force in the *exploration front*, and thus it is an important component. However, among EC researchers, population diversity is primarily considered as a component due to play a role in the important exploration of the landscape for the sake of obtaining a single solution, while its role in obtaining multiple solutions is typically considered as a secondary one.

Mahfoud's Formalism

Mahfoud constructed a formalism for characterizing *population diversity* in the framework of Evolutionary Algorithms (see Mahfoud 1995a, pp. 50–59). Mahfoud's formal framework was based on the partitioning of the search space into equivalence classes (set to *minima* in the search landscape), a descriptive relation (typically, *genotypic* or *phenotypic* mappings), and the measurement of distance between the current distribution of subpopulations to some predefined *goal distribution*.

Let P be a discrete distribution describing the current partitioning of the population into subpopulations, and let Q be the goal distribution of the population with respect to the defined sites. The formalism focuses in defining the *directed divergence*, or distance

of distribution P to distribution Q . Several well-known metrics follow this formalism by satisfying its various criteria, such as Shannon's information entropy, standard distance metrics, etc. (Mahfoud 1995a).

Diversity Loss

Subject to the complex dynamics of the various forces within an evolutionary algorithm, population diversity is typically lost, and the search is likely to converge into a single basin of attraction in the landscape. *Population diversity loss* within the population of solutions is the fundamental effect which niching methods aim to treat. In fact, from the historical perspective, the quest for diversity-promoting techniques was the main goal within the EC community for some time, and niching methods were merely obtained as *by-products*, so to speak, of that effort.

Due to the fundamental differences between GAs and ES, we choose to describe the effect of population diversity loss for each one of them separately.

3.1 Diversity Loss Within Genetic Algorithms

Mahfoud devoted a large part of his doctoral dissertation to studying population diversity within GAs (Mahfoud 1995a). He concluded that three main components can be attributed to the effect of population diversity loss within GAs.

Selection Pressure The traditional GA applies a probabilistic selection mechanism, namely, the *roulette-wheel selection* (RWS). This mechanism belongs to a broad set of selection mechanisms, which follow the fitness-proportionate selection principle. Selection pressure is thus associated with the first *moment of the selection operator*. It has been demonstrated by Mahfoud (1995a) that the selection pressure, or equivalently the nonzero expectation of the selection operator, prevents the algorithm from converging in parallel into more than a single attractor.

Selection Noise Selection noise is associated with the second *moment of the selection operator*, or its *variance*. Mahfoud (1995a) demonstrated that the high variance of the RWS, as well as of other selection mechanisms, is responsible for the fast convergence of a population into a single attractor, even when there exists a set of equally fit attractors. This effect can be considered as a *genetic drift* in its broad definition, that is, sampling error of a distribution.

Operator Disruption Evolutionary operators in general, and the *mutation* and *recombination* operators in particular, boost the evolution process toward exploration of the search space. In that sense, they have a constructive effect on the process, since they allow locating new and better solutions. However, their action also has a destructive effect. This is due to the fact that, by applying them, good solutions that have been previously located might be lost. Therefore, they may eliminate competition between highly fit individuals, and “assist” some of them to take over. The mutation operator usually has a small effect, since it acts in small steps – low mutation probability in the traditional GA, which means infrequent occurrence of bit flips. Thus, the mutation operator can be considered to have a negligible disruption. The recombination operator, on the other hand, has a more considerable effect. In the GA field, where the *crossover* operator is in use (single-point, two-point or n -point crossovers), the latter has been shown to have a disruptive nature by breaking desired patterns within the population (the well-known *schema theorem* discusses the schema disruption by the crossover operator

and states that schemata with high defining length will most likely be disrupted by the crossover operator; see, for example, Goldberg (1989)).

Corollary 1 *The traditional GA, which employs the standard set of operators, is exposed to statistical as well as disruptive effects that are responsible for the loss of population diversity. This outcome is likely to occur due to the first and second moments of the RWS operator, as well as to the disruptive nature of the crossover operator. We conclude that the traditional GA is expected to lose diversity among its candidate solutions.*

3.2 Diversity Loss Within Evolution Strategies

The defining mechanism of ES is strongly dictated by the mutation operator as well as by the deterministic selection operator. As defining operators, they have a direct influence on the diversity property of the population. The recombination operator, nevertheless, does not play a critical role in the ES mechanism.

We attribute two main components to the *population diversity loss* within ES: fast *take-over*, which is associated with the *selection* operator, and *genetic drift* (or *neutrality effect*), which is associated both with the *selection* and the *recombination* operators.

3.2.1 Selective Pressure: Fast Take-Over

Evolution strategies have a strictly deterministic, rank-based approach to selection. In the two traditional approaches (Bäck 1996) – (μ, λ) and $(\mu + \lambda)$ – the best individuals are selected, implying, rather intuitively, high *selective pressure*. Due to the crucial role of the selection operator within the evolution process, its impact within the ES field has been widely investigated.

Goldberg and Deb introduced the important concept of *take-over time* (Deb and Goldberg 1989), which gives a quantitative description of selective pressure **with respect only to the selection operator**.

Definition 2 The *take-over time*, τ^* , is the minimal number of generations until repeated application of the selection operator yields a uniform population filled with copies of the best individual.

The selective pressure has been further investigated by Bäck (1994), who analyzed all the ES selection mechanisms also with respect to take-over times. He concluded that upon employing the typical selection mechanisms, very short *take-over times* are yielded. This result implies that ES are typically subject to high *selective pressures*.

3.2.2 ES Genetic Drift

We consider two different ES neutral effects that could be together ascribed as a general ES genetic drift: *recombination drift*, and *selection drift*. We argue that these two components are responsible for the loss of population diversity within ES.

Recombination Drift

Beyer explored extensively the so-called *mutation-induced speciation by recombination* (MISR) principle (see, e.g., Beyer 1999). According to this important principle, repeated application of the mutation operator, subject to a dominant recombination operator, would lead to a stable distribution of the population, which resembles a species or a cloud of individuals. When fitness-based selection is applied, this cloud is likely to move together toward fitter regions of the landscape. Furthermore, Beyer managed to prove analytically (Beyer 1999) that the MISR principle is indeed universal when finite populations are employed, subject to sampling-based recombination. The latter was achieved by analyzing the ES dynamics without fitness-based selection, deriving the expected population variance, and showing that it is reduced with random sampling in finite populations. This result was also corroborated by numerical simulations. This study provides one with an analytical result that a sampling-based recombination is subject to genetic drift, and leads to loss of population diversity.

Selection Drift

A recent study on the extinction of subpopulations on a simple *bimodal equi-fitness* model investigated the drift effect of the selection operator (Schönemann et al. 2004). It considered the application of *selection* on finite populations, when the fitness values of the different attractors were equal (i.e., eliminating the possibility of a *take-over effect*), and argued that a neutral effect (*drift*) would occur, pushing the population into a single attractor. The latter study, indeed, demonstrated this effect of *selection drift* in ES, which resulted in a convergence to an equilibrium distribution around a single attractor. It was also shown that the time of extinction increases proportionally with μ . The analysis was conducted by means of Markov chain models, supported by statistical simulations.

Corollary 2 *Evolution Strategies that employ finite populations are typically underposed to several effects that are responsible for the loss of population diversity. It has been shown that the standard selection mechanisms may lead to a fast take-over effect. In addition, we argued that both the recombination and the selection operators experience their own drift effects that lead to population diversity loss. We concluded that an evolution strategy with a small population is likely to encounter a rapid effect of diversity loss.*

4 Evolutionary Algorithms Niching Techniques

Despite the fact that the motivation for multimodal optimization is beyond doubt, and the biological inspiration is real, there is no unique definition of the goal statement for *niching techniques*. There have been several attempts to provide a proper definition and functional specification for niching; we review here some of them:

1. Mahfoud (1995a) chose to put emphasis on locating as well as maintaining good optima, and formulated the following:
 - The litmus test for a niching method, therefore, will be whether it possesses the capability to find multiple, final solutions within a reasonable amount of time, and to maintain them for an extended period of time.

2. Beyer et al. (2002) put forward also the actual maintenance of population diversity:
 - *Niching*: process of separation of individuals according to their states in the search space or maintenance of diversity by appropriate techniques, for example, local population models, fitness sharing, or distributed EA.
3. Preuss (2006) considered the two definitions mentioned above, and proposed a third:
 - Niching in EAs is a two-step procedure that (a) concurrently or subsequently distributes individuals onto distinct basins of attraction and (b) facilitates approximation of the corresponding (local) optimizers.

We choose to adopt Preuss' mission statement and define **the challenge in niching** as follows:

- **Attaining the optimal interplay between partitioning the search space into niches occupied by stable subpopulations, by means of population diversity preservation, to exploiting the search in each niche by means of a highly efficient optimizer with local-search capabilities.**

Next, we shall provide an overview of existing niching techniques. Special attention will be given to a specific niching method which is based on a state-of-the-art evolution strategy, the so-called covariance matrix adaptation ES (CMA-ES). The latter is considered to be particularly efficient for high-dimensional continuous optimization, and will be described in greater detail.

4.1 GA Niching Methods

Niching methods within genetic algorithms have been studied during the past few decades, initially triggered by the necessity to promote *population diversity* within EAs. The research has yielded a variety of different methods, which are the vast majority of existing work on niching in general.

The remainder of this section will focus on GA niching techniques, by providing a short survey of the main known methods, with emphasis on the important concepts of *sharing* and *crowding*. This survey is mainly based on Mahfoud (1995a) and Singh and Deb (2006).

4.1.1 Fitness Sharing

The *sharing* concept was one of the pioneering niching approaches. It was first introduced by Holland (1975), and later implemented as a niching technique by Goldberg and Richardson (1987). This strong approach of **considering the fitness as a shared resource** has essentially become an important concept in the broad field of evolutionary algorithms, and laid the foundation for various successful niching techniques for multimodal function optimization, mainly within GAs. A short description of the *fitness sharing* mechanism follows.

The basic idea of *fitness sharing* is to consider the fitness of the landscape as a resource to be shared among the individuals, in order to decrease redundancy in the population. Given a similarity metric of the population, which can be *genotypic* or *phenotypic*, the *sharing function* is defined as follows:

$$sh(d_{i,j}) = \begin{cases} 1 - \left(\frac{d_{i,j}}{\rho}\right)^{\alpha_{sh}} & \text{if } d_{i,j} < \rho \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where $d_{i,j}$ is the distance between individuals i and j , ρ (traditionally noted as σ_{sh}) is the fixed radius of every niche, and $\alpha_{sh} \geq 1$ is a control parameter, typically set to 1. Using the *sharing function*, the *niche count* is given by

$$m_i = \sum_{j=1}^N sh(d_{i,j}) \quad (5)$$

where N is the number of individuals to be considered in the selection phase.

