The genes are the master programmers, and they are programming for their lives. They are judged according to the success of their programs in coping with all the hazards that life throws at their survival machines, and the judge is the ruthless judge of the court of survival.

The Selfish Gene; Richard Dawkins

Chapter 2

Introduction to Niching

2.1 Speciation Theory vs. Conceptual Designs

Evolutionary Algorithms have the tendency to lose diversity within their population of feasible solutions and to converge into a single solution [1, 36, 37], even if the search landscape has multiple globally optimal solutions.

Niching methods, the extension of EAs to finding multiple optima in multi-modal 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 basins of attraction in a multi-modal 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 two-fold - maintaining the diversity within a population-based stochastic algorithm from the computational perspective, but also having 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 the decision maker with multiple solutions which ideally represent different conceptual designs, for single-criterion or multi-criterion search spaces [38, 39]. The concept of "going optimal" is often extended now into the aim for "going multi-optimal", so to speak: Obtaining optimal results but also providing the decision maker with different choices. On this particular note, it is worth mentioning the so-called Second Toyota Paradox [40]:

"Delaying decisions, communicating ambiguously, and pursuing an excessive number of prototypes, can produce better cars faster and cheaper."

Niching methods have been studied in the past 35 years, mostly in the context of Genetic Algorithms, and the focus has been mainly on the theoretical aspect. As will be discussed here, niching methods have been mostly
a by-product of studying population diversity, and were hardly ever at the
front of the EC research.

This chapter, the gateway to niching, discusses a variety of introductory
topics - ranging from biological aspects of diversity and speciation, mathemati-
cal definitions of basins of attraction, to GA niching methods - which
reflect the strong interdisciplinary nature of this subject.

2.2 From DNA to Organic Diversity

In this section we introduce the biological elementary concepts that corre-
spond to the core of niching methods: population diversity. This section is
mainly based on [41] and personal lecture notes\footnote{Notes were taken in
the course "Evolutionary Biology" of Prof. David Stern (EEB309),
Princeton University, Fall 2007}.

A Preliminary 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 indi-
viduals 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.

2.2.1 Genetic Drift

Organic evolution can be broken down into four defining fundamental mecha-
nisms: 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 gen-
etics, and thus did not discuss this effect in his "Origin of Species" [42]. In
short, genetic drift is a stochastic process in which the diversity is lost in
finite populations. A distribution of genetic properties is transferred to the
next 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,
e.g., fixation of specific alleles when subject to equal fitness. This is why
genetic drift is often considered as a neutral effect. The smaller the popu-
lation, the faster and stronger this effect occurs. An analogy is occasionally
drawn between genetic drift to Brownian motion of particles in mechanics.
In order to demonstrate the genetic drift effect, we conducted simulations\(^2\) on the following basic model of population genetics: The evolution of random-mating populations with *two alleles*, namely, \(A\) and \(a\), equal fitnesses of the *three genotypes* (i.e., no preferences for \(AA\), \(Aa\), nor \(aa\)), no mutations, no migration between the replicate populations, and finite population size \(N\). We simulated ten simultaneously evolving populations, for three test-cases of population sizes: \(N_1 = 10\), \(N_2 = 100\), and \(N_3 = 1000\). Figure 2.1 offers an illustration for the three different simulations. It is easy to observe a clear trend in this simple experiment: Alleles’ loss/fixation is very likely to occur in small population sizes, and is not likely to occur in large population sizes.

The *genetic drift* effect had been originally recognized by R.A. Fisher [43] (referred to as *random survival*), and was explicitly mentioned by S. Wright when studying Mendelian populations [44]. It was, however, revisited and given a new interpretation in the *Neutral Theory of Molecular Evolution* of Kimura [45]. The *Neutral Theory* suggested that the *random genetic drift* effect is the main driving force within molecular evolution, rather than the *non-random 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.

### 2.2.2 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 review shortly both of them.

**Variations within a Species** Diversity of organisms within a single species stems from variance at the genotypic level, referred to as *genetic diversity*, or from the existence of 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

---

\(^2\)Simulations were conducted with the *PopG Genetic Simulation Program*, version 3.1.
Figure 2.1: Ten simultaneously evolving populations, for three test-cases of population sizes: $N_1 = 10$ [TOP], $N_2 = 100$ [CENTER], and $N_3 = 1000$ [BOTTOM]. The vertical axis corresponds to the allele frequency of A in the population, as a function of generations, indicated on the horizontal axis.
as quantitative variation. Roughly speaking, the main sources of quantitative variations [41, 46] are outlined here:

1. Genes have multiple loci, and hence are mapped into 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.

3. Phenotypic plasticity is the amount in which the genotypic expression vary in different environments\(^3\), and it is a direct source of variation at the phenotypic level.

