# Stopping Criteria for Multimodal Optimization

Simon Wessing[1], Mike Preuss[2], and Heike Trautmann[2]

[1] Department of Computer Science, TU Dortmund, Germany
simon.wessing@tu-dortmund.de
[2] Information Systems and Statistics Group, University of Münster, Germany
{mike.preuss,trautmann}@uni-muenster.de

**Abstract.** Multimodal optimization requires maintenance of a good search space coverage and approximation of several optima at the same time. We analyze two constitutive optimization algorithms and show that in many cases, a phase transition occurs at some point, so that either diversity collapses or optimization stagnates. But how to derive suitable stopping criteria for multimodal optimization? Experimental results indicate that an algorithm's population contains sufficient information to estimate the point in time when several performance indicators reach their optimum. Thus, stopping criteria are formulated based on summary characteristics employing objective values and mutation strength.

**Keywords:** Multimodal optimization, global optimization, multiobjective selection, convergence detection, stopping criteria.

## 1 Introduction

For quite some time, global optimization has been the predominant research direction in single-objective *evolutionary computation* (EC). While algorithms for obtaining more than one good solution at once have been investigated already in the 1970s (see [1] for a survey), the term *multimodal optimization* (MMO) has become publicly known only lately. It may be seen as superordinate concept that contains *niching* and related approaches, with the overall task to obtain a set of diverse but very good solutions. It is easy to imagine that such behavior is useful in many real-world applications, because it leaves more options to the decision maker (related arguments apply to multiobjective optimization).

In contrast to global optimization methods, MMO algorithms always employ populations and/or archives, and next to objective values, diversity is an important issue: the finally chosen best solutions should not at all be similar but be located in different search space regions. However, the interplay between reaching good objective values but keeping search diverse has not been investigated much from a general perspective, without focusing on a certain algorithm and/or optimization problem. Authors usually refer to exploitation/exploration balance, which means that there is a contradiction between improving solutions and covering the search space well. However, recent work on multiobjectivization-based selection criteria for MMO [2] suggests that it is possible to realize compromises

T. Bartz-Beielstein et al. (Eds.): PPSN XIII 2014, LNCS 8672, pp. 141–150, 2014.

between the two, such that diversity is kept and still the good solutions are improved further. In §3, we show that in most cases, the balance holds only temporarily: after some time, it usually breaks and multimodal performance (dealt with in §2) degrades again. Phenomenologically, this means that the algorithm moves in direction of one extreme (see Fig. 1): either it focuses on one or few attraction basins or it emphasizes diversity so that local optimization in the separate basins becomes ineffective. This naturally calls for stopping criteria as they are, e. g., known in multiobjective optimization (see §4). We do not state that at the determined point in time, optimization shall just be cancelled. But it undergoes a phase transition after which the algorithm does not sufficiently balance both goals any more, so that it may be supplemented with other techniques as local searches. It does not seem reasonable just to continue runs.

The first goal of this paper is to document this phase transition and provide data on where it can be expected in a run for different selection types, based on a simple model algorithm that may serve as blue print for more complex methods in MMO. The second goal is to suggest (§4), experimentally assess (§5) and discuss (§6) stopping criteria that detect the right time for a behavior change of the algorithm. Differently from the situation in single- or multiobjective optimization, the important indicators cannot be observed directly in a real-world application scenario. We would have to know in advance where the different optima are located to compute the indicators. However, we can offer criteria for mutation adaptive and non-adaptive optimization algorithms that provide a good estimation of the point in time when the phase transition occurs, so that measures against a degeneration of the optimization process can be taken.