Let an individual raw fitness be denoted by f_i , then the *shared fitness* is defined by:

$$f_i^{sh} = \frac{f_i}{m_i} \quad (6)$$

assuming that the fitness is *strictly positive* and subject to *maximization*. The evaluation of the shared fitness is followed by the selection phase, which is typically based on the RWS operator (Goldberg 1989); The latter takes into consideration the shared fitness. Thus, the *sharing mechanism* practically penalizes individuals that have similar members within the population via their fitness, and by that it aims at reducing redundancy in the gene pool, especially around the peaks of the fitness landscape.

One important auxiliary component of this approach is the *niche radius*, ρ . Essentially, this approach makes a strong assumption concerning the fitness landscape, stating that the optima are far enough from one another with respect to the *niche radius*, which is estimated for the given problem and remains fixed during the course of evolution. Furthermore, it is important to note that the formulas for determining the value of ρ , which will be given in [Sect. 4.3](#), are dependent on q , the number of peaks of the target function. Hence, a second assumption is that q can be estimated. In practice, an accurate estimation of the expected number of peaks q in a given domain may turn out to be extremely difficult. Moreover, peaks may vary in shape, and this would make the task of determining ρ rather complicated. The aforementioned assumptions pose the so-called *niche radius problem*, to be discussed in [Sect. 6](#).

In the literature, several GA niching *sharing*-based techniques, which implement and extend the basic concept of sharing, can be found (Mahfoud 1995a; Goldberg 1987; Yin and Germany 1993; Jelasity 1998; Miller and Shaw 1996; Petrowski 1996; Cioppa et al. 2004). Furthermore, the concept of sharing was successfully extended to other “yields of interest,” such as *concept sharing* (Avigad et al. 2004).

4.1.2 Dynamic Fitness Sharing

In order to improve the *sharing mechanism*, a dynamic approach was proposed. The *dynamic niche sharing* method (Miller and Shaw 1996), which extended the *fitness sharing* technique, aimed at dynamically recognizing the q peaks of the forming niches, and based on that information, classified the individuals as either members of one of the niches, or as members of the “non-peaks domain.”

Explicitly, let us introduce the *dynamic niche count*:

$$m_i^{\text{dyn}} = \begin{cases} n_j & \text{if individual } i \text{ is within dynamic niche } j \\ m_i & \text{otherwise (non-peak individual)} \end{cases} \quad (7)$$

where n_j is the size of the j th dynamic niche (i.e., the number of individuals which were classified to niche j), and m_i is the standard *niche count*, as defined in [Eq. 5](#).

The shared fitness is then defined as follows:

$$f_i^{\text{dyn}} = \frac{f_i}{m_i^{\text{dyn}}} \quad (8)$$

The identification of the dynamic niches can be carried out by means of a *greedy* approach, as proposed in Miller and Shaw (1996) as the dynamic peak identification (DPI) algorithm (see Algorithm 1). As in the original *fitness sharing* technique, the *shared fitness evaluation* is followed by the selection phase, typically implemented with the RWS operator. Thus, this technique does not fixate the peak individuals, but rather provides them with an advantage in the selection phase, which is probability-based within GAs.

4.1.3 Clearing

Another variation to the *fitness sharing* technique, called *clearing*, was introduced by Petrowski (1996) at the same time as the *dynamic fitness sharing* (Miller and Shaw 1996). The essence of this mechanism is the “winner takes it all” principle, and its idea is to designate a specific number of individuals per niche, referred to as *winners*, which could enjoy the resources of that niche. This is equivalent to the introduction of a “death penalty” to the *losers* of the niche, the individuals of each niche that lose the generational competition to the actual peak individuals. Following a *radius-based* procedure of identifying the winners and losers of each niche in each generation, the winners are assigned with their raw-fitness values, whereas all the other individuals are assigned with *zero* fitness (*maximization* was assumed). This is called the *clearing phase*. The selection phase, typically based on the RWS operator, considers, *de facto*, only the winners of the different niches. The allowed number of winners per niche, also referred to as the *niche capacity*, is a control parameter that reflects the degree of elitism. In any case, as in the previous techniques, the peak individuals are never fixated, and are subject to the probabilistic selection of the GA. This method was shown to outperform the *fitness sharing* technique on a specific set of low-dimensional test problems (Petrowski 1996).

Algorithm 1 Dynamic Peak Identification

```

input: population Pop, number of niches q, niche radius ρ
1: Sort Pop in increasing fitness order {minimization}
2: i := 1
3: NumPeaks := 0
4: DPS := ∅ {Set of peak elements in population}
5: while NumPeaks ≠ q and i ≤ popSize do
6:   if Pop[i] is not within sphere of radius ρ around peak in DPS then
7:     DPS := DPS ∪ {Pop[i]}
8:     NumPeaks := NumPeaks + 1
9:   end if
10:  i := i + 1
11: end while
output: DPS
    
```

4.1.4 Crowding

Crowding was one of the pioneering methods in this field, as introduced by De Jong (1975). It considered, and to some extent generalized, *preselection schemes*, which had been investigated in the doctoral dissertation of Cavicchio (1970). The latter had showed that certain preselection schemes boosted the preservation of population diversity. The *crowding* approach aimed at reducing changes in the population distribution between generations, in order to prevent *premature convergence*, by means of *restricted replacement*. Next, we will describe the method in more detail.

Given the traditional GA, a proportion G of the population is selected in each generation via fitness-proportionate selection to undergo variations (i.e., *crossover* and *mutation*) – out of which a part is chosen to die and to be replaced by the new offspring. Each offspring finds the individuals it replaces by taking a random sample of CF (referred to as **crowding factor**) individuals from the population, and replacing the **most similar individual** from the sample. An appropriate *similarity metric* should be chosen.

The crucial point of this niching mechanism is the calculation of the so-called *crowding distance between parents and offspring*, in order to control the *change rate* between generations. A different use of the *crowding distance*, applied among individuals of the same generation and assigned with reversed ranking, is widely encountered in the context of evolutionary multiobjective optimization (EMOA) (Deb 2001; Coello Coello et al. 2007). In the joint context of niching and EMOA see also Deb’s “omni-optimizer” (Deb and Tiwari 2005).

Mahfoud, who analyzed the *crowding* niching technique (1995a), concluded that it was subject to disruptive effects, mainly *drift*, which prevented it from maintaining more than two peaks. He then proposed a mechanism called *deterministic crowding*, as an improvement to the original *crowding* scheme. The proposed procedure applies variation operators to pairs of individuals in order to generate their offspring, which are all then evaluated with respect to the crowding distance, and undergo *replacement selection* (see Algorithm 2, which assumes *maximization*).

4.1.5 Clustering

The application of *clustering* for niching is very intuitive from the computational perspective, as well as straightforward in its implementation. Yin et al. (1993) proposed a clustering

Algorithm 2 Deterministic Crowding: Replacement Selection (*maximization*)

- 1: Select two parents, p_1 and p_2 , randomly, without replacement
- 2: Generate two variations, c_1 and c_2
- 3: **if** $d(p_1, c_1) + d(p_2, c_2) \leq d(p_1, c_2) + d(p_2, c_1)$ **then**
- 4: **if** $f(c_1) > f(p_1)$ **then** replace p_1 with c_1
- 5: **if** $f(c_2) > f(p_2)$ **then** replace p_2 with c_2
- 6: **else**
- 7: **if** $f(c_2) > f(p_1)$ **then** replace p_1 with c_2
- 8: **if** $f(c_1) > f(p_2)$ **then** replace p_2 with c_1
- 9: **end if**

framework for niching with GAs, which we describe here briefly. A clustering algorithm, such as the *K-Means* algorithm (Haykin 1999), first partitions the population into niches, and then considers the *centroids*, or center points of mass, of the newly partitioned subpopulations.

Let d_{ic} denote the distance between individual i and its *centroid*, and let f_i denote the raw fitness of individual i . Assuming that there are n_c individuals in the niche of individual i , its fitness is defined as:

$$f_i^{\text{Clustering}} = \frac{f_i}{n_c \cdot (1 - (d_{ic}/2d_{\max})^\alpha)} \quad (9)$$

where d_{\max} is the maximal distance allowed between an individual and its niche centroid, and α is a defining parameter. It should be noted that the clustering algorithm uses an additional parameter, d_{\min} , for determining the minimal distance allowed between centroids, playing an equivalent role to the *niche radius* ρ of the *sharing*-based schemes.

This method is often subject to criticism for its strong dependency on a relatively large number of parameters. However, this *clustering* technique has become a popular kernel for niching with EAs, and its application was reported in various studies (see, e.g., Schönemann et al. 2004, Hanagandi and Nikolaou 1998, Branke 2001, Gan and Warwick 2001, Aichholzer et al. 2000, Streichert et al. 2003, and Ando et al. 2005).

4.1.6 The Sequential Niche Technique

A straightforward approach of *iteration* can be used to sequentially locate multiple peaks in the landscape, by means of an *iterative local search* (Ramalhinho-Lourenco et al. 2000). This procedure is blind to any information gathered in previous searches, and sequentially restarts stochastic search processes, hoping to hit a different peak every run. Obviously, it is likely to encounter *redundancy*, and the number of expected iterations is then increased by a factor. A **redundancy factor** can be estimated if the peaks are of equal height (equi-fitness landscape), that is, the probability to converge into any of the q peaks is equal to $1/q$:

$$R = \sum_{i=1}^q \frac{1}{i}$$

For $q > 3$, this can be approximated by:

$$R \approx \gamma + \ln(q) \quad (10)$$

where $\gamma \approx 0.577$ is the Euler–Mascheroni constant. This *redundancy factor* remains reasonably low for any practical value of q , but is expected to considerably increase if all optima are not likely to be found equal.

On a related note, we would like to mention a multirestart with an increasing population size approach, that was developed with the CMA-ES algorithm (Auger and Hansen 2005a). The latter aims at attaining the global minimum, while possibly visiting local minima along the process and restarting the algorithm with a larger population size and a modified initial step-size. It is not defined as a niching technique and does not target optima other than the global minimum, but it can capture suboptimal minima during its search.

Beasley et al. extended the naive *iteration* approach, and developed the so-called *Sequential Niche* technique (Beasley et al. 1993). This method, in contrast to the other niching methods presented earlier, does not modify the genetic operators nor any characteristics of the

traditional GA, but rather creates a general search framework suitable for locating multiple solutions. By means of this method the search process turns into a sequence of independent runs of the traditional GA, where the basic idea is to suppress the fitness function at the observed optimum that was obtained in each run, in order to prevent the search from revisiting that optimum.