4. The plastic response of the genotype to the environment, i.e., 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 [44].

Speciation 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, and 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, e.g., migration or mountain building. It results in two geographically isolated populations.

\(^{3}\)Bradshaw [47] 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."
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 [48].

These four modes of speciation correspond to four levels of geographically decreasing linkages. Roughly 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).

### 2.3 "Ecological Optima": Basins of Attraction

We devote this section to the definition of basins of attraction. This section is mainly based on Törn and Zilinskas [8].

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\(^4\) [8].

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

\[
\frac{d\bar{x}(t)}{dt} = -\nabla f(\bar{x}(t))
\]  

(2.1)

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

\[
Y = \left\{ \bar{x} \in \mathbb{R}^n \left| \bar{x}(0) = \bar{x} \land \bar{x}(t)_{t \geq 0} \text{ satisfies Eq. 2.1} \land \lim_{t \to \infty} \bar{x}(t) \text{ exists} \right. \right\}
\]

(2.2)

\(^4\) Intuitively, and strictly metaphorically speaking, we may think of a *region of attraction* of \(\bar{x}_L\) as the region, where if water is poured, it will reach \(\bar{x}_L\). Accordingly, we may then think of the basin of \(\bar{x}_L\) as the maximal region that will be covered when the cavity at \(\bar{x}_L\) is filled to the lowest part of its rim.
2.4. Population Diversity within Evolutionary Algorithms

Definition 2.3.1. The region of attraction $A(x_L)$ of a local minimum $x_L$ is

$$A(x_L) = \left\{ \vec{x} \in \Upsilon \mid \vec{x}(0) = \vec{x} \wedge \| \vec{x}(t) \|_{t \geq 0} \right.$$ satisfies Eq. 2.1 $\wedge$ $\lim_{t \to \infty} \vec{x}(t) = \vec{x}_L \right\}.$$  \hspace{1cm} (2.3)

The basin of $x_L$ is the maximal level set that is fully contained in $A(x_L)$.

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 non-overlapping connected sets.

2.3.1 Classification of Optima: The Practical Perspective

On the note of the theoretical definition of the basin, it is worth mentioning the practical perspective for the classification of optima shapes, also referred to as global topology. This topic is strongly related to the emerging subfield of robustness study (see, e.g., [49]), which aims at attaining high-yield optima 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 Lunaček and Whitley for classifying different classes of multimodal landscapes with respect to algorithmic performance [50]. 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\(^5\). This classification to low dispersion versus high dispersion may be associated with the algorithmic trade-off between exploration of the landscape and exploitation of local structures. In the broad context of this work, it is interesting to note that the CMA was shown in [50] to perform well on low-dispersion landscapes, and was less efficient on high-dispersion landscapes.

2.4 Population Diversity within EAs

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 and speciation - the two different concepts from organic evolution that were discussed in Section 2.2 - at the same time. This is simply due to the fact that the differences between the two concepts do not have any practical effect on the evolutionary search and 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. Among EC

\(^5\)We deliberately avoid the definition of a funnel, as its definition is rather vague. We refer the reader to [51].
researchers, population diversity is first considered as a component due to play a role in a fruitful 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 [37], 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 given goal-distribution.

Let \( P = \{p_i\}_{i=1}^\ell \) be a discrete distribution describing the current partitioning of the population into subpopulations, i.e., \( p_i \) is the portion of the population located at the \( i^{th} \) site. Let \( Q = \{q_i\}_{i=1}^\ell \) be the goal-distribution of the population with respect to the defined sites. We demand that by construction we have \( \sum_{i=1}^\ell p_i = 1 \), as well as \( \sum_{i=1}^\ell q_i = 1 \). 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. We review some of them here.

1. The entropy of a system is a quantitative measurement of its disorder or randomness [52]. Although it had originated in Physics, in the Second Law of Thermodynamics, it also became an important criterion in information systems, also referred to as Shannon’s Information Entropy. Accordingly, this general concept has several definitions, where we choose here to introduce a relevant definition to probability distributions.

