Nat Comput (2009) 8:171–196
DOI 10.1007/s11047-007-9065-5

Niching with derandomized evolution strategies in artificial and real-world landscapes

Ofer M. Shir · Thomas Bäck

Published online: 13 February 2008

Abstract We introduce a framework of *derandomized* evolution strategies (ES) niching techniques. A survey of these techniques, based on 5 variants of *derandomized* ES, is presented, based on the fixed niche radius approach. The core mechanisms range from the very first *derandomized approach to self-adaptation of ES* to the sophisticated $(1 \dot{+} \lambda)$ Covariance Matrix Adaptation (CMA). They are applied to artificial as well as real-world multimodal continuous landscapes, of different levels of difficulty and various dimensions, and compared with the *maximum-peak-ratio* (MPR) performance analysis tool. While characterizing the performance of the different derandomized variants in the context of niching, some conclusions concerning the niching formation process of the different mechanisms are drawn, and the hypothesis of a trade-off between learning time and niching acceleration is numerically confirmed. Niching with $(1 + \lambda)$-CMA core mechanism is shown to experimentally outperform all the other variants, especially on the real-world problem. Some theoretical arguments supporting the advantage of a plus-strategy for niching are discussed. For the real-world application in hand, taken from the field of Quantum Control, we show that the niching framework can overcome some degeneracy in the search space, and obtain different conceptual designs using problem-specific diversity measurements.

Keywords Niching · Derandomized evolution strategies · CMA · Quantum control · Laser pulse shaping

O. M. Shir (✉) · T. Bäck
Natural Computing Group, Leiden University, Niels Bohrweg 1, Leiden, CA 2333, The Netherlands
e-mail: oshir@liacs.nl
URL: natcomp.liacs.nl

T. Bäck
NuTech Solutions, Martin-Schmeisser-Weg 15, 44227 Dortmund, Germany
e-mail: baeck@liacs.nl

1 Introduction

Evolutionary Algorithms (EAs), popular population-based stochastic search-methods, have the tendency to lose diversity within their population of feasible solutions and to converge into a single solution (Mahfoud 1995; Bäck 1994, 1996). *Niching methods*, the extension of EAs to multi-modal optimization, address this issue by maintaining the diversity of certain properties within the population—and this way they allow parallel convergence into multiple good solutions in multimodal domains. 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 the diversity within a population-based stochastic algorithm from the computational perspective, but also having an insight into *speciation theory* from the biological perspective. The practical aspect provides a real-world motivation for this problem—there is an increasing interest of the applications' community in providing the decision maker with multiple solutions with different conceptual designs, for single-criterion or multi-criteria search spaces (see, e.g., Avigad et al. 2004).

Niching techniques are often subject to criticism due to the so-called *niche radius problem*, as will be explained. The majority of the niching methods hold an assumption concerning the fitness landscape, stating that the optima are far enough from one another with respect to some threshold distance, called the *niche radius*, which is estimated for the given problem and remains fixed during the course of evolution. Obviously, there are landscapes for which this assumption is not applicable, and where this approach is most likely to fail. Generally speaking, 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 were only few attempts to tackle it theoretically (Törn and Zilinskas 1987). De facto, the niche-radius problem has been addressed at several directions, and a recent study offered a successful self-adaptive approach for an individual niche-radius (Shir and Bäck 2006).

Evolution Strategies (ES) (Beyer and Schwefel 2002) are a *canonical EA for continuous function optimization*, due to their straightforward encoding, their specific variation operators, the self-adaptation of their mutation distribution as well as to their high performance in this domain in comparison to other methods on benchmark problems. Even for large dimensions, an ES is a suitable method, and was shown to outperform other competing methods (Bäck 1996). However, the standard ES approaches are exposed to several disruptive effects, especially concerning the individual mutative step-size control. The family of derandomized evolution strategies (Hansen and Ostermeier 2001) offers an improved mutative ES mechanism, and its variants are considered to be the state of the art strategies.

This paper presents a survey of ES niching techniques, based on 5 variants of *derandomized* evolution strategies, applied to a set of continuous multimodal artificial test functions of different levels of difficulty, as well as to a real-world landscape from the field of *Quantum Control* (Warren et al. 1993; Rabitz et al. 2000; Weinacht and Bucksbaum 2002; Nuernberger et al. 2007). Quantum control theoretical problems typically have rich landscapes with an infinite number of optimal solutions, corresponding to perfect control (Demiralp and Rabitz 1993), and thus are a suitable framework for testing niching techniques.

As far as we know this is the first comparison of these ES variants, in particular in the context of niching. The performance of the algorithms is evaluated based on the so-called MPR analysis tool (Shir and Bäck 2005a), which allows to characterize to some degree the learning behavior, the niching formation process and the saturation profile of the different mechanisms. Niching with $(1 + \lambda)$-CMA core mechanism will be shown to experimentally outperform all the other variants. The numerical results are consistent, and support our

experimental conclusions. Some theoretical arguments supporting the advantage of a plus-strategy for niching are discussed.

The remainder of the paper is organized as follows. Section 2 presents the various evolutionary core mechanisms in use for this survey—the so-called family of derandomized evolution strategies. In Sect. 3 we introduce the ES niching framework, and its MPR performance analysis. This is followed in Sect. 4 by the description of the experimental procedure—for both artificial and real-world landscapes—as well as an analysis of the numerical results. In Sect. 5 we draw conclusions, summarize our study, and propose future directions in the domain of our research.

2 The family of derandomized evolution strategies

In standard evolution strategies, mutative step-size control tends to work well for the adaptation of a global step-size, but tends to fail when it comes to individual step-sizes or arbitrary normal mutations. This is due to several disruptive effects (Hansen and Ostermeier 2001) as well as to the fact that the selection of the *strategy parameters* setting is indirect, i.e. the vector of a successful mutation is not used to adapt the step-size parameters, but rather the parameters of the distribution that led to this mutation vector. The so-called *derandomized mutative step-size control* aims to treat those disruptive effects.

The first versions of *derandomized ES algorithms* introduced a controlled global step-size in order to monitor the individual step-sizes by decreasing the stochastic effects of a probabilistic sampling. The selection disturbance was completely removed with later versions by omitting the adaptation of strategy parameters by means of probabilistic sampling. This was combined with individual information from the last generation (the successful mutations, i.e. of selected offspring), and then adjusted to *correlated mutations*. Later on, the concept of *adaptation by accumulated information* was introduced, aiming to use wisely the past information for the purpose of step-size adaptation: instead of using the information from the last generation only, it was successfully generalized to a weighted average of the previous generations.

Note that the different variants of *derandomized-ES* hold different numbers of strategy parameters to be adapted, and this is a factor in the learning speed of the optimization routine. The different algorithms employ a number of strategy parameters that scale *linearly* (individual step-sizes) versus *quadratically* (arbitrary normal mutations, by means of a full covariance matrix) with the search space dimensionality, corresponding to first- or second-order information of the search landscape. As expected, there is typically a trade-off between the number of strategy parameters and the time needed for the learning of the step-sizes.

We hereby present briefly different derandomized-ES algorithms that are used in our niching framework.

DR1. The first derandomized attempt (Ostermeier et al. 1993) coupled the successful mutations to the selection of decision parameters, and learned the mutation step-size directly from the *difference vectors* between parents and selected offspring. The offspring generation, or the mutation step, is formulated for $k = 1 \ldots \lambda$:

$$\boldsymbol{x}^{(g+1)} = \boldsymbol{x}^{(g)} + \xi_k \delta^{(g)} \boldsymbol{\xi}^k_{scal} \boldsymbol{\delta}^{(g)}_{scal} \boldsymbol{z}_k \boldsymbol{z}_k \in \{-1, +1\}^n \tag{1}$$

The evaluation and selection are followed by the adaptation of the strategy parameters (subscripts *sel* refer to the selected individual):

$$\delta^{(g+1)} = \delta^{(g)} \cdot \left(\xi^{sel}\right)^{\beta} \tag{2}$$

$$\boldsymbol{\delta}_{scal}^{(g+1)} = \boldsymbol{\delta}_{scal}^{(g)} \cdot \left(\boldsymbol{\xi}_{scal}^{sel} + b\right)^{\beta_{scal}} \tag{3}$$

where $\boldsymbol{\xi}_{scal}^{k} \sim \mathcal{N}(0,1)^{+}$ (i.e., the positive part of the normal distribution), and $\mathcal{P}\left(\xi^{k} = \frac{7}{5}\right) = \mathcal{P}\left(\xi^{k} = \frac{5}{7}\right) = \frac{1}{2}$; $\beta = \sqrt{1/n}$, $\beta_{scal} = 1/n$, $b = 0.35$, and $\xi^{k} \in \left\{\frac{7}{5}, \frac{5}{7}\right\}$ are constants.

DR2. The second derandomized ES variant (Ostermeier et al. 1994) aimed to accumulate information about the correlation or anti-correlation of past mutation vectors in order to adapt the *global step-size* as well as the *individual step-sizes*—by introducing a quasi-memory vector. This accumulated information allowed omitting the stochastic element in the adaptation of the strategy parameters—mutating them only by means of successful mutations, rather than with random steps.

The mutation step for $k = 1...\lambda$ reads:

$$\boldsymbol{x}^{(g+1)} = \boldsymbol{x}^{(g)} + \delta^{(g)}\boldsymbol{\delta}_{scal}^{(g)}\boldsymbol{z}_k \quad \boldsymbol{z}_k \sim \mathcal{N}(0,1) \tag{4}$$

Introducing a quasi-memory vector $\boldsymbol{Z}$

$$\boldsymbol{Z}^{(g)} = c\boldsymbol{z}_{sel} + (1-c)\boldsymbol{Z}^{(g-1)} \tag{5}$$

The adaptation of the strategy parameters according to the selected offspring:

$$\delta^{(g+1)} = \delta^{(g)} \cdot \left(exp\left(\frac{\|\boldsymbol{Z}^{(g)}\|}{\sqrt{n}\sqrt{\frac{c}{2-c}}} - 1 + \frac{1}{5n}\right)\right)^{\beta} \tag{6}$$

$$\boldsymbol{\delta}_{scal}^{(g+1)} = \boldsymbol{\delta}_{scal}^{(g)} \cdot \left(\frac{\left|\boldsymbol{Z}^{(g)}\right|}{\sqrt{\frac{c}{2-c}}} + b\right)^{\beta_{scal}}, \quad \left|\boldsymbol{Z}^{(g)}\right| = \left(|Z_1^{(g)}|, |Z_2^{(g)}|, ..., |Z_n^{(g)}|\right) \tag{7}$$

with $\beta = \sqrt{1/n}$, $\beta_{scal} = 1/n$, $b = 0.35$, and the quasi-memory rate $c = \sqrt{1/n}$ as constants.