In further detail, the traditional GA is run multiple times sequentially: given the best solution of each run, it is first stored as a possible final solution and, second, the fitness function is artificially suppressed in all the points within the neighborhood of that optimum up to a desired radius. This modification is done immediately after each run. Its purpose is to discourage the following runs from revisiting these optima, and by that to encourage the exploration of other areas of the search landscape – aiming at obtaining all its optima. It should be noted that each function modification might yield artificial discontinuities in the fitness landscape. This method focuses only on locating multiple optima of the given search problem, without considering the concepts of parallel evolution and formation of subpopulations. In that sense, it has been claimed that it could not be considered as a niching method, but rather as a modified iterated search.

4.1.7 The Islands Model

This is probably the most intuitive niching approach from the biological perspective, directly inspired by organic evolution. Also referred to as the *regional population model*, this approach (see, e.g., Grosso 1985; Adamidis 1994; Martin et al. 1997) simulates the evolution of subpopulations on remote computational units (independent processors), aiming at achieving a speciation effect by **monitoring the gene flow**. The population is divided into multiple subpopulations, which evolve independently for a fixed number of generations, called *isolation period*. This is followed by a phase of controlled gene flow, or *migration*, when a portion of each subpopulation migrates to other nodes.

The genetic diversity and the amount of information exchange between subpopulations are determined by the following parameters – the number of exchanged individuals, the *migration rate*, the selection method of the individuals for migration (uniformly at random, or elitist fitness-based approach), and the scheme of migration, for example, complete net topology, ring topology, or neighborhood topology.

4.1.8 Other GA-Based Methods

Tagging (see, e.g., Spears 1994 and Deb and Spears 1997) is a mechanism that aims at improving the distance-based methods of *fitness sharing* and *crowding*, by labeling individuals with tag-bits. Rather than carrying out distance calculations, the tag-bits are employed for identifying the subpopulations, enforcing *mating restrictions*, and then implementing the *fitness sharing* mechanism. An individual is classified to a subpopulation by its genetic inheritance, so to speak, which is subject to generational variations, rather than by its actual spatial state. This concept simplifies the classification process, and obviously reduces the computational costs per generation, and at the same time it introduces a new bio-inspired approach into niching: individuals belong to a species because their parents did, and not because they are currently adjacent to a “peak individual,” for instance. This technique was shown in Spears (1994) to be a rather efficient implementation of the *sharing* concept.

A complex subpopulation differentiation model, the so-called **multinational evolutionary algorithm**, was presented by Ursem (1999). This original technique considers a world of “nations,” “governments,” and “politicians,” with dynamics dictated by migration of individuals, merging of subpopulations, and selection. Additionally, it introduces a topology-based auxiliary mechanism of *sampling*, which detects whether feasible solutions share the same basin of attraction. Due to the *curse of dimensionality*, this sampling-based mechanism is expected to lose its efficiency in high-dimensional landscapes.

Stoean et al. (2005) constructed the so-called **elitist generational genetic chromodynamics algorithm**. The idea behind this radius-based technique was the definition of a *mating region*, a *replacement region*, and a *merging region* — with appropriate mating, replacement, and merging radii — which dictates the dynamics of the genetic operations.

4.1.9 Miscellaneous: Mating Schemes

It has been observed that once the niche formation process starts, that is, when the population converges into the multiple basins of the landscape, crossbreeding between different niches is likely to fail in producing good offspring. In biological terms, this is the elimination of the divergence, by means of *hybridization*, in the **secondary contact phase**, as discussed in [Sect. 2.2](#).

Deb and Goldberg (1989) proposed a so-called *mating restriction scheme*, which poses a limitation on the choice of partners in the reproduction phase and prevents recombination between competing niches. They employed a distance measure, subject to a distance threshold, which was set to the niche radius, and showed that it could be used to improve the *fitness sharing* algorithm.

Mahfoud (1995a) proved that the mating restriction scheme of Deb and Goldberg was not sufficient, *per se*, in maintaining the population diversity in GA niching. A different approach of Smith and Bonacina (2003), however, considered an evolutionary computation multi-agent system, as opposed to the traditional *centralized* EA, and did manage to show that the same mating restriction scheme in an agent-based framework was capable of maintaining diversity and converging with stability into the desired peaks.

From the biological perspective, the mating restriction scheme is obviously equivalent to keeping the geographical isolation, or the barrier to gene flow, in order to allow the completion of the speciation phase. As discussed earlier, the geographical element in organic evolution is the crucial component which creates the conditions for speciation, and it is not surprising that artificial niching techniques choose to enforce it, by means of mechanisms such as the niche radius or the mating restriction scheme.

4.2 ES Niching Methods

Researchers in the field of evolution strategies initially showed no particular interest in the topic of niching, leaving it essentially for genetic algorithms. An exception would be the employment of island models. Generally speaking, classical niching schemes such as *fitness sharing*, which redefine the selection mechanism, are likely to interfere with the core of evolution strategies – the *self-adaptation mechanism* – and thus doomed to fail in a straightforward implementation. Any manipulation of the fitness value is usually not suitable for evolution strategies, as in the case of constraints handling: death penalty is typically the chosen

approach for a violation of a constraint in ES, rather than a continuous penalty as used in other EAs, in order to avoid the introduction of disruptive effects to the self-adaptation mechanism (see, e.g., Coello Coello 1999; Kramer and Schwefel 2006). Therefore, niching with evolution strategies would have to be addressed from a different direction. Moreover, the different nature of the ES dynamics, throughout the *deterministic selection* and the *mutation operator*, suggests that an alternative treatment is required here.

There are several, relatively new, niching methods that have been proposed within ES, mostly clustering-based (Schönemann et al. 2004; Aichholzer et al. 2000; Streichert et al. 2003). In addition, niching was also introduced to the mixed-integer ES framework (Li et al. 2008). A different approach, based on derandomized evolution strategies (DES) (for the latter see, e.g., Ostermeier et al. (1993, 1994) and Hansen et al. (1995)), was presented by Shir and Bäck (2008). One of its variants, which employs the popular covariance matrix adaptation evolution strategy (CMA-ES), will receive special attention here, and will be set as a detailed case study of niching techniques, to be outlined in the following section.

4.2.1 Case Study: Niching with CMA-ES

A niching framework for $(1 + \lambda)$, derandomized ES (DES) kernels subject to a fixed niche radius has been introduced recently (see, e.g., Shir and Bäck 2008). Following the *mission statement* presented earlier, the aim was the construction of a generic niching framework, which offers the combination of population diversity preservation and local search capabilities. Thus, DES were considered as an excellent choice for that purpose, as EAs with local search characteristics. Furthermore, DES typically employ small populations, which was shown to be a potential advantage for a niching technique, as it can boost the speciation effect (Shir 2008).

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

The CMA-ES (Hansen and Ostermeier 2001), is a DES variant that has been successful in treating correlations among object variables by efficiently learning matching mutation distributions. Explicitly, given an initial search point $\mathbf{x}^{(0)}$, λ offspring are generated by means of normally distributed variations:

$$\mathbf{x}^{(g+1)} \sim \mathcal{N}\left(\langle \mathbf{x} \rangle_W^{(g)}, \sigma^{(g)^2} \mathbf{C}^{(g)}\right) \quad (11)$$

Here, $\mathcal{N}(\mathbf{m}, \mathbf{C})$ denotes a normally distributed random vector with mean \mathbf{m} and a covariance matrix \mathbf{C} . The best μ search points out of these λ offspring undergo weighted recombination and become the parent of the next generation, denoted by $\langle \mathbf{x} \rangle_W$. The covariance matrix \mathbf{C} is initialized as the *unity matrix* and is learned during the course of evolution, based on cumulative information of successful past mutations (the *evolution path*). The global step-size, $\sigma^{(g)}$, is updated based on information extracted from *principal component analysis* of $\mathbf{C}^{(g)}$ (the *conjugate evolution path*). For more details, we refer the reader to Hansen and Ostermeier (2001).

An elitist sibling to the CMA comma strategy was also introduced (Igel et al. 2006), based upon the classical $(1+1)$ -ES (Bäck 1996).

A Detailed Description of the Algorithm

This niching technique is based upon interacting search processes, which simultaneously perform a derandomized $(1, \lambda)$ or $(1 + \lambda)$ search in different locations of the space. In case of

multimodal landscapes these search processes are meant to explore different attractor basins of local optima.

An important point in this approach is to strictly enforce the fixed allocation of the population resources, that is, number of offspring per niche. The idea is thus to prevent a scenario of a take-over, where a subpopulation located at a fitter optimum can generate more offspring. The biological idea behind this fixed allocation of resources stems from the concept of limited *hosting capacities* of given ecological niches, as previously discussed.

The *speciation interaction* occurs every generation when all the offspring are considered together to become niches' representatives for the following iteration, or simply the next search points, based on the rank of their fitness and their location with respect to higher-ranked individuals. The focus here is on a simple framework without recombination ($\mu = 1$).

Given q , the estimated/expected number of peaks, $q + p$ "D-sets" are initialized, where a D-set is defined as the collection of all the dynamically adapted strategy as well as decision parameters of the CMA algorithm, which uniquely define the search at a given point of time. (A D-set originally referred to the *derandomized* set of strategy parameters. When the CMA kernel is in use, it is sometimes referred to in the literature as a CMA-set.) These parameters are the current search point, the covariance matrix, the step-size, as well as other auxiliary parameters. At every point in time the algorithm stores exactly $q + p$ D-sets, which are associated with $q + p$ search points: q for the peaks and p for the "non-peaks domain." The $(q + 1)th \dots (q + p)th$ D-sets are associated with individuals, which are randomly re-generated every *epoch*, that is, a cycle of κ generations, as potential candidates for niche formation. This is basically a *quasi-restart* mechanism, which allows new niches to form dynamically. Setting the value of p should reflect the trade-off between applying a wide restart approach for exploring further the search space to exploiting computational resources for the existing niches. In any case, due to the *curse of dimensionality*, p loses its significance as the dimension of the search space increases.

Until the stopping criterion is met, the following procedure takes place. Each search point samples λ offspring, based on its evolving D-set. After the fitness evaluation of the new $\lambda \cdot (q + p)$ individuals, the classification into niches of the entire population is obtained in a *greedy* fashion, by means of the DPI routine (Miller and Shaw 1996) (➤ [Algorithm 1](#)). The latter is based on the fixed niche radius ρ . The peaks then become the new search points, while their D-sets are inherited from their parents and updated according to the CMA defining equations.

A pseudocode for a single iteration in this *niching routine* is presented as ➤ [Algorithm 3](#).