**Definition 2.4.1.** The entropy of a discrete probability distribution, \( \{p_i\}_{i=1}^\ell \), is defined as:

\[
S(P) = \sum_{i=1}^\ell p_i \cdot \ln \left( \frac{1}{p_i} \right) = - \sum_{i=1}^\ell p_i \cdot \ln (p_i)
\]

(2.4)

The following measure, developed by Kullback and Leibler [53], quantifies the directed divergence between the two distributions, \( P \) and \( Q \), as long as it is well defined (i.e., \( \forall i \ p_i > 0, \ q_i > 0 \)):

\[
D(P, Q) = \sum_{i=1}^\ell p_i \cdot \ln \left( \frac{p_i}{q_i} \right)
\]

(2.5)

Given a uniform goal-distribution, the Kullback-Leibler measure is re-
2.4. Population Diversity within Evolutionary Algorithms

duced to the following:

\[ D(P, U) = \sum_{i=1}^{\ell} p_i \cdot \ln \left( \frac{p_i}{1/\ell} \right) = \ln(\ell) - S(P) \quad (2.6) \]

Mahfoud shows that the Kullback-Leibler measure satisfies the criteria of his formalism, and can be used as a diversity measure.

2. The standard distance metrics are useful measures of directed divergence between the distributions.

**Definition 2.4.2.** A family of distance metrics is defined as follows:

\[ D(P, Q) = \sqrt{\sum_{i=1}^{\ell} |p_i - q_i|^k}, \quad 0 < k \leq \infty \quad (2.7) \]

Mahfoud shows that the family of distance metrics, with \( 0 < k \leq \infty \), satisfies the criteria and can be used as diversity measures.

This analytical framework, with its derived measurements of diversity, allowed Mahfoud to compare the role of population diversity among different GA niching techniques, and essentially became a performance criterion in his study.

**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. As will be argued here, population diversity is an important component in a population-based search, and it even becomes critical in extended techniques, such as *Evolutionary Multi-Objective* approaches (see Chapter 5).

Next, we describe the effect of *diversity loss* within Evolution Strategies. This will be followed by some conclusions drawn by the GA research concerning diversity loss within GAs, as a point of reference to ES.

**2.4.1 Diversity Loss in 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. In practice, especially in the context of derandomized ES, it is not an essential component.

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, respectively.

**Selective Pressure: Fast Take-Over**

Evolution Strategies have a strictly deterministic, rank-based approach, to selection. In the two traditional approaches, \((\mu, \lambda)\) 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 *takeover time* \([54]\), which gives a quantitative description of selective pressure with respect to the selection operator exclusively:

**Definition 2.4.3.** The *takeover time* \(\tau^*\) 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 \([36]\), who analyzed all the ES selection mechanisms also with respect to takeover times. Here, we introduce the results for the takeover times of the main selection mechanisms in the absence of mutation, where we chose to omit the derivations. See \([1]\) for the proofs.

**Theorem 2.4.4.** The takeover time of \((\mu, \lambda)\)-selection is:

\[
\tau^*_{(\mu, \lambda)} = \frac{\ln(\lambda)}{\ln\left(\frac{\mu}{\lambda}\right)}
\]  

\[(2.8)\]

**Theorem 2.4.5.** The takeover time of \((\mu + \lambda)\)-selection is given implicitly by:

\[
\lambda = \frac{\alpha_1^{\tau^*+1} - \alpha_2^{\tau^*+1}}{\sqrt{\frac{\lambda}{\mu} \left(\frac{\lambda}{\mu} + 4\right)}}
\]

\[
\alpha_{1,2} = \frac{\lambda}{2\mu} \pm \frac{1}{2} \sqrt{\left(\frac{\lambda}{\mu} \left(\frac{\lambda}{\mu} + 4\right)\right)}
\]

\[(2.9)\]
2.4. Population Diversity within Evolutionary Algorithms

**Corollary 2.4.6.** It is easy to verify that upon the substitution of the traditional population sizes of the standard-ES, one obtains very short takeover times for the given selection mechanisms, which imply high selective pressure.

The ratio $\frac{\lambda}{\mu}$ clearly plays a dominant role in the derived takeover times of the two selection approaches. Not surprisingly, the term *selective pressure* is occasionally associated with this ratio. It should be noted that the same ratio also governs the convergence velocity of the $(\mu + \lambda)$-ES for large population sizes, i.e., $\mu \gg 1$ (see [1] pp. 89-90).