DR3. This third variant (Hansen et al. 1995), usually referred to as the *Generation Set Adaptation* (GSA), considered the derandomization of arbitrary normal mutations for the first time, aiming to achieve invariance with respect to the scaling of variables and the rotation of the coordinate system. This naturally came with the cost of a quasi-memory matrix, $\mathbf{B} \in \mathbb{R}^{m \times n}$, setting the dimension of the strategy parameters space to $n^2 \leq m \leq 2n^2$. The adaptation of the global step-size is *mutative* with stochastic variations, just like in the *DR1*.

The mutation step for $k = 1...\lambda$ now reads:

$$\boldsymbol{x}^{(g+1)} = \boldsymbol{x}^{(g)} + \delta^{(g)}\xi_k\boldsymbol{y}_k \quad \boldsymbol{z}_k \sim \mathcal{N}(0,1) \tag{8}$$

$$\boldsymbol{y}_k = c_m\mathbf{B}^{(g)} \cdot \boldsymbol{z}_k \tag{9}$$

The update of the *memory matrix* is formulated as:

$$\begin{gathered} \mathbf{B}^{(g)} = \left(\boldsymbol{b}_1^{(g)}, \ldots, \boldsymbol{b}_m^{(g)}\right) \\ \boldsymbol{b}_1^{(g+1)} = (1-c) \cdot \boldsymbol{b}_1^{(g)} + c \cdot (c_u\xi_{sel}\boldsymbol{y}_{sel}), \quad \boldsymbol{b}_{i+1}^{(g+1)} = \boldsymbol{b}_i^{(g)} \end{gathered} \tag{10}$$

The step-size is updated as follows:

$$\delta^{(g+1)} = \delta^{(g)}(\xi_{sel})^{\beta} \tag{11}$$

where $\mathcal{P}(\xi_k = \frac{3}{2}) = \mathcal{P}(\xi_k = \frac{2}{3}) = \frac{1}{2}$; $\beta = \sqrt{1/n}$,$c_m = (1/\sqrt{m})(1 + 1/m)$,$c = \sqrt{1/n}$, $\xi_k \in \{\frac{3}{2}, \frac{2}{3}\}$ and $c_u = \sqrt{(2-c)/c}$ are constants.

$(1, \lambda)$ - CMA - ES. We consider the covariance matrix adaptation evolution strategy (CMA-ES) (Hansen and Ostermeier 2001) *(rank-one update with cumulation)*. This advanced method applies *principal component analysis* (PCA) to the *selected* mutations during the evolution, also referred to as *"the evolution path"*, for the adaptation of the covariance matrix of the distribution.

$\boldsymbol{p}_c^{(g)} \in \mathbb{R}^n$ is the so-called *evolution path*, the crucial component for the adaptation of the covariance matrix, and $\boldsymbol{p}_\sigma^{(g)} \in \mathbb{R}^n$ is the *conjugate evolution path*, which is responsible for the step-size control. $\mathbf{C}^{(g)} \in \mathbb{R}^{n \times n}$, is the covariance matrix $\left(\mathbf{C}^{(g)} = \mathbf{B}^{(g)}\mathbf{D}^{(g)}\left(\mathbf{B}^{(g)}\mathbf{D}^{(g)}\right)^T\right)$.

The mutation step for $k = 1...\lambda$ is then defined:

$$\boldsymbol{x}_k^{(g+1)} = \boldsymbol{x}^{(g)} + \sigma^{(g)}\mathbf{B}^{(g)}\mathbf{D}^{(g)}\boldsymbol{z}_k^{(g+1)} \tag{12}$$

with $\boldsymbol{z}_k \sim \mathcal{N}(\boldsymbol{0}, \mathbf{I})$.

The *evolution path*, initialized $\boldsymbol{p}_c^{(0)} = \boldsymbol{0}$, is explicitly updated as follows:

$$\boldsymbol{p}_c^{(g+1)} = (1 - c_c) \cdot \boldsymbol{p}_c^{(g)} + H_\sigma^{(g+1)}\sqrt{c_c(2 - c_c)} \cdot \mathbf{B}^{(g)}\mathbf{D}^{(g)}\boldsymbol{z}_{sel}^{(g+1)} \tag{13}$$

$$H_\sigma^{(g+1)} = \begin{cases} 1 & \text{if } \frac{\left\|\boldsymbol{p}_\sigma^{(g+1)}\right\|}{\sqrt{1-(1-c_\sigma)^2}} < H_{thresh} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where $H_{thresh} = \left(1.5 + \frac{1}{n-0.5}\cdot\right)E[\|\mathcal{N}(0, \mathbf{I})\|]$.

The covariance matrix, initialized as identity $\mathbf{C}^{(0)} = \mathbf{I}$, is adapted accordingly:

$$\mathbf{C}^{(g+1)} = (1 - c_{cov}) \cdot \mathbf{C}^{(g)} + c_{cov} \cdot \boldsymbol{p}_c^{(g+1)}\left(\boldsymbol{p}_c^{(g+1)}\right)^T \tag{15}$$

The calculation of the "conjugate" evolution path, initialized $\boldsymbol{p}_\sigma^{(0)} = \boldsymbol{0}$, reads:

$$\boldsymbol{p}_\sigma^{(g+1)} = (1 - c_\sigma) \cdot \boldsymbol{p}_\sigma^{(g)} + \sqrt{c_\sigma(2 - c_\sigma)} \cdot \mathbf{B}^{(g)}\boldsymbol{z}_{sel}^{(g+1)} \tag{16}$$

and then followed by the update of the global step-size:

$$\sigma^{(g+1)} = \sigma^{(g)} \cdot \exp\left(\frac{c_\sigma}{d_\sigma} \cdot \left(\frac{\left\|\boldsymbol{p}_\sigma^{(g+1)}\right\|}{E[\|\mathcal{N}(0, \mathbf{I})\|]} - 1\right)\right) \tag{17}$$

The various learning coefficients are typically set as $c_c = 4/(n + 4)$, $c_{cov} = 2/(n + 1.4)^2$, $c_\sigma = 3/(n + 4)$, and $d_\sigma = 1 + c_\sigma$. The expectation of the length of a normally distributed random vector is approximated by:

$$E[\|\mathcal{N}(0, \mathbf{I})\|] \approx \sqrt{n}\left(1 - \frac{1}{4n} + \frac{1}{21n^2}\right)$$

$(1 + \lambda)$ - CMA - ES This elitist version (Igel et al. 2006) of the CMA-ES algorithm, which had been originally derived for the sake of a *multi-criteria* CMA algorithm (Igel et al. 2007), combined the classical concept of the (1 + 1) ES strategy, and in particular the *success probability* and *success rule* components (see, e.g., Bäck 1996), with the covariance matrix adaptation concept. The so-called *success rule based step size control* replaces the *path length control* of the CMA-comma strategy. The same notation as in the comma version is used here:

$$\boldsymbol{x}_k^{(g+1)} = \boldsymbol{x}^{(g)} + \sigma^{(g)}\mathbf{B}^{(g)}\mathbf{D}^{(g)}\boldsymbol{z}_k^{(g+1)}, \quad k = 1, \ldots, \lambda \tag{18}$$

After the evaluation of the new generation, the success rate is updated $p_{succ} = \lambda_{succ}^{(g+1)}/\lambda$, where:

$$\bar{p}_{succ} = (1 - c_p) \cdot \bar{p}_{succ} + c_p \cdot p_{succ} \tag{19}$$

$$\sigma^{(g+1)} = \sigma^{(g)} \cdot \exp\left(\frac{1}{d} \cdot \left(\bar{p}_{succ} - \frac{p_{succ}^{target}}{1 - p_{succ}^{target}}(1 - \bar{p}_{succ})\right)\right) \tag{20}$$

The covariance matrix is updated only if the selected offspring is better than the parent. Then,

$$\boldsymbol{p}_c = \begin{cases} (1 - c_c)\boldsymbol{p}_c + \sqrt{c_c(2 - c_c)} \cdot \frac{\boldsymbol{x}_{sel}^{(g+1)} - \boldsymbol{x}^{(g)}}{\sigma_{parent}^{(g)}} & \text{if } \bar{p}_{succ} < p_\Theta \\ (1 - c_c)\boldsymbol{p}_c & \text{otherwise} \end{cases} \tag{21}$$

$$\mathbf{C}^{(g+1)} = \begin{cases} (1 - c_{cov}) \cdot \mathbf{C}^{(g)} + c_{cov} \cdot \boldsymbol{p}_c\boldsymbol{p}_c^T & \text{if } \bar{p}_{succ} < p_\Theta \\ (1 - c_{cov}) \cdot \mathbf{C}^{(g)} + c_{cov} \cdot \left(\boldsymbol{p}_c\boldsymbol{p}_c^T + c_c(2 - c_c)\mathbf{C}^{(g)}\right) & \text{otherwise} \end{cases} \tag{22}$$

The default parameters are set as follows: $d = 1 + \frac{n}{2}$, $p_{succ}^{target} = \frac{2}{11}$, $c_p = \frac{1}{12}$, $c_c = \frac{2}{n+2}$, $c_{cov} = \frac{2}{n^2+6}$, and $p_\Theta = 0.44$.