Natural Interpretation: Alpha-Males Competition

We would like to point out the nature of the subpopulations dynamics in the aforementioned niching scheme. Due to the *greedy* classification to niches, which is carried out every generation, some niches can merge in principle, while all the individuals, except for the *peak individual*, die out in practice. Following the posed principle of fixed resources per niche, only the peak individual will be sampled λ times in the following generation. In socio-biological terms, the peak individual could be then associated with an **alpha-male**, which wins the local competition and gets all the sexual resources of its ecological niche. The algorithm as a whole can be thus considered as a competition between $q + p$ alpha-males, each of which is fighting for one of the available q "computational resources," after winning its local competition at the "ecological optimum" site. The domination battles take place locally every cycle, as dictated by the DPI scheme. An elitist-CMA kernel will allow aging alpha-males to keep participating in the ongoing competitions, whereas a comma strategy will force their replacement by fresh

Algorithm 3 $(1 + \lambda)$ -CMA-ES Niching with Fixed Niche Radius (A Single Iteration)

```

1: for  $i = 1 \dots (q + p)$  search points do
2:   Generate  $\lambda$  samples based on the D-set of  $i$ 
3: end for
4: Evaluate fitness of the population
5: Compute the Dynamic Peak Set with the DPI Algorithm
6: for all elements of  $DPS$  do
7:   Set peak as a search point
8:   Inherit the D-set and update it respectively
9: end for
10: if  $N_{DPS} = \text{size of } DPS < q$  then
11:   Generate  $q - N_{DPS}$  new search points, reset D-sets
12: end if
13: if  $gen \bmod \kappa \equiv 0$  then
14:   Reset the  $(q + 1)th \dots (q + p)th$  search points
15: end if

```

blood of new alpha-males. At the global level, the value of p determines the selection pressure of alpha-males; setting $p = 0$ will then eliminate global competition and will grant automatically λ offspring to each alpha-male in the following generation, respectively.

4.3 Niche-Radius Calculation

The original formula for ρ for *phenotypic sharing* in GAs was derived by Deb and Goldberg (1989). Analogously, by considering the decision parameters as the decoded parameter space of the GA, the same formula can be applied, using the Euclidean metric, to ES. Given q , the number of target peaks in the solution space, every niche is considered to be surrounded by an n -dimensional hypersphere with radius ρ , which occupies $\frac{1}{q}$ of the entire volume of the space. The volume of the hypersphere which contains the entire space is

$$V = cr^n \quad (12)$$

where c is a constant, given explicitly by:

$$c = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} \quad (13)$$

with $\Gamma(n)$ as the Gamma function. Given lower and upper bound values $x_{k,\min}$, $x_{k,\max}$ of each coordinate in the decision parameters space, r is defined as follows:

$$r = \frac{1}{2} \sqrt{\sum_{k=1}^n (x_{k,\max} - x_{k,\min})^2} \quad (14)$$

If we divide the volume into q parts, we may write

$$c\rho^n = \frac{1}{q} cr^n \quad (15)$$

which yields

$$\rho = \frac{r}{\sqrt[n]{q}} \quad (16)$$

Hence, by applying this niche radius approach, two assumptions are made:

1. The number of target peaks, q , is given or can be estimated.
2. All peaks are at least at distance 2ρ from each other, where ρ is the fixed radius of every niche.

5 Experimental Methodology

Since the topic of niching has not drawn considerable attention from the mainstream EC community, there have been no constructive attempts to generate a generalized experimental framework, for testing niching methods to be agreed upon. This section will focus on that, and propose an experimental methodology for EA niching algorithms. It will present a suite of synthetic test functions, discuss possible performance criteria, and conclude by revisiting the niching-CMA algorithm and some of its experimental observations.

In the broad context of function optimization, EA niching techniques are obviously within the general framework of stochastic algorithms, and as such should be treated carefully upon reporting their experiments. We recommend following Bartz-Beielstein (2006) when conducting experiments, and especially following the 7-points scheme of Preuss (2007) when reporting them.

5.1 Multimodal Test Functions

The choice of a numerical testbed for evaluating the performance of search or optimization methods is certainly one of the core issues among the scholars in the community of algorithms and operations research.

In a benchmark article, Whitley et al. (1996) criticized the commonly tested artificial landscapes in the evolutionary algorithms community, and offered general guidelines for constructing test problems. A remarkable effort was made almost a decade after that document, when a large group of scholars in the EC community joined their efforts and compiled an agreed test suite of single-objective artificial landscapes (Suganthan et al. 2005), to be tested in an open performance competition reported at the 2005 IEEE Congress on Evolutionary Computation (CEC) (Auger and Hansen 2005b). The latter also included multimodal functions.

The issue of developing a multimodal test suite received even less attention, likely due to historical reasons. Since multimodal domains were mainly treated by GA-based niching methods, their corresponding test suites were limited to low-dimensional continuous landscapes, typically with two decision parameters to be optimized ($n = 2$) (see, e.g., Goldberg and Richardson 1987; Mahfoud 1995a).

When compiling our proposed test suite, we aim at following Whitley's guidelines, and to include some traditional GA-niching test functions as well as functions from the 2005 CEC inventory (Suganthan et al. 2005). Some of the landscapes have symmetric or equal distributions of minima, and some do not. Some of the functions are *separable*, that is, they can be optimized by solving n 1-dimensional problems separately (Whitley et al. 1996), while some of them are non-separable.

■ Table 1

Test functions to be *minimized* and initialization domains. For some of the non-separable functions, we apply translation and rotation: $\mathbf{y} = \mathcal{O}(\mathbf{x} - \mathbf{r})$ where \mathcal{O} is an orthogonal rotation matrix, and \mathbf{r} is a shifting vector

Separable:			
Name	Function	Init	Niches
\mathcal{M}	$\mathcal{M}(\mathbf{x}) = -\frac{1}{n} \sum_{i=1}^n \sin^\alpha(5\pi x_i)$	$[0, 1]^n$	100
\mathcal{A} [Ackley]	$\mathcal{A}(\mathbf{x}) = -c_1 \cdot \exp\left(-c_2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(c_3 x_i)\right) + c_1 + e$	$[-10, 10]^n$	$2n+1$
\mathcal{L}	$\mathcal{L}(\mathbf{x}) = -\prod_{i=1}^n \sin^k(l_1 \pi x_i + l_2) \cdot \exp\left(-l_3 \left(\frac{x_i - l_4}{l_5}\right)^2\right)$	$[0, 1]^n$	$n+1$
\mathcal{R} [Rastrigin]	$\mathcal{R}(\mathbf{x}) = 10n + \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i))$	$[-1, 5]^n$	$n+1$
\mathcal{G} [Griewank]	$\mathcal{G}(\mathbf{x}) = 1 + \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right)$	$[-10, 10]^n$	5
\mathcal{S} [Shekel]	$\mathcal{S}(\mathbf{x}) = -\sum_{i=1}^{10} \frac{1}{k_i(\mathbf{x}-a_i)(\mathbf{x}-a_i)^T + c_i}$	$[0, 10]^n$	8
\mathcal{V} [Vincent]	$\mathcal{V}(\mathbf{x}) = -\frac{1}{n} \sum_{i=1}^n \sin(10 \cdot \log(x_i))$	$[0.25, 10]^n$	50
Non-separable:			
Name	Function	Init	Niches
\mathcal{F} [Fletcher-Powell]	$\mathcal{F}(\mathbf{x}) = \sum_{i=1}^n (A_i - B_i)^2$ $A_i = \sum_{j=1}^n (a_{ij} \cdot \sin(\alpha_j) + b_{ij} \cdot \cos(\alpha_j))$ $B_i = \sum_{j=1}^n (a_{ij} \cdot \sin(x_j) + b_{ij} \cdot \cos(x_j))$ $a_{ij}, b_{ij} \in [-100, 100]; \alpha \in [-\pi, \pi]^n$	$[-\pi, \pi]^n$	10
\mathcal{R}_{SR} [S.R. Rastrigin]	$\mathcal{R}_{SR}(\mathbf{x}) = 10n + \sum_{i=1}^n (y_i^2 - 10 \cos(2\pi y_i))$	$[-5, 5]^n$	$n+1$
\mathcal{G}_{SR} [S.R. Griewank]	$\mathcal{G}_{SR}(\mathbf{x}) = 1 + \sum_{i=1}^n \frac{y_i^2}{4000} - \prod_{i=1}^n \cos\left(\frac{y_i}{\sqrt{i}}\right)$	$[0, 600]^n$	5

We propose for consideration a set of multimodal test functions, as indicated in ► Table 1. The table summarizes the proposed unconstrained multimodal test functions as well as their initialization intervals and the number of target niches. Next, we provide the reader with an elaborate description of the test functions, corresponding to the notation of ► Table 1:

- \mathcal{M} is a basic hypergrid multimodal function with uniformly distributed minima of equal function value of -1 . It is meant to test the stability of a particularly large number of niches: in the interval $[0, 1]^n$ it has 5^n minima. We set $\alpha = 6$.
- The well-known Ackley function has one global minimum, regardless of its dimension n , which is surrounded isotropically by $2n$ local minima in the first hypersphere, followed by an exponentially increasing number of minima in successive hyperspheres. Ackley's function has been widely investigated in the context of *evolutionary computation* (see, e.g., Bäck 1996). We set $c_1 = 20$, $c_2 = 0.2$, and $c_3 = 2\pi$.

- \mathcal{L} – also known as $F2$, as originally introduced by Goldberg and Richardson (1987) – is a sinusoid trapped in an exponential envelope. The parameter k determines the sharpness of the peaks in the function landscape; We set it to $k = 6$. \mathcal{L} has one global minimum, regardless of n and k . It has been a popular test function for GA niching methods. We set $l_1 = 5.1$, $l_2 = 0.5$, $l_3 = 4 \cdot \ln(2)$, $l_4 = 0.0667$ and $l_5 = 0.64$.
- The Rastrigin function (Törn and Zilinskas 1987) has one global minimum, surrounded by a large number of local minima arranged in a lattice configuration.

We also propose its shifted-rotated variant (Suganthan et al. 2005), with a linear transformation matrix of condition number 2 as the rotation operator.

- The Griewank function (Törn and Zilinskas 1987) has its global minimum ($f^* = 0$) at the origin, with several thousand local minima in the area of interest. There are four suboptimal minima $f \approx 0.0074$ with $\mathbf{x}^* \approx (\pm\pi, \pm\pi\sqrt{2}, 0, 0, 0, \dots 0)$.

We also propose its shifted-rotated variant (Suganthan et al. 2005), with a linear transformation matrix of condition number 3 as the rotation operator.

- The Vincent function is a sine function with a decreasing frequency. It has 6^n global minima in the interval $[0.25, 10]^n$.
- The Shekel function, suggested by Törn and Zilinskas (1987), introduces a landscape with a dramatically uneven spread of minima. It has one global minimum, and seven ordered local minima.
- The function after Fletcher and Powell (Bäck 1996) is a non-separable *nonlinear parameter estimation problem*, which has a nonuniform distribution of 2^n minima. It has non-isotropic attractor basins.

5.2 Performance Criteria

The performance criteria of niching methods include numerous possible attractive options and, generally speaking, vary in different experimental reports. The common ground is typically the examination of the population in its final stage, while aiming to observe the algorithm's ability to capture as well as *maintain* peaks of high-quality. Stability of good niches is thus considered as one of the implicit criteria.