**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 directly responsible to the loss of population diversity in ES.

**Recombination Drift** Beyer explored extensively the so-called *mutation-induced speciation by recombination* (MISR) principle (see, e.g., [55]). 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 towards fitter regions of the landscape. Furthermore, Beyer managed to prove analytically [55] 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. That study provides us with an analytical result that a sampling-based recombination is subject to genetic drift, and leads to loss of population diversity.

**Selection Drift** At the same time, a recent study on the extinction of subpopulations on a simple *bimodal equi-fitness* model investigated the drift effect of the selection operator [56]. 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 $\mu$. The analysis was conducted by means of Markov chain models, supported by statistical simulations.
Corollary 2.4.7. 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 conclude that an Evolution Strategy with a small population is likely to encounter a rapid effect of diversity loss.

2.4.2 Point of Reference: Diversity Loss within GAs

Mahfoud devoted a large part of his thesis to studying population diversity within GAs [37]. 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 1st moment of the selection operator. It has been demonstrated by Mahfoud [37] that the selection pressure, or equivalently the non-zero 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 2nd moment of the selection operator, or its variance. Mahfoud [37] 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. We consider this effect as a genetic drift in its broad definition - sampling error of a distribution - although it was not explicitly referred to as such by Mahfoud.

- **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 located previously might be lost. In that sense, they 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
2.4. Population Diversity within Evolutionary Algorithms

Effect. In the GA field, where the crossover operator is in use (single-point, two-point or n-point crossovers), it 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, e.g., [22]).

It should be noted that an equivalent ES disruptive-recombination effect was analyzed in [57], and was shown to boost the extinction of subpopulations located around a basin of attraction. Furthermore, it was observed that by omitting the recombination operator the stability of the subpopulations was indeed strengthened.

2.4.3 Neutrality in ES Variations: Mutation Drift

The mutation operator, the defining operator of Evolution Strategies, applies normally-distributed variations of finite sample sizes, and thus is expected to experience sampling errors as the sample sizes decrease. These sampling errors lead to an undirected movement of the population center of mass, with speed which depends on the population size. We shall call this effect mutation drift.

Simulations. In order to demonstrate and analyze this mutation drift effect, we conducted simulations on the following basic ES model: The parallel evolution of several populations in an n-dimensional space, based on sequential normally-distributed variations (with a fixed identity matrix as the covariance of the distribution), without selection nor recombination. The ES variation can be then considered as a continuous random walk of μ individuals in an n-dimensional space. Essentially, this corresponds to mutation-only ES of multiple populations.

We simulated 10 simultaneously evolving populations, for three test-cases of population sizes: μ1 = 10, μ2 = 100, and μ3 = 1000, subject to three space dimensions: n1 = 1, n2 = 10, and n3 = 1000. For each simulation, we measured the distance of the population mean, or center of mass, to the starting point, as a function of generational steps. More precisely, we measured the location of the population mean for n1, and the Euclidean distance from the origin for {n2, n3}. Figure 2.2 presents the outcome of these calculations. It is easy to observe in those simulations a similar trend to the equivalent simulations of Section 2.2.1: The center of mass strongly drifts away from the origin when the population is small, and shows the contrary behavior when the population is large. We therefore conclude that mutation drift is very likely to occur in small population sizes, and is not likely to occur in large population sizes.
Chapter 2. Introduction to Niching

Figure 2.2: Illustration of the mutation drift effect in ES, for 10 simultaneously evolving populations, as a function of population size [$\mu_1 = 10$ (left), $\mu_2 = 100$ (center), and $\mu_3 = 1000$ (right)] and landscape dimensionality [$n_1 = 1$ (top), $n_2 = 10$ (center), and $n_3 = 1000$ (bottom)]. The vertical axes correspond to the location of the center of mass of the population (for $n_1 = 1$, top row) or distance from the origin to the center of mass of the population (for $n_2 = 10$ or $n_3 = 1000$, in the center or bottom rows, respectively). The horizontal axis corresponds to the generational step of the calculation.

We thus demonstrated here that the center of mass of a small ES population is subject to a so-called mutation drift. This is an equivalent effect to the genetic drift of alleles, as described in Section 2.2.1. We claim that it allows for easy translation of small populations from one location to another, having the potential to boost fast and efficient speciation. Therefore, we argue that drift in this context can be a blessing for the fast formation of species in niching.

Since small populations are typically employed by Evolution Strategies, and especially by the derandomized variants, we consider this effect of mutation drift as a positive potential component for niching with ES. This result
provides us with further motivation to introduce DES with small populations into the niching framework.