3 ES dynamic niching

The advent of modern evolution strategies allows successful global optimization with minimal requirements concerning exogenous parameters, mostly without recombination, and with a low number of function evaluations. In particular, consider the $(1 \overset{+}{,} \lambda)$ derandomized ES variants presented in the previous section. In the context of niching, this generation of modern ES variants allows the construction of fairly simple and elegant niching algorithms. Note that this study is limited to niching with a fixed niche radius, assuming that the landscapes under investigation would not dramatically suffer from the so-called niche-radius assumptions. Also, we choose to apply derandomized evolution strategies *without recombination* as the core mechanisms of the framework.

We provide here the reader with some details concerning our niching framework, but first begin with the motivation for ES niching.

3.1 The motivation: ES diversity

The promotion of *diversity* in the traditional GA had been originally the main motivation for the development of niching methods, as was deeply investigated by Mahfoud (1995). In this section we give a brief review of ES diversity, with respect to the tools given by Mahfoud, and by that provide the motivation for niching within ES.

We consider three main effects which cause the standard ES to lose diversity: *selective pressure, operator disruption* and *random genetic drift*.

Selective pressure. The standard evolution strategy (Bäck 1996) has a strictly deterministic, rank-based approach, to selection. In the two traditional strategies, (μ,λ) and $(\mu + \lambda)$, an approach of deterministically selecting the best individuals (out of the appropriate set—the next generation or the union of the two generations, respectively) is

applied, which intuitively implies 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. It should be noted that the term selective pressure is occasionally associated with the ratio λ/μ.

Furthermore, Goldberg and Deb introduced the important concept of *takeover time* (Deb and Goldberg 1989), which gives a quantitative description of selective pressure with respect only to the selection operator:

Definition 1 *The takeover 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 takeover times. Let us introduce the results for the takeover times of the main selection mechanisms, where we chose to omit the derivations.

Theorem The takeover time of (μ, λ)-selection is:

$$\tau^*_{(\mu,\lambda)} = \frac{\ln(\lambda)}{\ln\left(\frac{\lambda}{\mu}\right)} \tag{23}$$

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

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

where

$$\alpha_{1,2} = \frac{\lambda}{2\mu} \pm \frac{1}{2} \cdot \sqrt{\left(\frac{\lambda}{\mu}\left(\frac{\lambda}{\mu} + 4\right)\right)} \tag{24}$$

It is easy to verify that with the substitution of the traditional values of the standard ES, we get very short takeover times for the given selection mechanisms, which imply high *selective pressure*.

Operator disruption. In the standard ES the mutation operator typically has a small effect, which means "staying in the neighborhood". In that sense, the mutation operator can be regarded in the standard ES as an operator with negligible disruption. The recombination operator, on the other hand, has a bigger effect. In the standard ES, where *discrete* and *intermediate recombination operators* are in use (Bäck 1996), the disruptive nature is also highly intuitive—modifying a coordinate of the decision parameters to be optimized, not in a local manner (averaging or taking a value from a different individual), has the potential to shift the offspring not in a negligible way. The outcome would be losing good candidate solutions, while boosting the takeover of other highly-fit individuals.

Random genetic drift. *Genetic drift* is a stochastic process in which the diversity is lost in finite populations (Kimura 1983). A distribution of genetic properties is transferred to the next generation in a limited manner, due to the finite number of offspring. As a result, the distribution tends to approach an equilibrium distribution. In small populations this

process can occur fast and become significant. Since small population sizes are used in the standard ES, the effect of *random genetic drift* occurs and causes the loss of diversity within the population. In multimodal functions, it was shown that the effect of *genetic drift* in ES causes a convergence to an equilibrium distribution around a single attractor (Schönemann et al. 2004).

ES diversity: conclusions. The standard ES is exposed to several strong effects which interrupt the formation and maintenance of multiple solutions and push the evolution process toward a rapid convergence into a single solution.

3.2 The niching routine

We consider a niching technique with individual search points, which independently and simultaneously perform a derandomized $(1, \lambda)$ or $(1 + \lambda)$ search in different locations of the space. The *speciation interaction* occurs every generation when all the offspring are considered together to become the niches' representatives for the next iteration, or simply the next search points, based on the rank of their fitness and their spatial location with respect to higher-ranked individuals.

Explicitly, 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 dynamic variables of the derandomized algorithms which uniquely define the search at a given point of time. Such dynamic variables are the current search point, the mutation vector / 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 $(q + 1)^{th}...(q + p)^{th}$ D-sets are individuals which are randomly re-generated in every cycle of generations (denoted as κ) as potential candidates for niche formation. This is basically a *quasi-restart* mechanism, which allows new niches to form dynamically. We stress that the total number of function evaluations allocated for a run should depend on the number of desired peaks, q, and not on p. Setting the value of p should reflect the dilemma between applying a wide restart approach for exploring further the search space and 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 problem increases.

Alogorithm 1 Dynamic Peak Identification

input: Pop, q, ρ

Sort *Pop* in decreasing fitness order
$i := 1$
$NumPeaks := 0$
$DPS := \emptyset$
while $NumPeaks \neq q$ and $i \leq popSize$ **do**
  if $Pop[i]$ is not within ρ of peak in DPS **then**
    $DPS := DPS \cup \{Pop[i]\}$
    $NumPeaks := NumPeaks + 1$
  end if
  $i := i + 1$
end while

output: DPS

Until stopping criteria are 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 done using the DPI routine (Miller and Shaw 1996) (see Algorithm 1)—based on the fixed niche radius ρ—and the peaks then become the new search points. Their D-sets are inherited from their parents and updated respectively.

A pseudo-code for the *niching routine* is presented as Algorithm 2.

3.3 The niche radius

The original formula for ρ for *phenotypic sharing* in GAs was derived by Deb and Goldberg (Deb and Goldberg 1989). By following the trivial analogy and considering the decision parameters as the decoded parameter space of the GA, the same formula can be applied, using the Euclidean metric. Given q, the number of 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.

Algorithm 2 Dynamic ES Niching: A Single Generation

```
for i = 1...(q + p) search points do
    Generate λ samples based on the D-set of i
end for
Evaluate fitness of the population
Compute the Dynamic Peak Set with Algo. 1
for all elements of DPS do
        Set peak as a search point
        Inherit the D-set and update it respectively
end for
if N_DPS = size of DPS < q then
    Generate q − N_dps new search points, reset D-sets
end if
if gen mod κ ≡ 0 then
        Reset the (q + 1)^th...(q + p)^th search points
end if
```

The volume of the hypersphere which contains the entire space is

$$V = cr^n \tag{25}$$

where c is a constant, given explicitly by:

$$c = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}, \quad \Gamma(n) = \int_0^\infty x^{n-1}\exp(-x)dx \tag{26}$$

Given lower and upper boundary values $x_{k,min}$, $x_{k,max}$ of each coordinate in the decision parameters space, r is defined as follows:

$$r = \frac{1}{2}\sqrt{\sum_{k=1}^{n}(x_{k,max} - x_{k,min})^2} \tag{27}$$

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

$$c\rho^n = \frac{1}{q}cr^n \tag{28}$$

which yields

$$\rho = \frac{r}{\sqrt[n]{q}} \tag{29}$$

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

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

3.4 MPR analysis

Our research focuses 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. Thus, as has been done in earlier studies of GA niching (Miller and Shaw 1996), we adopt the performance metric called the *maximum peak ratio 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 of the optima in the final population $\{\tilde{f}_i\}_{i=1}^{q}$, and the real optima of the objective function $\{\hat{\mathcal{F}}_i\}_{i=1}^{q}$, the *maximum peak ratio* is defined as follows:

$$MPR = \frac{\sum_{i=1}^{q}\hat{\mathcal{F}}_i}{\sum_{i=1}^{q}\tilde{f}_i} \tag{30}$$

Also, given a maximization problem the MPR is defined as the obtained optima divided by 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.

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

The logistic equation. A simple modeling of the human population growth is often described by the following differential equation:

$$\frac{dy}{dt} = cy\left(1 - \frac{y}{a}\right), \tag{31}$$

with the solution

$$y(t) = \frac{a}{1 + \exp\{c(t - T)\}} \tag{32}$$

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 in 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*.

3.5 Previous results

This MPR time-dependent analysis was applied in Shir and Bäck (2005a) to two ES-based niching techniques: the Standard-ES Schwefel-approach niching (Shir and Bäck 2005b), and the CMA-ES niching. Here, some of the conclusions of that study are outlined:

1. The *niching formation acceleration*, expressed as the absolute value of c, had larger values for the CMA-ES mechanism for all the 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, the lower the absolute value of c.
3. The *learning period*, expressed as the value of T in the curve fitting, got 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 was larger in all of the test-cases for the CMA-ES mechanism. This result also supported the claim that the CMA-ES had a faster convergence, as it got better fitness values earlier.

The 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).

4 Experimental procedure

In the following section we shall describe an experimental setup for comparing the 5 derandomized variants with respect to the MPR analysis, and present the numerical results. We emphasize the fact that our set of core mechanisms consists of two classes:

- First-order mechanisms with $\mathcal{O}n$ strategy parameters: DR1, DR2.
- Second-order mechanisms with $\mathcal{O}n^2$ strategy parameters, which aim at achieving invariance with respect to the rotation operator: DR3, CMA, CMA+.

Thus, the CPU time profile differs, respectively, among the different variants. We, in any case, are interested in the convergence behavior subject to the same number of function evaluations and population settings.

4.1 Artificial test functions

We consider the following artificial multimodal continuous functions:

- $\mathcal{M}$: a basic hyper-grid 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.
- $\mathcal{A}$: 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 the up-going hyper-spheres. Ackley's function has been widely investigated in the context of *evolutionary computation* (Bäck 1996).
- $\mathcal{L}$: also known as $F2$, as had been originally introduced in Goldberg and Richardson (1987), is a sinusoid trapped in an exponent envelope. The parameter k determines the sharpness of the peaks in the function landscape. $\mathcal{L}$ has one global minimum, regardless of n and k. It has been a popular test function for GA niching methods. We set $k = 6$.
- $\mathcal{R}$: 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 consider its non-separable shifted-rotated variant (Suganthan et al. 2005).
- $\mathcal{G}$: the Griewank function (5) has its global minimum ($f^* = 0$) at the origin, with several thousand global minima in the area of interest. There are 4 sub-optimal minima $f \approx 0.0074$ with $\boldsymbol{x}^* \approx \left(\pm\pi, \pm\pi\sqrt{2}, 0, 0, 0, \ldots 0\right)$. We also consider its non-separable shifted-rotated variant (Suganthan et al. 2005).
- $\mathcal{F}$: the function after Fletcher and Powell (Bäck 1996) is a non-separable *non-linear parameter estimation problem*, which has a non-uniform distribution of 2^n minima.