Studies of traditional GA niching methods had been strongly interested in the distribution of the final population compared to a goal-distribution, as formalized by Mahfoud (1995a). While Mahfoud's formalism introduced a generic theoretical tool, being derived from information theory, other studies considered, *de facto*, specific performance calculations. For example, a very popular niching performance measurement, which satisfies Mahfoud formalism's criteria, is the *Chi-square-like performance statistic* (see, e.g., Deb and Goldberg 1989). The latter estimates the deviation of the actual distribution of individuals N_i from an ideal distribution (characterized by mean μ_i and variance σ_i^2) in all the $i = 1 \dots q + 1$ subspaces (q peak subspaces and the non-peak subspace):

$$\chi^2 = \sqrt{\sum_{i=1}^{q+1} \left(\frac{N_i - \mu_i}{\sigma_i} \right)^2} \quad (17)$$

where the ideal-distribution characteristic values are derived per function.

Most of the existing studies have focused on the ability to identify global as well as local optima, and to converge in these directions through time, with no particular interest in the

distribution of the population. One possible performance criterion is often defined as the *success-rate* of the niching process, which refers to the percentage of optima attained by the end of the run with respect to the target peaks, as defined *a priori*. Furthermore, as has been employed in earlier studies of GA niching (Miller and Shaw 1996), another performance criterion is called the *maximum peak ratio* (MPR) statistic. This metric measures the quality as well as the number of optima given as a final result by the evolutionary algorithm. Explicitly, assuming a *minimization problem*, given the fitness values of the subpopulations in the final population $\{\tilde{f}_i\}_{i=1}^q$, and the fitness values of the real optima of the objective function $\{\hat{\mathcal{F}}_i\}_{i=1}^q$, the *maximum peak ratio* is defined as follows:

$$\text{MPR} = \frac{\sum_{i=1}^q \hat{\mathcal{F}}_i}{\sum_{i=1}^q \tilde{f}_i} \quad (18)$$

where all values are assumed to be *strictly positive*. If this is not the case in the original parametrization of the landscape, the latter should be scaled accordingly with an additive constant for the sake of this calculation. Also, given a maximization problem, the MPR is defined as the sum of the obtained optima divided by the sum of the real optima. A drawback of this performance metric is that the real optima need to be known *a priori*. However, for many artificial test problems these can be derived analytically, or tight numerical approximations to them are available. We adopt the MPR performance criterion and recommend it for reporting experimental results of EA niching techniques.

5.2.1 Another Perspective: MPR Versus Time

Although the MPR metric was originally derived to be analyzed by means of its saturation value in order to examine the niches' stability, a new perspective was introduced by Shir and Bäck (2005a). That study investigated the MPR as a function of time, focusing on the early stages of the run, in addition to the saturation value. It was shown experimentally that the time-dependent MPR data fits a theoretical function, *the logistic curve*:

$$y(t) = \frac{a}{1 + \exp\{c(t - T)\}} \quad (19)$$

where a is the saturation value of the curve, T is its time shift, and c (in this context always negative) determines the shape of the exponential rise. This equation, known as the *logistic equation*, describes many processes in nature. All those processes share the same pattern of behavior: growth with *acceleration*, followed by *deceleration* and then a *saturation* phase. In the context of evolutionary niching methods, it was argued (Shir and Bäck 2005a) that the logistic parameters should be interpreted in the following way: T as the *learning period* of the algorithm, and the absolute value of c as its *niching formation acceleration*. a is clearly the MPR saturation value.

5.3 Experimental Observation Examples: Niching-CMA Revisited

We revisit the niching-CMA algorithm with three examples of experimental observations.

5.3.1 MPR Time-Dependent Analysis: Reported Observations

The MPR time-dependent analysis was applied in Shir and Bäck (2005a) to two ES-based niching techniques: niching with the CMA-ES and niching with the Standard-ES according to the Schwefel-approach (Shir and Bäck 2005b). In short, the latter method applies the same niching framework as the niching-CMA (● Sect. 4.2.1) except for one conceptual difference: It employs a (μ, λ) strategy in each niche, subject to *restricted mating*. Otherwise, it applies the standard ES operators (Bäck 1996).

We outline some of the conclusions of that study:

1. The **niching formation acceleration**, expressed as the absolute value of c , had larger values for the CMA-ES kernel for all the observed test cases. That implied stronger niching acceleration and faster convergence.
2. A trend concerning the absolute value of c as a function of the dimensionality was observed: The higher the dimensionality of the search space, the lower the absolute value of c , that is, the slower the niching process.
3. The **learning period**, expressed as the value of T in the curve fitting, has negative as well as positive values. Negative values mean that the niches' formation process, expressed as the exponential rise of the MPR, started immediately from generation zero.
4. The averaged **saturation value** a , that is, the MPR saturation value, was larger in all of the test cases for the CMA-ES mechanism. In that respect, the CMA kernel outperformed the standard-ES on the tested landscapes.

That study concluded with the claim that there was a clear *trade-off*: Either a long learning period followed by a high niching acceleration (CMA-ES), or a short learning period followed by a low niching acceleration (Standard-ES). A hypothesis concerning the existence of a general trade-off between the learning period T and the niching acceleration, c , was numerically assessed in a following study (Shir and Bäck 2008). It was shown that this trade-off stands for various DES niching variants on two synthetic landscapes (one separable, one non-separable) in a large spectrum of search space dimensions.

5.3.2 Performance on the Synthetic Multimodal Test-Suite

The proposed CMA-ES niching framework has been successfully applied to a suite of synthetic multimodal *high-dimensional* continuous landscapes, as reported by Shir and Bäck (2008), which in principle followed the recommended test functions proposed earlier in this chapter. That study addressed various *research questions* concerning the generalized DES niching framework, such as *which DES variant captures and maintains most desired optima* (i.e., best saturation MPR values), *what are the differences in the niching formation acceleration among the DES variants*, and others. The CMA-ES kernels were observed to perform very well among the DES variants, with the $(1+10)$ kernel performing best, while typically obtaining most of the desired basins of attraction of the various landscapes at different dimensions. Furthermore, upon carrying out behavioral analysis of the simulations, e.g., MPR analysis, some characteristic patterns for the different algorithmic kernels were revealed. For instance, it was observed that the elitist-CMA has consistently the lowest niching acceleration. A straightforward and rather intuitive explanation for the excellent behavior of the elitist-CMA variant would be its tendency to maintain convergence in any basin of attraction, versus a higher

probability for the comma strategy to escape them. Moreover, another argument for the advantage of an elitist strategy for niching was suggested. The niching problem can be considered as an optimization task with constraints, that is, the formation of niches that restricts competing niches and their optimization routines from exploring the search space freely. It has been suggested in previous studies (see, e.g., Kramer and Schwefel 2006) that ES self-adaptation in constrained problems would tend to fail with a comma strategy, and thus an elitist strategy is preferable for such problems.

We choose to omit here specific details concerning the numerical simulations, and refer the reader to Shir and Bäck (2008) and Shir (2008).

5.3.3 Real-World Application: Quantum Control

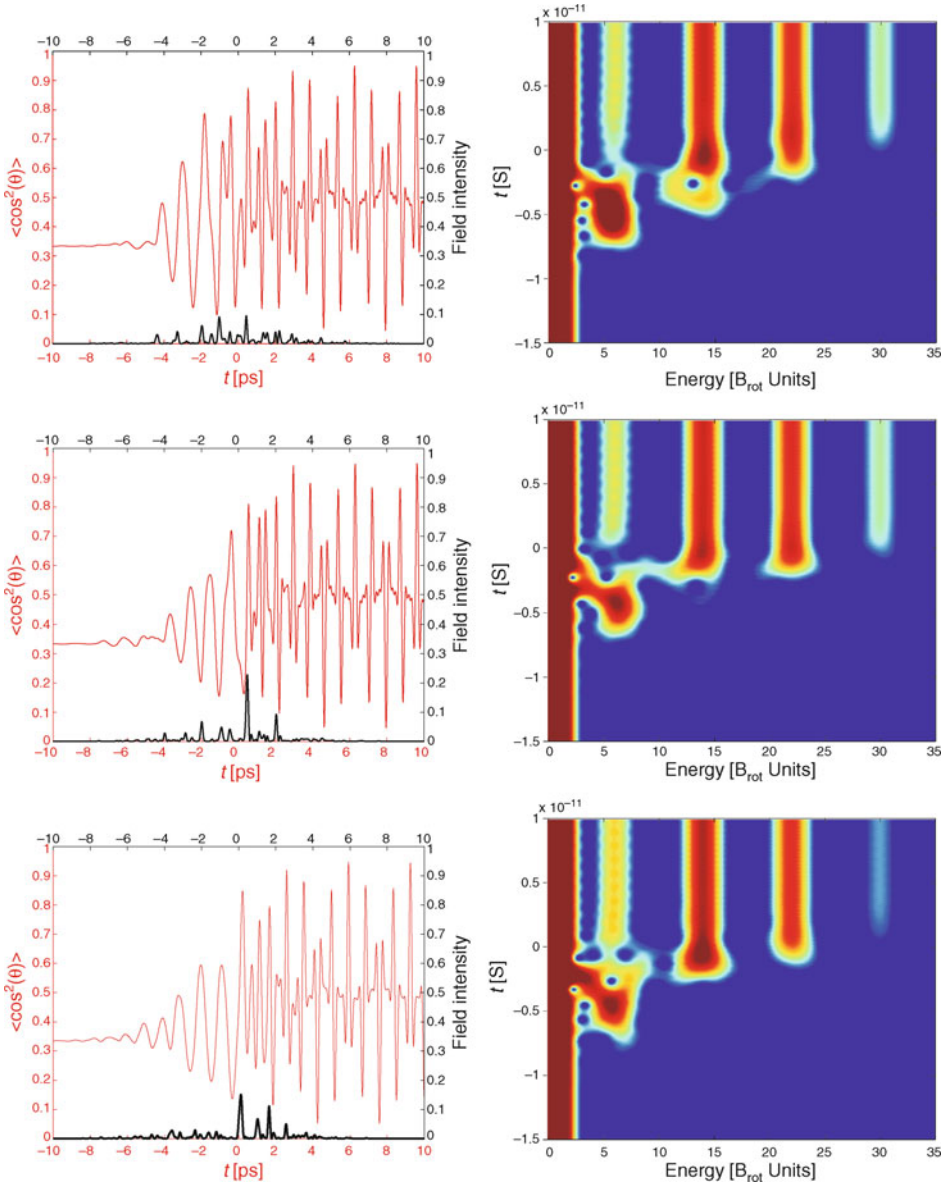
As was, furthermore, reported by Shir and Bäck (2008), the proposed DES niching framework was successfully applied to a real-world landscape from the field of Quantum Control. (Symbolically, this interdisciplinary study forms a *closed natural computing circle*, where biologically oriented investigation of organic evolution and speciation helps to develop methods for solving applications in physics in general, and in quantum control in particular. By our reckoning, this symbolism is even further strengthened upon considering the stochastic nature of evolutionary algorithms. This process can be thus considered as throwing dice in order to solve quantum mechanics, sometimes referred to as the *science of dice*.) Namely, the dynamic molecular alignment problem (for a review see Shir et al. 2008). In short, the goal of this application is the maximization of the alignment of diatomic molecules after the interaction with an electric field arising from a laser source. The black-box noise-free simulator, which is designed in a lab-oriented fashion, provides a reliable physics simulation with a duration of 35 s per trial solution. The 80 object variables constitute a parametrized curve, to be learned, that undergoes spectral transformation for the construction of the electric field. This challenging high-dimensional application required the definition of a tailor-made diversity measure, due to certain invariance properties of the control function that stem from the spectral transformations. The resulting metric employed was the Euclidean distance in the second derivative space of the control function. DES niching variants were shown to perform well, and to obtain different pulse shapes of high quality, representing different conceptual designs. Also in this case, the elitist-CMA kernel was observed to perform best. While referring the reader to Shir and Bäck (2008) and Shir (2008) for more details on this particular study, we would like to conclude with a visualization of the resulting niching process. Three laser-pulse niches, obtained in a typical elitist-CMA run, are plotted for illustration in ► Fig. 1 (left column): The thick line describes the attained pulse shape, while the thin line is the molecular reaction, also referred to as the *revival structure*. The right column provides the equivalent quantum pictures, in terms of population of the energy levels: The reader can observe three different modes of quantum energy-level population for the different attained niches.