### 2.5 Classical 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 mission statement for niching techniques. There have been several attempts to provide a proper definition and functional specification for niching; we review some of them here:

1. Mahfoud [37] 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. [58] put forward also the actual maintenance of population diversity:

   **Naching**: process of separation of individuals according to their states in the search space or maintenance of diversity by appropriate techniques, e.g. local population models, fitness sharing, or distributed EA.

3. Preuss [59] 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.

**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 overview of the main known methods, with emphasis on the important concepts of *Sharing* and *Crowding*. This survey is mainly based on [37] and [60].
2.5.1 Fitness Sharing

The sharing concept was one of the pioneering niching approaches. It was first introduced by Holland in 1975 [4], and later implemented as a niching technique by Goldberg and Richardson [61]. 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 foundations 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 the similarity metric of the population, which can be genotypic or phenotypic, the sharing function is defined as follows:

\[ sh(d_{i,j}) = \begin{cases} 
1 - \left( \frac{d_{i,j}}{\rho} \right)^{\alpha_{sh}} & \text{if } d_{i,j} < \rho \\
0 & \text{otherwise} 
\end{cases} \] (2.10)

where \( d_{i,j} \) is the distance between individuals \( i \) and \( j \), \( \rho \) (traditionally noted as \( \sigma_{sh} \)) 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

\[ m_i = \sum_{j=1}^{N} sh(d_{i,j}) \] (2.11)

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

\[ f_{i}^{sh} = \frac{f_i}{m_i} \] (2.12)

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 roulette wheel selection (RWS) operator [22]; The latter takes into consideration the shared fitness. Thus, the sharing mechanism practically punishes 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, \( \rho \). 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. This poses the so-called niche radius problem, to be discussed later, especially in Chapters 3 and 4.

It is important to note that the formulas for determining the value of \( \rho \), which will be given in Chapter 3, are dependent on \( q \), the number of peaks of the target function. Hence, a second assumption is that \( q \) can be estimated.
2.5. Classical Niching Techniques

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 $\rho$ rather complicated. This provides us with the motivation to treat the issue of niche shapes in Chapter 4.

In the literature, several GA niching sharing-based techniques, which implement and extend the basic concept of sharing, can be found [37, 61, 62, 63, 64, 65, 66]. Furthermore, the concept of sharing was successfully extended to other "yields of interest", such as concept sharing [38].

2.5.2 Dynamic Fitness Sharing

In order to improve the sharing mechanism, a dynamic approach was proposed. The dynamic niche sharing method [64], 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:

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

(2.13)

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. 2.11.

The shared fitness is then defined as follows:

$$f_i^{dyn} = \frac{f_i}{m_i^{dyn}}$$

(2.14)

The identification of the dynamic niches can be carried out by means of a greedy approach, as proposed in [64] as the Dynamic Peak Identification (DPI) algorithm (see Algorithm 4). 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.

2.5.3 Clearing

Another variation to the fitness sharing technique, called clearing, was introduced by Petrowski [65] at the same time as the dynamic fitness sharing [64]. 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
Algorithm 4 Dynamic Peak Identification (DPI)

**Input:** population Pop, number of niches q, niche radius ρ
1: Sort Pop in decreasing fitness order
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

individuals of each niche which 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. 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 previous techniques, the peaks are never fixed, 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 [65].

### 2.5.4 Crowding

Crowding was one of the pioneering methods in this field, as introduced by de Jong in 1975 [67]. The crowding approach aimed at reducing changes in the population distribution between generations, in order to prevent premature convergence; it does so by applying 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
### 2.5. Classical Niching Techniques

**Algorithm 5** Deterministic Crowding: Replacement Selection

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

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, will be revisited in the context of Evolutionary Multi-Objective Optimization in Chapter 5; in the context of niching see also Deb’s "Omni-Optimizer" ([68] and Section 5.2.1).

Mahfoud, who analyzed the crowding niching technique [37], 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 niching technique. The proposed procedure applies variation operators to pairs of individuals in order to generate their offspring, who are then all evaluated with respect to the crowding distance, and undergo replacement selection (see Algorithm 5, which assumes maximization).

#### 2.5.5 Clustering

The application of clustering for niching is very intuitive from the computational perspective, as well as straightforward in its implementation. Yin et al. [62] proposed a clustering framework for niching with GAs, which we describe here briefly. A clustering algorithm, such as the *K-Means* algorithm [69], 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 individuals $i$, its fitness is then defined as:

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

(2.15)

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

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., [56, 70, 71, 72, 74, 75]).