Table 1 summarizes the unconstrained multimodal test functions as well as their initialization intervals.

Modus operandi.The 5 niching algorithms are tested on the specified functions for various dimensions.[1] Each test case includes 100 runs per algorithm. All runs are performed with a core mechanism of a $(1 \stackrel{+}{,} 10)$-strategy per niche and initial points are sampled uniformly within the initialization intervals. Initial step-sizes are set to $\frac{1}{4}$ of the intervals. The parameter q is set based on a-priori knowledge when available, or arbitrarily otherwise.

Function evaluations: the idea is to allocate a fixed number of evaluations per peak ($n \cdot 10^4$), and thus each run is stopped after $q \cdot n \cdot 10^4$ function evaluations.

As mentioned earlier, setting the parameter p reflects the trade-off between further sampling the search-space, on the expense of exploiting the function evaluations at the existing niches. Here, we set $p = 1$.

A curve fitting routine is applied to each run in order to retrieve the characteristic parameters of its logistic curve. This routine uses the least-squared-error method, and runs an optimization procedure to minimize it.

Numerical results. The numerical results are presented at several levels:

Niching acceleration. Table 2 presents the mean and the standard deviations for the parameter c over the 100 runs, as obtained by the curve fitting routine. There is a clear trend in the given numerical results—in the vast majority of the test cases, the DR2 algorithm has the highest absolute values of c, whereas the CMA+ has the lowest absolute values. This trend corresponds to having the highest niching acceleration and the lowest niching acceleration,

[1] Matlab source-code of the 5 routines is available at: http://www.liacs.nl/home/oshir/NichingES/

Table 1 Test functions to be *minimized*, initialization domains and number of desired peaks. For some of the non-separable functions, we apply translation and rotation: $\boldsymbol{y} = \mathcal{O}(\boldsymbol{x} - \boldsymbol{r})$, where $\mathcal{O}$ is an orthogonal rotation matrix, and $\boldsymbol{r}$ is a shifting vector

Name	Function	Init	Niches
Separable			
$\mathcal{M}$	$\mathcal{M}(\boldsymbol{x}) = -\frac{1}{n}\sum_{i=1}^{n} \sin^6(5\pi x_i)$	$[0,1]^n$	100
Ackley	$\mathcal{A}(\boldsymbol{x}) = -c_1 \cdot \exp\left(-c_2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right)$ $-\exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(c_3 x_i)\right) + c_1 + e$	$[-10,10]^n$	$2n+1$
$\mathcal{L}$	$\mathcal{L}(\boldsymbol{x}) = -\prod_{i=1}^{n} \sin^k(l_1\pi x_i + l_2)\exp\left(-l_3\left(\frac{x_i - l_4}{l_5}\right)^2\right)$	$[0,1]^n$	$n+1$
Rastrigin	$\mathcal{R}(\boldsymbol{x}) = 10n + \sum_{i=1}^{n}\left(x_i^2 - 10\cos(2\pi x_i)\right)$	$[-1,5]^n$	$n+1$
Griewank	$\mathcal{G}(\boldsymbol{x}) = 1 + \sum_{i=1}^{n}\frac{x_i^2}{4000} - \prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}}\right)$	$[-10,10]^n$	5
Non-separable			
Fletcher-Powell	$\mathcal{F}(\boldsymbol{x}) = \sum_{i=1}^{n}(A_i - B_i)^2$ $A_i = \sum_{j=1}^{n}\left(a_{ij}\cdot\sin(\alpha_j) + b_{ij}\cdot\cos(\alpha_j)\right)$ $B_i = \sum_{j=1}^{n}\left(a_{ij}\cdot\sin(x_j) + b_{ij}\cdot\cos(x_j)\right)$ $a_{ij}, b_{ij} \in [-100,100]; \boldsymbol{\alpha} \in [-\pi,\pi]^n$	$[-\pi,\pi]^n$	10
Shift-Rot Rast.	$\mathcal{R}_{RS}(\boldsymbol{x}) = 10n + \sum_{i=1}^{n}\left(y_i^2 - 10\cos(2\pi y_i)\right)$	$[-5,5]^n$	$n+1$
Shift-Rot Grie.	$\mathcal{G}_{RS}(\boldsymbol{x}) = 1 + \sum_{i=1}^{n}\frac{y_i^2}{4000} - \prod_{i=1}^{n}\cos\left(\frac{y_i}{\sqrt{i}}\right)$	$[0,600]^n$	5

respectively. Moreover, the 4 comma strategies have c values in the same order of magnitude, where the CMA+ usually has the lowest absolute value among them.

MPR saturation. This scalar value represents, to some degree, the quality of the obtained minima, and thus the final result of the niching process. Table 3 presents the mean and the standard deviation of the saturation MPR values for the different test cases. As can be seen in this table, the CMA-$\left(\begin{smallmatrix}+\\,\end{smallmatrix}\right)$ algorithms achieve the highest MPR values, and as far as the niching process is concerned—together they outperform the other methods. However, for the given test cases, there is no clear winner for the MPR value.

Global minimum. Table 4 contains the percentage of runs in which the global minimum was located. $\mathcal{M}$ is discarded from the table, as its global minimum was always found, by all algorithms, for every dimension n under investigation. Generally speaking, the CMA-$\left(\begin{smallmatrix}+\\,\end{smallmatrix}\right)$ routines, and in particular the CMA+ strategy, are superior with respect to the other derandomized variants.

One can also observe a strong correlation between Tables 3 and 4: routines that obtain a high MPR saturation value, i.e. locate the high-quality peaks, usually perform well globally and locate the global minimum in a high percentage of the runs.

The c–T tradeoff hypothesis. We would like to numerically assess the hypothesis claiming the existence of a tradeoff between the learning period T and the niching acceleration c (Shir and Bäck 2005a), with respect to the 5 algorithms under investigation.

We consider two test functions of the suite, one per class: the *separable* $\mathcal{M}$ and the *non-separable* $\mathcal{G}_{RS}$ (the *Shifted Rotated Griewank*). For each we run the algorithms for an

Table 2 The parameter c obtained from the curve fitting: mean and standard deviation over 100 runs

Test-case	DR1	DR2	DR3	CMA	CMA+
$\mathcal{M} : n = 3$	-0.107 ± 0.006	-0.138 ± 0.009	-0.106 ± 0.010	-0.069 ± 0.005	-0.054 ± 0.003
$\mathcal{M} : n = 10$	-0.059 ± 0.002	-0.072 ± 0.002	-0.071 ± 0.003	-0.040 ± 0.001	-0.015 ± 0.001
$\mathcal{M} : n = 40$	-0.027 ± 0.001	-0.033 ± 0.001	-0.024 ± 0.001	-0.013 ± 0.001	-0.003 ± 0.001
$\mathcal{A} : n = 3$	-0.153 ± 0.038	-0.226 ± 0.058	-0.167 ± 0.006	-0.135 ± 0.033	-0.048 ± 0.006
$\mathcal{A} : n = 10$	-0.063 ± 0.009	-0.079 ± 0.013	-0.071 ± 0.011	-0.055 ± 0.011	-0.017 ± 0.001
$\mathcal{L} : n = 3$	-0.164 ± 0.070	-0.194 ± 0.124	-0.151 ± 0.064	-0.148 ± 0.047	-0.063 ± 0.030
$\mathcal{L} : n = 10$	-0.150 ± 0.015	-0.186 ± 0.024	-0.143 ± 0.057	-0.147 ± 0.016	-0.040 ± 0.003
$\mathcal{R} : n = 3$	-0.022 ± 0.032	-0.035 ± 0.042	-0.009 ± 0.012	-0.030 ± 0.024	-0.010 ± 0.011
$\mathcal{R} : n = 10$	-0.046 ± 0.007	-0.049 ± 0.010	-0.039 ± 0.017	-0.022 ± 0.007	-0.016 ± 0.002
$\mathcal{G} : n = 3$	-0.012 ± 0.014	-0.025 ± 0.017	-0.012 ± 0.003	-0.023 ± 0.040	-0.006 ± 0.012
$\mathcal{G} : n = 10$	-0.031 ± 0.027	-0.102 ± 0.020	-0.031 ± 0.030	-0.023 ± 0.003	-0.019 ± 0.015
$\mathcal{F} : n = 3$	-0.022 ± 0.023	-0.042 ± 0.017	-0.024 ± 0.024	-0.023 ± 0.025	-0.015 ± 0.012
$\mathcal{F} : n = 10$	-0.054 ± 0.093	-0.087 ± 0.105	-0.078 ± 0.123	-0.044 ± 0.083	-0.022 ± 0.021
$\mathcal{R}_{RS} : n = 3$	-0.157 ± 0.036	-0.254 ± 0.053	-0.178 ± 0.047	-0.200 ± 0.041	-0.055 ± 0.008
$\mathcal{R}_{RS} : n = 10$	-0.072 ± 0.026	-0.095 ± 0.019	-0.083 ± 0.025	-0.072 ± 0.027	-0.020 ± 0.002
$\mathcal{G}_{RS} : n = 3$	-0.108 ± 0.067	-0.126 ± 0.074	-0.118 ± 0.064	-0.113 ± 0.069	-0.050 ± 0.007
$\mathcal{G}_{RS} : n = 10$	-0.056 ± 0.015	-0.072 ± 0.015	-0.085 ± 0.020	-0.090 ± 0.012	-0.020 ± 0.004