6 The Niche-Radius Problem

While the motivation and usefulness of niching cast no doubt, the relaxation of assumptions and limitations concerning the hypothetical landscape is much needed if niching methods are

■ Fig. 1

Experimental results for niching-CMA with a fixed niche-radius on the real-world problem of dynamic molecular alignment (quantum control). *Left column: alignment and revival-structure of the three niches obtained by the (1+10)-CMA. Thin red line: alignment; thick black line: intensity of the laser pulse. Right column: Quantum picture of the solutions; a Fourier transform applied to the revival structures of the optimal solutions (the thin red alignment curves). The values are log scaled, and represent how high the rotational levels of the molecules are populated as a function of time. Note that the quality of the laser pulse cannot be measured in those plots.*



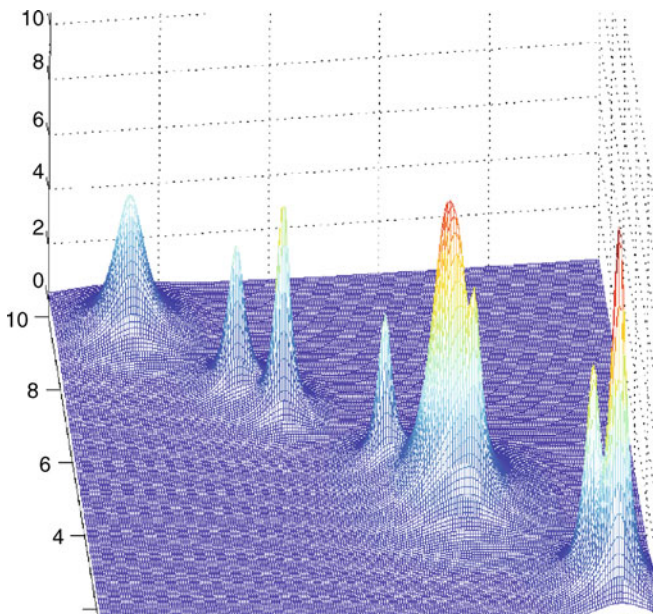
to be valid in a broader range of applications. In particular, consider the assumptions made in [Sect. 4.3](#) concerning the fitness landscape, stating that the optima are far enough from one another with respect to the *niche radius*, which is estimated for the given problem and remains fixed during the course of evolution. Most prominently, the niche radius is used in the *sharing function*, which penalizes fitness values of individuals whose distance to other individuals is below that threshold ([Sect. 4.1.1](#)). Obviously, there are landscapes for which this assumption is not applicable, and where this approach is most likely to fail (see [Figs. 2](#) and [3](#) for illustrations). This topic is directly linked to the task of defining a generic basin of attraction, which was discussed in [Sect. 2.3](#).

6.1 Treating the Niche Radius Problem: Existing Work

There were several GA-oriented studies which addressed this so-called *niche radius problem*, aiming to relax the assumption specified earlier, or even to drop it completely. Jelasity (1998) suggested a cooling-based mechanism for the niche-radius, also known as the UEGO, which adapts the global radius as a function of time during the course of evolution. Gan and Warwick (2001) introduced the so-called dynamic niche clustering, to overcome the radius problem by using a clustering mechanism. A complex subpopulation differentiation model, the so-called multinational evolutionary algorithm, was presented by Ursem (1999). It introduces a topology-based auxiliary mechanism of sampling, which detects whether feasible solutions share the same basin of attraction. A recent study by Stoean et al. (2007) considered

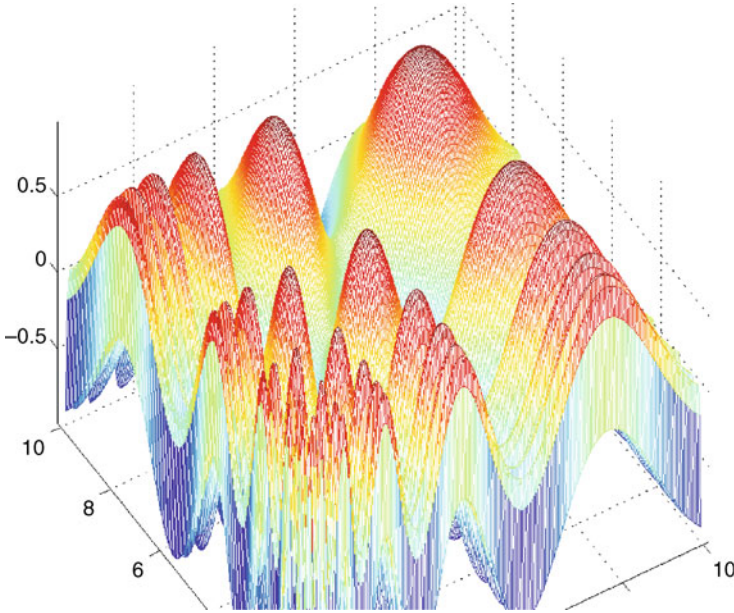
■ Fig. 2

The Shekel function (see, e.g., Törn and Zilinskas 1987) in a 2D decision space: introducing a dramatically uneven spread of optima.



■ Fig. 3

The Vincent function in a 2D decision space: a sine function with a decreasing frequency.



the hybridization of the latter with a radius-based niching method proposed in Stoean et al. (2005). An iterative statistical-based approach was introduced (Cioppa et al. 2004) for learning the optimal niche radius, without *a priori* knowledge of the landscape. It considers the *fitness sharing* strategy, and optimizes it as a function of the population size and the niche radius, without relaxing the landscape assumption specified earlier – that is, the niches are eventually obtained using a single fixed niche radius. Finally, two ES-based niche-radius adaptation niching algorithms were proposed recently (Shir and Bäck 2006; van der Goes 2008). They both considered individual niche radii for the evolving population members, based upon original learning schemes. While Shir and Bäck (2006) relied on coupling the niche radius to the niche’s global step size in combination with a secondary selection scheme, van der Goes et al. (2008) introduced a pure self-adaptive niche-radius approach, independent of any coupling to strategy parameters. The latter considered inner and outer niche counts, and applied the original sharing function for the sake of niches classification. Both approaches were shown to successfully tackle landscapes with challenging distributions as well as shapes of attraction basins.

6.2 Learning Niche Shapes: Exploiting CMA Information

As an extension to the previously discussed niching-CMA case study (➤ Sect. 4.2.1), a self-adaptive niche-shaping approach was derived in a recent study (Shir et al. 2010). The latter introduced the *Mahalanobis distance metric* into the niching mechanism, aiming to allow a

more accurate spatial classification of niches by considering rotatable ellipsoid shapes as classification regions. The shapes of these ellipsoids were obtained from the covariance matrix of the evolving multivariate normal distribution adapted by the CMA mutation scheme, replacing the default classification by means of Euclidean hyperspheres. This approach was implemented in a straightforward manner using the Mahalanobis metric as a replacement for the Euclidean, due to the fact that the former metric, the hyperspheres of which are ellipsoids when viewed in the Euclidean space, is parametrized by a covariance matrix that is obtained without additional cost from the CMA. The proposed approach was tested on high-dimensional artificial landscapes at several levels of difficulty, and was shown to be robust and to achieve satisfying results. ● [Figure 4](#) provides an illustration for the numerical results of this self-adaptive approach. It presents a snapshot gallery of the niching algorithm with the elitist-CMA kernel, employing the Mahalanobis distance, performing on the 2-dimensional Fletcher-Powell landscape.

7 Discussion and Outlook

Niching studies, following somehow various *mission statements*, introduce a large variety of approaches, some of which are more biologically inspired, whereas others are multimodal-optimization oriented. In both cases, those techniques were mainly tested on *low-dimensional synthetic landscapes*, and the application of these methods to real-world landscapes was hardly ever reported to date. We claim that niching methods should be implemented also for attaining multiple solutions in high-dimensional real-world problems, serving decision makers by providing them with the choice of optimal solutions, and representing well evolutionary algorithms in multimodal domains. By our reckoning, the *multimodal front* of real-world applications, that is, multimodal real-world problems, which pose the demand for multiple optimal solutions, should also enjoy the powerful capabilities of evolutionary algorithms, as other fronts do, for example, multi-objective or constrained domains.