2.5.6 The Sequential Niche Technique

The straightforward approach of iteration can be used to locate sequentially multiple peaks in the landscape, by means of an iterative local search [76]. This procedure is blind to any information gathered in previous searches, and sequentially restarts stochastic searches, 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), i.e., 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:

$$R \approx \gamma + \ln (q), \quad (2.16)$$

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 equally likely to be found.

On a related note, we would like to mention a multi-restart with increasing population size approach that was developed with the CMA algorithm [77]. 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 sub-optimal minima during its search.

Beasley et al. extended the naive iteration approach, and developed the so-called Sequential Niche technique [78]. 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.
2.5. Classical Niching Techniques

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 secondly 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 subpopulations formation. In that sense, it has been claimed that it could not be considered as a niching method, but rather as a modified iterated search.

2.5.7 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., [79, 80, 81]) 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, e.g., complete net topology, ring topology, or neighborhood topology.

2.5.8 Other GA-Based Methods

Tagging (see, e.g., [82, 83]) 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, but it also 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 [82] to be a rather efficient implementation of the sharing technique.

A complex subpopulation differentiation model, the so-called Multinational Evolutionary Algorithm, was presented in [84]. This original technique considers a world of "nations", "governments", and "politicians", with dynamics dictated by migration of individuals, merging of sub-populations, 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. [85] 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 dynamic of the genetic operations.

Chapter 4 will elaborate furthermore on specific GA-based niching techniques in the context of the so-called niche radius problem.

2.5.9 Miscellaneous: Mating Schemes

It has been observed that once the niche formation process starts, i.e., when the population converges into the multiple basins in the landscape, cross-breeding 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 Section 2.2.2.

Deb and Goldberg [54] 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 used 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 [37] 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 [86], 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 in maintaining diversity and converging with stability to 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.

2.6 Niching in Evolution Strategies

Researchers in the field of Evolution Strategies initially showed no particular interest in the field of niching, leaving it essentially for Genetic Algorithms. An exception would be the employment of island models. Roughly speaking, classical niching mechanisms 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 experience problems in a straightforward implementation. Manipulations of fitness values are usually not suitable for Evolution Strategies, as in the case of constraints handling: death-penalty is typically the preferred approach for constraints violation in ES, rather than a continuous punishment as used in other EAs, in order to avoid the introduction of disruptive effects to the self-adaptation mechanism (see, e.g., [34, 87]). 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 as well that a different treatment is required here.

There are several, relatively new, niching methods that have been proposed within ES, mostly clustering-based [56, 73, 74]. A different approach, which preceded this thesis, was presented in [88, 89, 90].

2.7 Discussion and Mission Statement

Niching techniques, 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 usually tested on low-dimensional artificial landscapes, and the application of these methods to real-world landscapes was hardly ever reported. We claim that niching methods should be implemented also for attaining multiple solutions in high-dimensional real-world problems, serving the decision makers by providing them with the choice of optimal solutions, and representing well Evolutionary Algorithms in multimodal domains. By our humble reckoning, the multimodal front of real-world applications, i.e. multimodal real-world problems which demand multiple optimal solutions, should also enjoy the powerful capabilities of Evolution Strategies, as other fronts do, e.g., multi-objective domains and constrained domains.

On a different note, Preuss, in an important paper [59], raised the question: “Under what conditions can niching techniques be faster than iterated local search algorithms?”. Considering a simplified model, and assuming the
existence of an efficient basin identification method, he managed to show
that it pays off to employ Evolutionary Algorithms niching techniques on
landscapes whose basins of attraction vary significantly in size. However,
the original question in its general form remained open.

Mahfoud [91] drew a comparison of parallel versus sequential niching
methods, while considering fitness sharing, deterministic crowding, sequen-
tial niching, and parallel hillclimbing. Generally speaking, he concluded that
parallel niching GAs outperform parallel hillclimbers on a hard set of prob-
lems, and that sequential niching is always outperformed by the parallel
approaches.

Obviously, there is no free lunch, and there is no best technique, espe-
cially in niching. In this context, local search capabilities should not be
underestimated, and population diversity preservers should not be overesti-
mated. 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.

We thus choose to adopt Preuss' mission statement, and define the chal-
lenge in niching as follows:

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