Table 3 The saturation MPR value: mean and standard deviation over 100 runs

Test-case	DR1	DR2	DR3	CMA	CMA+
$\mathcal{M} : n = 3$	1 ± 0	1 ± 0	1 ± 0	1 ± 0	1 ± 0
$\mathcal{M} : n = 10$	1 ± 0	1 ± 0	1 ± 0	1 ± 0	1 ± 0
$\mathcal{M} : n = 40$	0.997 ± 0.002	1 ± 0	0.988 ± 0.003	1 ± 0	1 ± 0
$\mathcal{A} : n = 3$	0.971 ± 0.029	0.966 ± 0.028	0.960 ± 0.030	0.977 ± 0.024	0.992 ± 0.017
$\mathcal{A} : n = 10$	0.901 ± 0.024	0.905 ± 0.025	0.901 ± 0.025	0.920 ± 0.023	0.942 ± 0.023
$\mathcal{L} : n = 3$	0.963 ± 0.028	0.945 ± 0.038	0.953 ± 0.029	0.962 ± 0.027	0.996 ± 0.006
$\mathcal{L} : n = 10$	0.505 ± 0.163	0.379 ± 0.153	0.167 ± 0.129	0.596 ± 0.148	0.562 ± 0.109
$\mathcal{R} : n = 3$	0.263 ± 0.314	0.245 ± 0.036	0.233 ± 0.042	0.143 ± 0.046	0.481 ± 0.124
$\mathcal{R} : n = 10$	0.052 ± 0.007	0.063 ± 0.007	0.055 ± 0.005	0.057 ± 0.009	0.053 ± 0.005
$\mathcal{G} : n = 3$	0.115 ± 0.168	0.526 ± 0.470	0.366 ± 0.050	0.223 ± 0.288	0.761 ± 0.098
$\mathcal{G} : n = 10$	0.024 ± 0.042	0.026 ± 0.047	0.066 ± 0.018	0.015 ± 0.017	0.079 ± 0.029
$\mathcal{F} : n = 3$	0.002 ± 0.002	0.002 ± 0.002	0.002 ± 0.002	0.003 ± 0.004	0.002 ± 0.001
$\mathcal{F} : n = 10$	0.001 ± 0.001	0.001 ± 0.001	0.001 ± 0.001	0.001 ± 0.001	0.001 ± 0.001
$\mathcal{R}_{RS} : n = 3$	0.409 ± 0.111	0.463 ± 0.067	0.423 ± 0.117	0.469 ± 0.103	0.563 ± 0.098
$\mathcal{R}_{RS} : n = 10$	0.085 ± 0.015	0.099 ± 0.019	0.078 ± 0.015	0.108 ± 0.017	0.071 ± 0.014
$\mathcal{G}_{RS} : n = 3$	0.072 ± 0.043	0.078 ± 0.044	0.085 ± 0.048	0.082 ± 0.036	0.108 ± 0.041
$\mathcal{G}_{RS} : n = 10$	0.134 ± 0.038	0.144 ± 0.037	0.122 ± 0.035	0.161 ± 0.034	0.045 ± 0.013

increasing dimensionality of $n = 3, 4, \ldots, 30$, and obtain the MPR parameters for 100 runs—in order to plot c as a function of T.

Figures 1 and 2 present the c–T curves for $\mathcal{M}$ and $\mathcal{G}_{RS}$, respectively. The curves reflect a clear trade-off between c and T over the dimensions for the algorithms for both cases (an

Table 4 Global minimum reached in 100 runs

Test-case	DR1 (%)	DR2 (%)	DR3 (%)	CMA (%)	CMA+ (%)
$\mathcal{A} : n = 3$	100	100	100	100	100
$\mathcal{A} : n = 10$	90	91	90	92	95
$\mathcal{L} : n = 3$	93	74	92	97	100
$\mathcal{L} : n = 10$	9	2	0	17	13
$\mathcal{R} : n = 3$	20	19	13	16	48
$\mathcal{R} : n = 10$	0	0	0	0	0
$\mathcal{G} : n = 3$	13	21	32	13	88
$\mathcal{G} : n = 10$	8	16	4	16	2
$\mathcal{F} : n = 3$	100	100	100	100	100
$\mathcal{F} : n = 10$	14	12	15	23	15
$\mathcal{R}_{RS} : n = 3$	45	40	39	54	72
$\mathcal{R}_{RS} : n = 10$	0	0	0	0	0
$\mathcal{G}_{RS} : n = 3$	4	2	4	12	8
$\mathcal{G}_{RS} : n = 10$	6	1	3	14	0

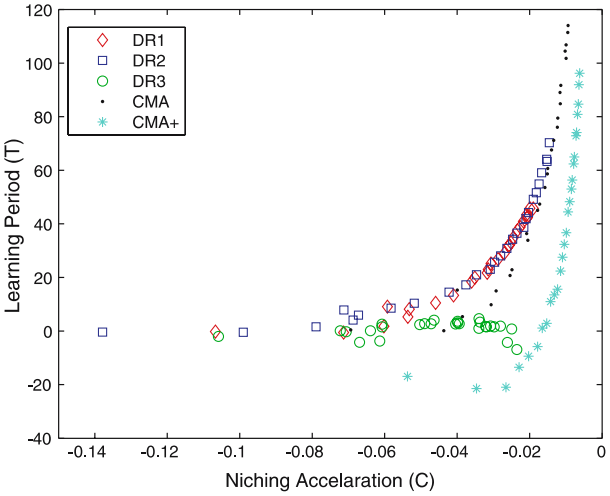

Fig. 1 The $c-T$ curve for $\mathcal{M}$: clear trade-off for the different algorithms, except for DR3, which has a flat curve

exception—the DR3 over $\mathcal{M}$). We consider this a numerical assessment for the hypothesis—the longer the learning period, the lower the niching acceleration.

Intermediate discussion. We would like to suggest an explanation for the advantage of a plus strategy for niching. The niching problem can be considered as an optimization task with constraints, i.e., the formation of niches that restricts competing niches and their optimization routine of 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 will tend to fail with a comma-strategy, and thus a plus-strategy is preferable for such problems. We might link this argumentation to the observation of our numerical results here, and suggest that a plus-strategy is preferable for niching.

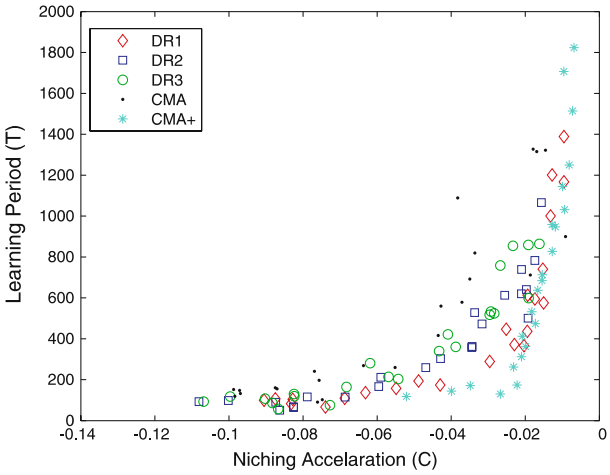

Fig. 2 The $c-T$ curve for $\mathcal{G}_{RS}$: clear trade-off for the 5 different algorithms

4.2 Quantum control application: molecular alignment

In this section we describe a challenging optimization problem from the field of *quantum control*, to which we apply our DES niching techniques. The reader who wants to abstract from the physics details can view the problem as a single-criterion 80-dimensional optimization task, subject to maximization, with a punishment term for handling a physics constraint.

This work is based on a joint study with the XUV-group at Amolf-FOM, *Institute for Atomic and Molecular Physics*, Amsterdam.

Dynamic alignment: physics background. The interaction of a generic linear molecule with a laser field is described within the framework introduced in Rosca-Pruna and Vrakking (2002). Briefly, we calculate the time evolution of molecules quantum mechanically by performing calculations for the ground initial rotational level, i.e., $J_{initial} = 0$, at zero Kelvin temperature. We take the molecule to be a rigid rotor. Two electronic states are taken into account, the ground state denoted by X and an off-resonant excited state denoted by A. The wavefunction is thus expanded as

$$\Psi(t) = \sum_{J=0}^{N_{rot}} \alpha_{XJ}(t)\psi_{XJ} + \alpha_{AJ}(t)\psi_{AJ} \tag{33}$$

with $N_{rot} = 20$ rotational levels, where the ψ_{XJ} correspond to even Legendre polynomials and the ψ_{AJ} correspond to odd Legendre polynomials. This expansion was confirmed to give converged results in the present calculations.

The time dependence of the molecular wave function is given by

$$i\frac{\partial \Psi}{\partial t} = (H_0 + V)\Psi(t) \tag{34}$$

he Hamiltonian consists of a molecular part H_0 and the interaction with the laser field, given by

$$V = \boldsymbol{\mu} \cdot \boldsymbol{E}(t)\cos(\omega t) \tag{35}$$

The Eigenenergies of H_0 are given by

$$E(J) = hcBJ(J+1) \tag{36}$$

where B is the *rotational constant* of the molecule. The laser field induces transitions between the rotational states which, in the off-resonant case, occur via subsequent Raman processes. The transitions between X and A were assumed to proceed via the selection rules $\Delta J = \pm 1$, $\Delta M = 0$.

The envelope of the laser field (which completely determines the dynamics after the transition to the rotating frame has been performed) is described by

$$E(t) = \mathbb{R}\left\{ \int A(\omega)\exp(i\phi(\omega))\exp(i\omega t)\ d\omega \right\} \tag{37}$$

The spectral function $A(\omega)$ is taken to be a Gaussian with a width chosen such that the *full-width-at-half-maximum* (FWHM) length of the *Fourier transform limited* (FTL) pulse (obtained by setting $\phi(\omega) \equiv 0$) equals 100 fs. The control function is the phase function $\phi(\omega)$, which defines the phase at a set of n frequencies that are equally distributed across the spectrum of the pulse.

These n values $\{\phi(\omega_i)\}_{i=1}^{n}$ are the decision parameters to be optimized. In order to achieve a good trade-off between high resolution and optimization efficiency, the value of $n = 80$ turned to be a good compromise. The search space is therefore an 80-dimensional hypercube spanning a length of 2π in each dimension. Real-laboratory pulse shapers use typically 64 or 128 pixels for the spectral phase, so in that respect our simulation reflects a feasible realization. Furthermore, phase-only shaping was observed to be sufficient for obtaining optimal control in this problem, and thus we restrict our work to phase modulation.