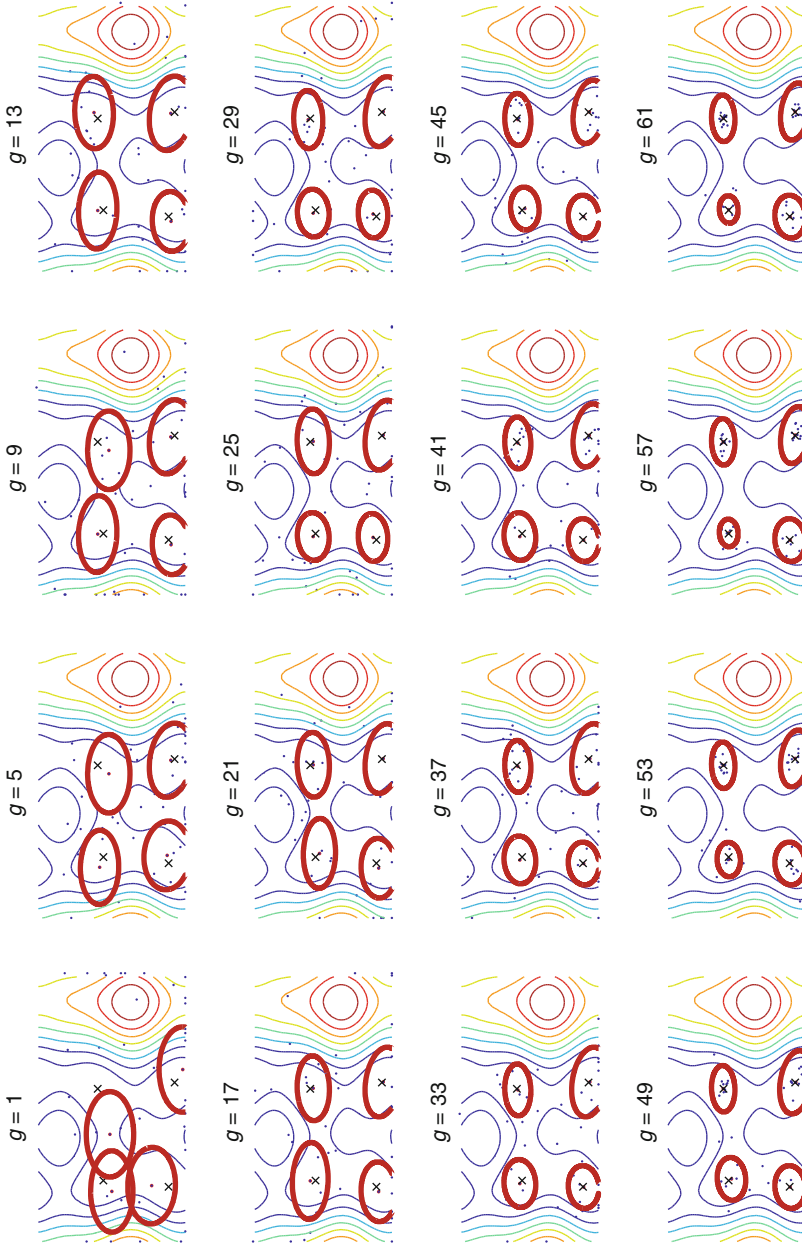
On a different note, Preuss, in an important paper (Preuss 2006), raised the question: “*Under what conditions can niching techniques be faster than iterated local search algorithms?*” Upon considering a simplified model, and assuming the existence of an efficient basin identification method, he managed to show that it does pay off to employ EA niching techniques on landscapes whose basins of attraction vary significantly in size. However, the original question in its general form remained open. Mahfoud (1995b) drew a comparison of *parallel* versus *sequential* niching methods, while considering *fitness sharing*, *deterministic crowding*, *sequential niching*, and *parallel hillclimbing*. Generally speaking, he concluded that parallel niching GAs outperform parallel hillclimbers on a hard set of problems, and that *sequential niching* is always outperformed by the parallel approaches.

Obviously, there is *no free lunch*, and there is no best technique, especially in the domain of multimodal search spaces. In this respect, *local search* capabilities should not be underestimated, and *population diversity preservers* should not be overestimated. We claim that like any other complex component in organic as well as artificial systems, the success of niching is about the subtle interplay between the different, sometime conflicting, driving effects.

In this chapter, we have introduced *niching* as an algorithmic framework following the evolutionary process of *speciation*. Upon providing the practical motivation for such a framework, in terms of conceptual designs for better decision making, we outlined in detail the essential biological background, as well as the computational perspective of multimodal function

■ Fig. 4

A snapshot gallery: the adaptation of the *classification-ellipses*, subject to the Mahalanobis metric with the updating covariance matrix, in the elitist-CMA strategy for a 2D Fletcher-Powell problem. Images are taken in the box $[-\pi, \pi]^2$. Contours of the landscape are given as the background, where the Xs indicate the real optima, the dots are the evolving individuals, and the ellipses are plotted centered about the peak individual. A snapshot is taken every four generations (i.e., every 160 function evaluations), as indicated by the counter.



optimization. This was followed by a survey of existing EA niching techniques, mainly from the GA field, and a detailed description of a specific case-study technique, based on the CMA-ES. We highlighted the *natural computing* aspects of these techniques, especially the biologically oriented components in each scheme, e.g., hosting capacity, gene flow, alpha-males, etc. We then proposed an experimental framework for testing niching methods, and presented some experimental observations of the CMA-ES case-study technique, including results from a real-world problem. We would like to use this opportunity to encourage other scholars to apply the proposed experimental framework, and at the same time to consider ways to join forces for the construction of a general framework to be agreed upon. Finally, we discussed the important topic of the niche-radius problem, gave an overview of existing methods to treat it, and revisited the CMA-ES case study for an extended self-adaptive scheme, which employs the Mahalanobis metric for obtaining niches with more complex geometrical shapes. We conclude that even though the assumptions made for radius-based niching techniques are problematic, there are good proposed solutions that treat the problem at different levels.

We would like to propose possible directions for future research in this domain:

- Transferring existing niching algorithms into additional real-world applications in general, and into experimental optimization in particular
- Proceeding with the effort to tackle the niche radius problem, in order to develop state-of-the-art niching techniques, which are not subject to the niche radius assumptions
- Extending the study of niching to environments with uncertainty
- Developing theoretical frameworks for the investigation of niching, for example, by means of simplified modeling

References

-
- Adamidis P (1994) Review of parallel genetic algorithms bibliography. Tech. rep., Automation and Robotics Lab., Dept. of Electrical and Computer Eng., Aristotle University of Thessaloniki, Greece
- Aichholzer O, Aurenhammer F, Brandstätter B, Ebner T, Krasser H, Magele C (2000) Niching evolution strategy with cluster algorithms. In: Proceedings of the 9th biennial IEEE conference on electromagnetic field computations. IEEE Press, New York
- Ando S, Sakuma J, Kobayashi S (2005) Adaptive isolation model using data clustering for multimodal function optimization. In: Proceedings of the 2005 conference on genetic and evolutionary computation, GECCO 2005. ACM, New York, pp 1417–1424
- Angus D (2006) Niching for population-based ant colony optimization. In: Second international conference on e-science and grid technologies (e-science 2006), December 4–6, 2006, Amsterdam, The Netherlands, IEEE Computer Society, p 115
- Auger A, Hansen N (2005a) A restart CMA evolution strategy with increasing population size. In: Proceedings of the 2005 congress on evolutionary computation CEC 2005. IEEE Press, Piscataway, NJ, pp 1769–1776
- Auger A, Hansen N (2005b) Performance evaluation of an advanced local search evolutionary algorithm. In: Proceedings of the 2005 congress on evolutionary computation CEC 2005. IEEE Press, Piscataway, NJ, pp 1777–1784
- Avigad G, Moshaiov A, Brauner N (2004) Concept-based interactive brainstorming in engineering design. J Adv Comput Intell Informatics 8(5): 454–459
- Avigad G, Moshaiov A, Brauner N (2005) Interactive concept-based search using MOEA: the hierarchical preferences case. Int J Comput Intell 2 (3):182–191
- Bäck T (1994) Selective pressure in evolutionary algorithms: a characterization of selection mechanisms. In: Michalewicz Z, Schaffer JD, Schwefel HP, Fogel DB, Kitano H (eds) Proceedings of the first IEEE conference on evolutionary computation (ICEC'94). Orlando FL. IEEE Press, Piscataway, NJ, pp 57–62
- Bäck T (1996) Evolutionary algorithms in theory and practice. Oxford University Press, New York

- Bartz-Beielstein T (2006) Experimental research in evolutionary computation – the new experimentalism. Natural computing series. Springer, Berlin
- Beasley D, Bull DR, Martin RR (1993) A sequential niche technique for multimodal function optimization. *Evolut Comput* 1(2):101–125
- Beyer HG (1999) On the dynamics of GAs without selection. In: Banzhaf W, Reeves C (eds) Foundations of genetic algorithms 5. Morgan Kaufmann, San Francisco, CA, pp 5–26
- Beyer HG, Schwefel HP (2002) Evolution strategies a comprehensive introduction. *Nat Comput Int J* 1(1):3–52
- Beyer HG, Brucherseifer E, Jakob W, Pohlheim H, Sendhoff B, To TB (2002) Evolutionary algorithms - terms and definitions. <http://ls11-www.cs.uni-dortmund.de/people/beyer/EA-glossary/>
- Blum C (2005) Ant colony optimization: introduction and recent trends. *Phys Life Rev* 2:353–373
- Bradshaw A (1965) Evolutionary significance of phenotypic plasticity in plants. *Adv Genet* 13:115–155
- Branke J (2001) Evolutionary optimization in dynamic environments. Kluwer, Norwell, MA
- Brits R, Engelbrecht AP, Bergh FVD (2002) A niching particle swarm optimizer. In: The fourth Asia-Pacific conference on simulated evolution and learning (SEAL2002). Singapore, pp 692–696
- Cavicchio D (1970) Adaptive search using simulated evolution. Ph.D. thesis, University of Michigan, Ann Arbor, MI
- Cioppa AD, Stefano CD, Marcelli A (2004) On the role of population size and niche radius in fitness sharing. *IEEE Trans Evolut Comput* 8(6):580–592
- Coello Coello CA, Lamont GB, Van Veldhuizen DA (2007) Evolutionary algorithms for solving multi-objective problems. Springer, Berlin
- Coello Coello CA (1999) A survey of constraint handling techniques used with evolutionary algorithms. Tech. Rep. Lania-RI-99-04, Laboratorio Nacional de Informática Avanzada. Xalapa, Veracruz, México
- Cristiano JJ, White CC, Liker JK (2001) Application of multiattribute decision analysis to quality function deployment for target setting. *IEEE Trans Syst Man Cybern Part C* 31(3):366–382
- Darwin CR (1999) The origin of species: by means of natural selection or the preservation of favoured races in the struggle for life. Bantam Classics, New York
- De Jong KA (1975) An analysis of the behavior of a class of genetic adaptive systems. Ph.D. thesis, University of Michigan, Ann Arbor, MI
- Deb K (2001) Multi-objective optimization using evolutionary algorithms. Wiley, New York
- Deb K, Goldberg DE (1989) An investigation of niche and species formation in genetic function optimization. In: Proceedings of the third international conference on genetic algorithms. Morgan Kaufmann, San Francisco, CA, pp 42–50
- Deb K, Spears WM (1997) Speciation methods. In: Bäck T, Fogel D, Michalewicz Z (eds) The handbook of evolutionary computation. IOP Publishing and Oxford University Press, Bristol
- Deb K, Tiwari S (2005) Omni-optimizer: a procedure for single and multi-objective optimization. In: Evolutionary multi-criterion optimization, third international conference, EMO 2005, Lecture notes in computer science, vol 3410. Springer, Guanajuato, Mexico, pp 47–61
- Dorigo M, Stützle T (2004) Ant colony optimization. MIT Press, Cambridge, MA
- Doye J, Leary R, Locatelli M, Schoen F (2004) Global optimization of Morse clusters by potential energy transformations. *INFORMS J Comput* 16(4):371–379
- Engelbrecht A (2005) Fundamentals of computational swarm intelligence. New York
- Fisher RA (1922) Darwinian evolution of mutations. *Eugen Rev* 14:31–34
- Fogel LJ (1966) Artificial intelligence through simulated evolution. Wiley, New York
- Freeman S, Herron JC (2003) Evolutionary analysis. Benjamin Cummings, 3rd edn. Redwood City, CA
- Gan J, Warwick K (2001) Dynamic niche clustering: a fuzzy variable radius niching technique for multimodal optimisation in GAs. In: Proceedings of the 2001 congress on evolutionary computation CEC2001, IEEE Press, COEX, World Trade Center, 159 Samseong-dong. Gangnam-gu, Seoul, Korea, pp 215–222
- van der Goes V, Shir OM, Bäck T (2008) Niche radius adaptation with asymmetric sharing. In: Parallel problem solving from nature – PPSN X, Lecture notes in computer science, vol 5199. Springer, pp 195–204
- Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Reading, MA
- Goldberg DE, Richardson J (1987) Genetic algorithms with sharing for multimodal function optimization. In: Proceedings of the second international conference on genetic algorithms and their application. Lawrence Erlbaum, Mahwah, NJ, pp 41–49
- Grosso PB (1985) Computer simulations of genetic adaptation: parallel subcomponent interaction in a multilocus model. Ph.D. thesis, University of Michigan, Ann Arbor, MI
- Hanagandi V, Nikolaou M (1998) A hybrid approach to global optimization using a clustering algorithm in a genetic search framework. *Comput Chem Eng* 22(12):1913–1925