In this study we consider a numerical modeling of the dynamic molecular alignment, driven by a known Hamiltonian, but designed in a lab-oriented manner. In essence, it is Quantum Control Theory combined with some Quantum Control Experiments characteristics. We choose to restrict this study to noise-free simulations, as we are mainly interested in the physics of the system, rather than the actual simulation of a real laboratory experiments.

Optimization. We consider the goal of optimizing the alignment of a sample of generic diatomic molecules undergoing irradiation by a shaped femtosecond laser. We have used the maximum $\langle \cos^2(\theta) \rangle$ that occurs under field free conditions after the laser pulse, where θ is the angle between the molecular and the laser polarization axis, as a measure of the alignment. The rotational constant was chosen to be $B = 5\ \mathrm{cm}^{-1}$. The peak Rabi frequency between the two electronic states X and A, that determines the interaction strength, was $\Omega_{XA} = 1.6 \cdot 10^{14}\ \mathrm{s}^{-1}$.

Since we want to achieve a high degree of alignment with a peak intensity as low as possible, an additional constraint was introduced as a punishment for pulses that are too intense. We have used

$$I_p = \int E^2(t)\Theta(E^2(t) - I_{thr})\ dt \tag{38}$$

(where $\Theta(x)$ is the Heaviside step function) for this purpose, so that the fitness function assigned to a pulse shape was

$$F = \max_{E(t)} \langle \cos^2(\theta) \rangle - I_p. \tag{39}$$

The calculation is done in the time range of [−15 ps,15 ps]. Figure 3 provides an illustrative overview of the numerical process.

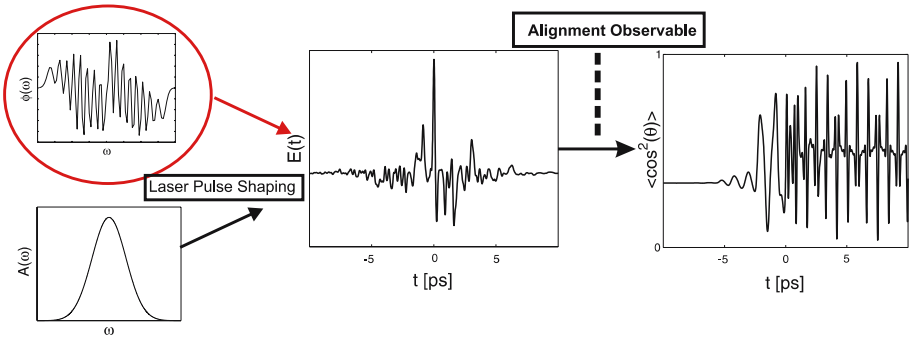

Fig. 3 The numerical process: an overview. The control function is the phase (circled, top left), the amplitude function is fixed and approximated by a Gaussian (bottom left). The shaping process (Eq. 37) generates the electric field, $E(t)$ (center). The "Schrödinger Box" of the alignment observable represents the numerical calculation of the interaction between the electric field with the molecules, based on the specified quantum dynamics equations. The revival structure (right) is the observed simulated behavior of the molecules, upon which the yield value is based

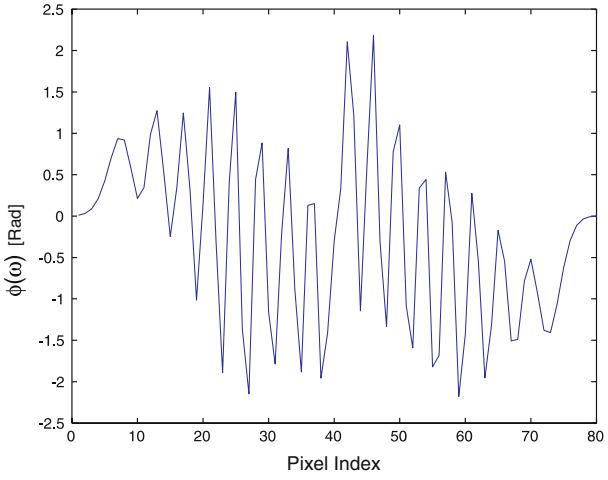

Fig. 4 A phase function obtained by an evolutionary search (frequency domain)

A typical phase function and a typical laser pulse, obtained by an evolutionary optimization, are given for illustration as Figs. 4 and 5, respectively.

Previous work. A variant of this dynamic molecular alignment optimization problem, with

$$T = 100\,K, \quad J_{initial}^{max} = 7,$$

has been investigated in (Siedschlag et al.2006). That recent study showed that it payed off to use more elaborated optimization schemes, and in particular derandomized evolution strategies, for such a high-dimensional optimization task. In particular, the DR2 algorithm outperformed the other methods, obtaining the best cosine-squared alignment.

Application of niching to the problem. Aiming to apply niching to the dynamic alignment problem, we have been required to define an appropriate *distance metric*. We should note at this point that the function undergoing the evaluation in the simulation is actually

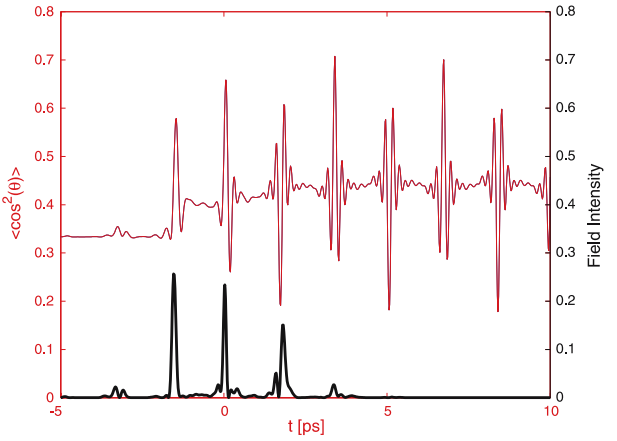

Fig. 5 Given the phase function in Fig. 4—thin red line: alignment; thick black line: intensity of the laser pulse (time domain)

$E^2(t)$, i.e., the *time-dependent laser intensity*. Hence, the outcome of the calculations was invariant under the following transformations:

- $\tilde{\phi}(\omega) = \phi(\omega) + \phi_0$: this would add a multiplication constant after the Fourier transform is calculated.
- $\tilde{\phi}(\omega) = \phi(\omega) + c \cdot \omega$: this would simply shift the entire pulse with respect to the time origin and therefore has no observable effect.

These invariance properties had to be taken into account when defining a distance metric between two individuals $\phi_i(\omega)$ and $\phi_j(\omega)$, as it is clear that using the straightforward approach, i.e. the Euclidean distance metric, would not achieve the goal: due to the fact that $\phi(\omega)$ is invariant under the specified transformations, calculating the distance between two feasible solutions $\phi(\omega)$, $\tilde{\phi}(\omega)$ would not guarantee that the derived pulses $f_i(t)$, $f_j(t)$ would have a different profile. Thus, a new distance metric that would remove this degeneracy is needed here.

Our proposed solution, concluded from the specified invariance properties, is to apply the distance metric in the *second-derivative space* of $\phi(\omega)$. Explicitly, given that the discretization is to n function values, the distance between $\phi_i(\omega)$, $\phi_j(\omega)$ was given by:

$$d_{i,j} = \sqrt{\sum_{k=1}^{n} \left(\left(\frac{\partial^2 \phi_i(\omega)}{\partial \omega^2} \right)_k - \left(\frac{\partial^2 \phi_j(\omega)}{\partial \omega^2} \right)_k \right)^2} \tag{40}$$

Experimental setup. We consider here three niching strategies:

1. The (1, λ)-DR2—for being the best method to perform on a different variant of this problem, and also as a representative of first-order strategies.
2. The (1, λ)-CMA—as a representative of second-order strategies.
3. The (1 + λ)-CMA—as a representative of plus strategies.

We conduct 10 runs per method, searching for $q = 3$ niches, subject to phase-function parameterization of $n = 80$. Moreover, we discuss additional technical aspects of our experimental setup:

- In this study we focus in the simplified variant of $T = 0$, with only $J^{max}_{initial} = 0$ in the initial distribution (versus $T = 100\ K$ and $J^{max}_{initial} = 7$ in Siedschlag et al. (2006). From the technical perspective, this means a dramatic reduction in the evaluation time: 6

seconds in the simplified model, versus 35 seconds in the original case—per function evaluation.
- The cosine-squared alignment gets values within [0,1]; the best value known to us in the current settings is $f^* = 0.9622$.
- *Each run was limited to* 15,000 *function evaluations per niche*.
- The niche radius was initially set to $\rho = 110$; see the Appendix for the explicit calculation.
- Implementation was done in Fortran for the numerical simulation, where the niching was based on the Matlab code mentioned earlier.

Experimental observation. The results of our experiments are discussed at several levels:

Niche-radius. The initial setting of the niche-radius, $\rho = 110$, failed to obtain satisfying performance. The DR2 as well as the CMA-comma routines did not succeed in delivering good solutions. The CMA-plus, however, managed to find good solutions for the first niche only; the second and third niches were not populated by good solutions. By dividing the niche radius by half, i.e. $\tilde{\rho} = 55$, we managed to get satisfying results, as will be reported here. Thus, consider all the reported results from now on as such to be obtained with $\tilde{\rho} = 55$.

Success-rate, i.e., the *cosine-squared alignment* of the three methods, for the three obtained niches, is presented at Table 5. One can observe a clear trend—the CMA+ mechanism outperforms the other mechanisms, with consistent location of three good niches on average. However, the DR2 mechanism managed to obtain the best solution for the best and 2^{nd}-best niches, in consistency with previous results (Siedschlag et al. 2006). The latter usually fails to locate a 3^{rd} good niche. The CMA comma-strategy, on the other hand, simply fails to deliver satisfying niching results on this landscape.