- Hansen N, Ostermeier A (2001) Completely derandomized self-adaptation in evolution strategies. *Evolut Comput* 9(2):159–195
- Hansen N, Ostermeier A, Gawelczyk A (1995) On the adaptation of arbitrary normal mutation distributions in evolution strategies: the generating set adaptation. In: Proceedings of the sixth international conference on genetic algorithms (ICGA6). Morgan Kaufmann, San Francisco, CA, pp 57–64
- Haykin S (1999) *Neural networks: a comprehensive foundation*, 2nd edn. Prentice Hall, NJ, USA
- Holland JH (1975) *Adaptation in natural and artificial systems*. The University of Michigan Press, Ann Arbor, MI
- Igel C, Sutton T, Hansen N (2006) A computational efficient covariance matrix update and a (1+1)-CMA for evolution strategies. In: Proceedings of the genetic and evolutionary computation conference, GECCO 2006. ACM, New York, pp 453–460
- Jelasy M (1998) UEGO, an abstract niching technique for global optimization. In: Parallel problem solving from nature - PPSN V, Lecture notes in computer science, vol 1498. Springer, Amsterdam, pp 378–387
- Kennedy J, Eberhart R (2001) *Swarm intelligence*. Morgan Kaufmann, San Francisco, CA
- Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge
- Kramer O, Schwefel HP (2006) On three new approaches to handle constraints within evolution strategies. *Nat Comput Int J* 5(4):363–385
- Li R, Eggermont J, Shir OM, Emmerich M, Bäck T, Dijkstra J, Reiber J (2008) Mixed-integer evolution strategies with dynamic niching. In: Parallel problem solving from nature - PPSN X, Lecture notes in computer science, vol 5199. Springer, pp 246–255
- Lunacek M, Whitley D (2006) The dispersion metric and the CMA evolution strategy. In: Proceedings of the genetic and evolutionary computation conference, GECCO 2006. ACM, New York, pp 477–484
- Lunacek M, Whitley D, Sutton A (2008) The impact of global structure on search. In: Parallel problem solving from nature - PPSN X, Lecture notes in computer science, vol 5199. Springer, pp 498–507
- Mahfoud SW (1995a) Niching methods for genetic algorithms. Ph.D. thesis, University of Illinois at Urbana Champaign, IL
- Mahfoud SW (1995b) A comparison of parallel and sequential niching methods. In: Eshelman L (ed) Proceedings of the sixth international conference on genetic algorithms. Morgan Kaufmann, San Francisco, CA, pp 136–143
- Martin W, Lienig J, Cohoon J (1997) Island (migration) models: evolutionary algorithms based on punctuated equilibria. In: Bäck T, Fogel DB, Michalewicz Z (eds) *Handbook of evolutionary computation*. Oxford University Press, New York, and Institute of Physics, Bristol, pp C6.3:1–16
- McPherson BA, Smith DC, Berlocher SH (1988) Genetic differences between host races of *Rhagoletis pomonella*. *Nature* 336:64–66
- Miller B, Shaw M (1996) Genetic algorithms with dynamic niche sharing for multimodal function optimization. In: Proceedings of the 1996 IEEE international conference on evolutionary computation (ICEC'96). New York, pp 786–791
- Ostermeier A, Gawelczyk A, Hansen N (1993) A derandomized approach to self adaptation of evolution strategies. Tech. rep., TU Berlin
- Ostermeier A, Gawelczyk A, Hansen N (1994) Step-size adaptation based on non-local use of selection information. In: Parallel problem solving from nature - PPSN III, Lecture notes in computer science, vol 866. Springer, Berlin, pp 189–198
- Parrott D, Li X (2006) Locating and tracking multiple dynamic optima by a particle swarm model using speciation. *Evolut Comput, IEEE Trans* 10(4): 440–458, doi: 10.1109/TEVC.2005.859468
- Petrowski A (1996) A clearing procedure as a niching method for genetic algorithms. In: Proceedings of the 1996 IEEE international conference on evolutionary computation (ICEC'96). New York, pp 798–803
- Preuss M (2006) Niching prospects. In: Proceedings of the international conference on bioinspired optimization methods and their applications, BIOMA 2006. Jožef Stefan Institute, Slovenia, pp 25–34
- Preuss M (2007) Reporting on experiments in evolutionary computation. Tech. Rep. CI-221/07, University of Dortmund, SFB 531
- Ramalhinho-Lourenco H, Martin OC, Stützle T (2000) Iterated local search. *Economics Working Papers* 513, Department of Economics and Business, Universitat Pompeu Fabra
- Scheiner SM, Goodnight CJ (1984) The comparison of phenotypic plasticity and genetic variation in populations of the grass *Danthonia spicata*. *Evolution* 38(4):845–855
- Schönemann L, Emmerich M, Preuss M (2004) On the extinction of sub-populations on multimodal landscapes. In: Proceedings of the international conference on bioinspired optimization methods and their applications, BIOMA 2004. Jožef Stefan Institute, Slovenia, pp 31–40
- Shir OM (2008) Niching in derandomized evolution strategies and its applications in quantum control. Ph.D. thesis, Leiden University, The Netherlands
- Shir OM, Bäck T (2005a) Dynamic niching in evolution strategies with covariance matrix adaptation. In: Proceedings of the 2005 congress on evolutionary computation CEC-2005. IEEE Press, Piscataway, NJ, pp 2584–2591

- Shir OM, Bäck T (2005b) Niching in evolution strategies. Tech. Rep. TR-2005-01, LIACS, Leiden University
- Shir OM, Bäck T (2006) Niche radius adaptation in the CMA-ES niching algorithm. In: Parallel problem solving from nature - PPSN IX, Lecture notes in computer science, vol 4193. Springer, pp 142–151
- Shir OM, Bäck T (2008) Niching with derandomized evolution strategies in artificial and real-world landscapes. *Nat Comput Int J* (2008), doi: 10.1007/s11047-007-9065-5
- Shir OM, Beltrani V, Bäck T, Rabitz H, Vrakking MJ (2008) On the diversity of multiple optimal controls for quantum systems. *J Phys B At Mol Opt Phys* 41(7):(2008). doi: 10.1088/0953-4075/41/7/074021
- Shir OM, Emmerich M, Bäck T (2010) Adaptive niche-radii and niche-shapes approaches for niching with the CMA-ES. *Evolut Comput* 18(1):97–126. doi: 10.1162/evco.2010.18.1.18104
- Singh G, Deb K (2006) Comparison of multi-modal optimization algorithms based on evolutionary algorithms. In: Proceedings of the 2006 annual conference on genetic and evolutionary computation, GECCO 2006. ACM Press, New York, pp 1305–1312
- Smith RE, Bonacina C (2003) Mating restriction and niching pressure: results from agents and implications for general EC. In: Proceedings of the 2003 conference on genetic and evolutionary computation, GECCO 2003, Lecture notes on computer science, vol 2724. Springer, Chicago, IL, pp 1382–1393
- Spears WM (1994) Simple subpopulation schemes. In: Proceedings of the 3rd annual conference on evolutionary programming, World Scientific. San Diego, CA, Singapore, pp 296–307
- Stoean C, Preuss M, Gorunescu R, Dumitrescu D (2005) Elitist generational genetic chromodynamics – a new radii-based evolutionary algorithm for multimodal optimization. In: Proceedings of the 2005 congress on evolutionary computation (CEC'05). IEEE Press, Piscataway NJ, pp 1839–1846
- Stoean C, Preuss M, Stoean R, Dumitrescu D (2007) Disburdening the species conservation evolutionary algorithm of arguing with radii. In: Proceedings of the genetic and evolutionary computation conference, GECCO 2007. ACM Press, New York, pp 1420–1427
- Streichert F, Stein G, Ulmer H, Zell A (2003) A clustering based niching EA for multimodal search spaces. In: Proceedings of the international conference evolution artificielle, Lecture notes in computer science, vol 2936. Springer, Heidelberg, Berlin, pp 293–304
- Suganthan PN, Hansen N, Liang JJ, Deb K, Chen YP, Auger A, Tiwari S (2005) Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization. Tech. rep., Nanyang Technological University, Singapore
- Törn A, Zilinskas A (1987) Global optimization, Lecture notes in computer science, vol 350. Springer, Berlin
- Tsui K (1992) An overview of Taguchi method and newly developed statistical methods for robust design. *IIE Trans* 24:44–57
- Ursem RK (1999) Multinational evolutionary algorithms. In: Proceedings of the 1999 congress on evolutionary computation (CEC 1999). IEEE Press, Piscataway NJ, pp 1633–1640
- Whitley D, Mathias KE, Rana SB, Dzuberka J (1996) Evaluating evolutionary algorithms. *Artif Intell* 85(1–2):245–276
- Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97–159
- Yin X, Germany N (1993) A fast genetic algorithm with sharing using cluster analysis methods in multimodal function optimization. In: Proceedings of the international conference on artificial neural nets and genetic algorithms, Innsbruck. Austria, 1993, Springer, pp 450–457