Laser pulse designs. Our definition of the distance metric for this problem has been proved to be successful. The obtained pulses in the time-domain had indeed different characteristics, representing different conceptual laser-pulse designs. Note that an optimal solution is represented by its phase function $\phi(\omega)$, and is actually examined by its well-defined electric field and the revival structure of the ensemble of the molecules. Three niches, obtained in a typical CMA+ run, are given as illustration as Fig. 6a, c, e.

Conceptual physical structure. We would like to offer an additional analysis for our niching solutions. *Only due to our simplified variant, i.e.* $J^{max}_{initial} = 0$ *at initialization, it is possible to study the* population *of the rotational levels as a function of time*.[2] Otherwise, in the general case, all levels are initially populated, and a thermal averaging is applied. Explicitly, the wavefunction can be expressed as a superposition of those levels,

$$\psi = \sum_j a_j^{(t)} \cdot |j\rangle \cdot e^{-i\frac{E_j t}{\hbar}} \tag{41}$$

the expectation of the cosine-squared alignment (the objective function) is calculated directly from these complex amplitudes $a_j^{(t)}$, whereas the *population* of the rotational levels is $\left|a_j^{(t)}\right|^2$. This population of rotational levels can be analyzed in a fairly simple technique, known as the *Sliding Window Fourier Transform* (SWFT), which provides us with a

[2] The careful reader should note that *'population'* is used here exclusively in the context of *quantum mechanics*, e.g., populating quantum levels.

Table 5 Three niches obtained in 10 runs—averaged fitness values (in parentheses—best value found)

Ranked-niches	DR2	CMA	CMA+
Best niche	0.9417 (0.9605)	0.8553 (0.9029)	0.9517 (0.9585)
2nd-best niche	0.8477 (0.9552)	0.8229 (0.8561)	0.9493 (0.9525)
3rd-best niche	0.8053 (0.8558)	0.7966 (0.8161)	0.9365 (0.9484)

powerful visual tool. Given the revival structure of an optimal solution, a sliding time window is Fourier transformed, to produce the frequency picture through the alignment process. This windowing creates a transformation which is localized in time. Due to the *quantization* of the rotational levels, only certain frequencies (or *energy* levels, respectively) are expected to appear.

Figure 6b, d, f provide the SWFT picture for the obtained best solutions. The reader can observe three pictures of quantum energy-levels population for the different solutions. However, these three solutions represent the same conceptual physical structure of population. This SWFT observation reinforces a recent study which showed a correlation between the optimization routine and the parameterization to specific conceptual structures (Shir et al. 2007). Thus, based on that study, it is not surprising that all three pulses share the same 'behind-the-scenes physics', due to the fact that they were all obtained with the same core mechanism (e.g., CMA+), subject to the same parameterization. This observation does not contradict our primary conclusion that the niching process has been successful in locating three different pulse shapes in the time domain. It simply reveals an additional, well-hidden, degeneracy among the solutions. In the next section we will offer a way to remove this second degeneracy completely.

Removing the second degeneracy. Given the additional degeneracy which was encountered in the SWFT space, one can further develop a problem-specific diversity measurement. In this case, our idea is to consider the *wavepackets space*: the differences between the population of rotational levels, $\left|a_j^{(t)}\right|^2$, as the measurement of diversity between niches—and by that to allow obtaining new conceptual designs of laser pulses. The implementation is straightforward, as the vector of population coefficients is given by the alignment-routine. Due to the normalization of the coefficients, subject to the basic postulate of Quantum Mechanics, it is fairly simple to estimate the niche radius also in this case: We set it to $\rho = 0.5$ (see the Appendix for the calculation). We choose to apply only the CMA+ routine in this case, aiming to show feasibility of the defined distance metric.

This newly-defined diversity-measurement for the laser problem has been observed to be successful. The obtained pulses in the time domain had indeed different characteristics, and in particular their shapes differed in a satisfying manner. The best niche obtained in every run is usually of the optimal type known to us—both the cosine-squared alignment yield, as well as the pulse shape and the population profile are associated with best solutions obtained in the past by unimodal optimization. The second-best niche is a representative of a sub-optimal set of solutions—it has a lower value of cosine-squared alignment yield and a different profile of population. However, note that the third-best located niche is usually not an interesting solution, as it has dramatically lower alignment values in comparison to the first two niches. The numerical values of the runs are omitted here, as they are fully consistent with the CMA+ results given earlier for the first two best niches.

Typical solutions of best and second-best niches are plotted in Figs. 7a, b and 8a, b—the pulse-shapes as well as their SWFT pictures.

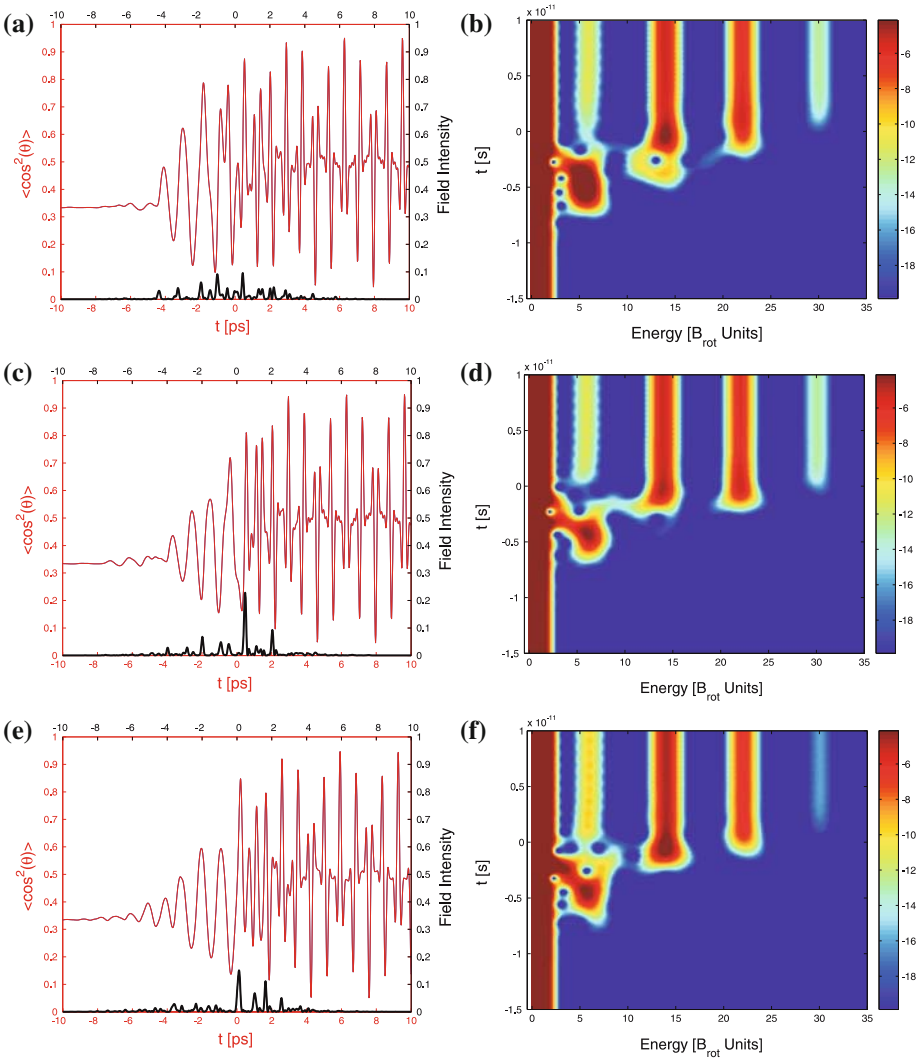

Fig. 6 (**a**) Best niche: $\langle \cos^2(\theta) \rangle = 0.9524$. (**b**) SWFT of best niche's solution. (**c**) 2nd-best niche: $\langle \cos^2(\theta) \rangle = 0.9513$. (**d**) SWFT of 2nd-best niche's solution. (**e**) 3rd-best niche: $\langle \cos^2(\theta) \rangle = 0.9466$. (**f**) SWFT of 3rd-best niche's solution. *Notes*: (**a**, **c**, **e**) *alignment* and *revival-structure* of the three niches obtained by the $(1 + \lambda)$-CMA. Thin red line: alignment; thick black line: intensity of the laser pulse. (**b**, **d**, **f**) SWFT of the solutions: a *Fourier transform* applied to the revival strictures of the optimal solutions (the thin-red alignment curves). The values are log-scaled 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

5 Discussion and outlook

We have introduced a family of advanced ES niching techniques, based on 5 variants of derandomized ES, with a fixed niche radius approach. We then presented a survey of these niching methods to a suite of artificial as well as real-world test problems.

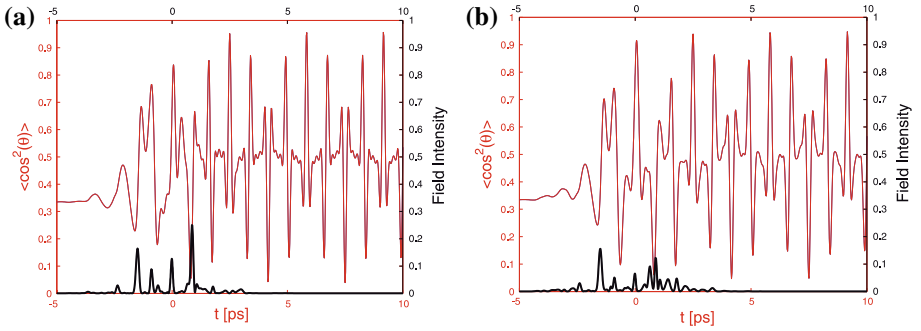

Fig. 7 (**a**) Optimal niche: $\langle \cos^2(\theta) \rangle = 0.9596$. (**b**) Sub-optimal niche: $\langle \cos^2(\theta) \rangle = 0.9472$. *Notes*: Alignment and revival structure of a typical run—the two leading niches. Thin red line: alignment; thick black line: intensity of the laser pulse

Generally speaking, the CMA-$\left(\begin{smallmatrix} + \\ , \end{smallmatrix}\right)$ routines, and in particular the CMA+ strategy, were found to be superior with respect to the other derandomized variants on the artificial landscapes. The DR2 variant has the highest niching acceleration, as opposed to the CMA+ with the lowest niching acceleration.

The low niching acceleration of the *plus-strategy* seems to be the key for the successful niching, allowing it to obtain the best location of the global minimum, and to reach the highest MPR saturation values. The *niching acceleration* seems to originate, to our best understanding, in the adaptation profile of the step-size, and apparently the CMA+ mechanism offers a profile which suits niching very well.

Moreover, the hypothesis claiming that there exists a trade-off between the learning period and the niching acceleration has been numerically assessed in this study.

On the real-world problem, the so-called dynamic molecular alignment optimization problem, a degeneracy in the default diversity-measurement was observed, due to some invariance properties of the Fourier Transform. We offered a problem-specific diversity measurement to overcome it, which was successful.

Regarding the performance of the different routines on this application, the CMA+ performed the best when obtaining always three niches of high-quality laser pulses. The DR2 found the best solution, in consistency with past results, but did not perform well on the 2nd-best or the 3rd-best niches. The CMA comma-strategy failed to deliver satisfying niching results. The original calculation of the niche radius was not successful at the practical level. The results reported here were obtained only after introducing a factor of 0.5 to the original value. We believe that this suggests a search space with a limited regime of good solutions. Essentially, based on the argument given in Sect. 4.1, we would argue that introducing a large niche radius could introduce an impossible constrained problem. This should remind us that the given calculation of the niche radius is merely a proposed approximation, and moreover, we should keep in mind that the niche radius is a sensitive yet crucial component of this mechanism.

Furthermore, we have applied a physics numerical assessment, at the quantum energy level, with the so-called SWFT technique. The latter has supported previous observations concerning the correlation between optimization routines and parameterizations to conceptual physical structures. This observation revealed that all three niches of a given run, which differ dramatically at the laser-pulse design level (temporal space), usually share the same conceptual physical structure at the quantum-energy level (wavepacket space). We

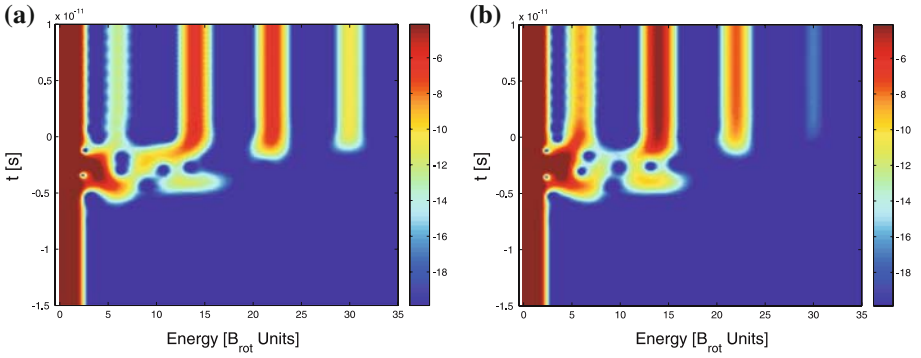

Fig. 8 (**a**) Optimal niche: 4th rotational level is mostly populated after the interaction. (**b**) Sub-optimal niche: 3rd rotational level is mostly populated after the interaction. Notes: The *Fourier transform* applied to the revival structures of the optimal solutions (the thin-red alignment curves Fig. 7a, b). The values are log-scaled, and represent how high the rotational levels of the molecules are populated as a function of time

offered another diversity measurement, which relies on the physics information, in order to overcome this second degeneracy. This approach managed to remove the degeneracy, and offer multiple solutions corresponding to different conceptual designs.

Overall, we would like to stress that given a challenging real-world optimization problem, it is possible to apply niching techniques with a fixed niche radius approach, in order to provide a *decision maker* with multiple good conceptual designs. We managed to show that when degeneracy is revealed in the landscape, it is possible to design problem-specific diversity measurements in order to overcome it.

Here are some directions for future study in this domain:

- Developing a DR2 algorithm with a plus-strategy (DR2+). Such core mechanism can be powerful in the context of niching for the real-world problem discussed here, as well as for other problems. The DR2 algorithm seems to be a strong first-order approach, and its plus-strategy may allow it to outperform the other variants.
- Proceeding with the effort to tackle the niche radius problem, in order to develop state-of-the-art ES niching techniques which are not subject to the niche radius assumptions.

Acknowledgments The authors would like to thank Christian Siedschlag and Marc Vrakking, of Amolf-FOM Amsterdam, for the collaboration on the laser problem, and Michael Emmerich, of Leiden University, for the fruitful discussions on niching.This work is part of the research programme of the 'Stichting voor Fundamenteel Onderzoek de Materie (FOM)', which is financially supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)'.

Appendix: Niche radius calculation for the laser problem

We provide here the explicit calculations for the niche radii in the dynamic alignment problem.

Second-derivative space

The distance metric in the dynamic alignment problem is applied in the second-derivative space. Numerically, the derivative is simply implemented by the MATLAB command

`diff`. Thus, after the double-application of `diff` to the original phase-vector of dimension $n = 80$, the modified vector $\boldsymbol{y}$ is reduced to dimension $n^* = n-2 = 78$. Given the original upper and lower bounds of the decision parameters,

$$x_{k,min} = 0, x_{k,max} = +2\pi \quad k = 1\ldots80,$$

the first application of `diff` will make new bounds of

$$\tilde{x}_{k,min} = -2\pi, \tilde{x}_{k,max} = +2\pi \quad k = 1\ldots79,$$

and the second application will make it

$$y_{k,min} = -4\pi,\ y_{k,max} = +4\pi \quad k = 1\ldots78.$$

Plugging this into Eq. 29, we obtain:

$$\rho = \frac{\frac{1}{2}\sqrt{78 \cdot (8\pi)^2}}{3^{\frac{1}{78}}} \approx 110 \tag{42}$$

Wavepacket space

According to a Quantum Mechanics postulate, the wavepacket coefficients in the N-dimensional Hilbert space are normalized:

$$\sum_{j}^{N} \left| a_j^{(t)} \right|^2 = 1$$

In the wavepacket treatment for removing the second degeneracy, these coefficients play the role of the decision parameters for the diversity measurement. The calculation of r of Eq. 27 reads:

$$r = \frac{1}{2}\sqrt{\sum_{j=1}^{N_{rot}} \left| a_j^{(t)} \right|^2} = \frac{1}{2}$$

With $q = 3$ and $N_{rot} = 20$, Eq. 29 yields:

$$\rho = \frac{\frac{1}{2}}{3^{\frac{1}{20}}} \approx 0.47 \tag{43}$$

References

Avigad G, Moshaiov A, Brauner N (2004) Concept-based interactive brainstorming in engineering design. J Adv Comput Intell Intell Inform 8(5):454–459

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 evolutionary computation (ICEC'94), Orlando FL, vol 1. IEEE Press, Piscataway, NJ, USA, pp 57–62

Bäck T (1996) Evolutionary algorithms in theory and practice. Oxford University Press, New York, NY, USA

Beyer HG, Schwefel HP (2002) Evolution strategies a comprehensive introduction. Nat Comput Int J 1(1):3–52

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 Publishers Inc., San Francisco, CA, USA

Demiralp M, Rabitz H (1993) Optimally controlled quantum molecular dynamics: a perturbation formulation and the existence of multiple solutions. Phys Rev A 47(2):809–816

Goldberg DE, Richardson J (1987) Genetic algorithms with sharing for multimodal function optimization. In: Proceedings of the second international conference on genetic algorithms on genetic algorithms and their application. Lawrence Erlbaum Associates Inc, Mahwah, NJ, USA, pp 41–49

Hansen N, Ostermeier A (2001) Completely derandomized self-adaptation in evolution strategies. Evol 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)

Igel C, Suttorp 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 2005), ACM Press, pp 453–460

Igel C, Hansen N, Roth S (2007) Covariance matrix adaptation for multi-objective optimization. Evol Comput 15(1):1–28

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

Mahfoud S (1995) Niching methods for genetic algorithms. PhD thesis, University of Illinois at Urbana, Champaign

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, NY, USA

Nuernberger P, Vogt G, Brixner T, Gerber G (2007) Femtosecond quantum control of molecular dynamics in the condensed phase. Phys Chem Chem Phys 9(20):2470–2497

Ostermeier A, Gawelczyk A, Hansen N (1993) A derandomized approach to self adaptation of evolution strategies. Technical report, TU, Berlin

Ostermeier A, Gawelczyk A, Hansen N (1994) Step-size adaptation based on non-local use of selection information. In: PPSN. Volume 866 of lecture notes in computer science, Springer

Rabitz H, de Vivie-Riedle R, Mutzkus M, Kompa K (2000) Whither the future of controlling quantum phenomena? Science 288:824–828

Rosca-Pruna F, Vrakking M (2002) Revival structures in picosecond laser-induced alignment of i2 molecules. J Chem Phys 116(15):6579–6588

Schönemann L, Emmerich M, Preuss M (2004) On the extiction 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, 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, USA

Shir OM, Bäck T (2005b) Niching in evolution strategies. Technical report, Technical Report 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, 9th international conference, Reykjavik, Iceland, September 9–13, 2006, Proceedings. Volume 4193 of lecture notes in computer science, Springer, pp 142–151

Shir OM, Kok JN, Vrakking MJ, Bäck T (2007) Gaining insight into laser pulse shaping by evolution strategies. In: IWINAC. Volume 4527 of lecture notes in computer science, Springer

Siedschlag C, Shir OM, Bäck T, Vrakking MJJ (2006) Evolutionary algorithms in the optimization of dynamic molecular alignment. Opt Commun 264:511–518

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. Technical report, Nanyang Technological University, Singapore

Törn A, Zilinskas A (1987) Global optimization, vol 350. Springer

Warren WS, Rabitz HA, Dahleh M (1993) Coherent control of quantum dynamics: the dream is alive. Science 259:1581–1589

Weinacht TC, Bucksbaum PH (2002) Using feedback for coherent control of quantum systems. J Opt B 4:R35–R52