## 2 MMO Performance Indicators and Model Algorithms

Several different approaches exist to measure performance of multimodal optimization algorithms [3]. To precisely assess the approximation of the optima, at least their locations and objective values have to be known. This information, and above all the exact shape of corresponding attraction basins, is of course only available for benchmarking purposes. In this case, the goal in one way or another is to measure deviations from the local optima in objective and/or in decision space. After carrying out our initial investigations (see §3) for all indicators in [3], we are focusing the presentation on the quality indicator *averaged Hausdorff distance* (AHD) [4], which is a natural advancement of the well-known indicator peak distance (PD) [3,5]. It is defined as

$$\mathrm{AHD}(\mathcal{P}, \mathcal{Q}) = \max\left\{\frac{1}{m}\sum_{i=1}^{m} d_{\mathrm{nn}}(\boldsymbol{z}_i, \mathcal{P}), \frac{1}{\mu}\sum_{i=1}^{\mu} d_{\mathrm{nn}}(\boldsymbol{x}_i, \mathcal{Q})\right\},$$

where $\mathcal{P} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_\mu\}$ is the approximation set, $\mathcal{Q} = \{\boldsymbol{z}_1, \dots, \boldsymbol{z}_m\}$ is the set of optima, and $d_{\mathrm{nn}}(\boldsymbol{x}, \mathcal{P}) = \min\{\|\boldsymbol{x} - \boldsymbol{y}\|_2 \mid \boldsymbol{y} \in \mathcal{P} \setminus \{\boldsymbol{x}\}\}$. PD is equivalent to the first term inside the maximum (which is also known as inverted generational distance). AHD is attractive, because it exhibits a continuous behavior over the

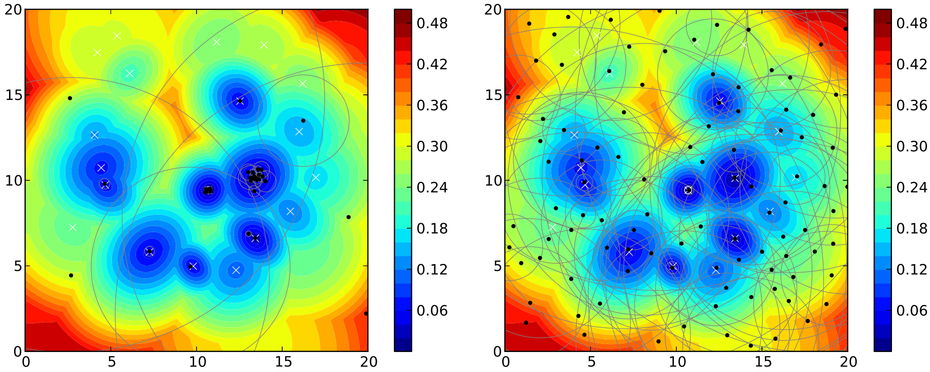


**Fig. 1.** Populations (black dots) after 5000 function evaluations with different selection variants (left: SV4, right: SV7). White crosses mark the local optima, a white circle the global one. Gray circles denote the mutation strengths of the respective individuals.

whole range from very bad to very good approximations. This is in contrast to, e. g., the basin ratio (BR), which measures the fraction of attraction basins that contain a point of the population. However, BR is still an informative indicator, as it has known minimal and maximal values. Other indicators, as PD or peak inaccuracy, are somewhat correlated to AHD, and thus our developments should be transferable to some extent.

Two simple evolutionary algorithms (EA) are considered in this paper. While both employ nearest-better distances $d_{\mathrm{nb}}(\boldsymbol{x}, \mathcal{P}) = \min\{\|\boldsymbol{x} - \boldsymbol{y}\|_2 \mid f(\boldsymbol{y}) < f(\boldsymbol{x}) \wedge \boldsymbol{y} \in \mathcal{P}\}$ in their selection and use gaussian mutations for variation, they exhibit very different behaviors. The first algorithm uses a multiobjective selection with $d_{\mathrm{nb}}(\boldsymbol{x}, \mathcal{P})$ as a second objective. The ranking is established by non-dominated sorting (and each non-dominated front is sorted by objective value). The second algorithm uses truncation selection on a lexicographic ordering according to the tuples $(-d_{\mathrm{nb}}(\boldsymbol{x}, \mathcal{P}), f(\boldsymbol{x}))$. (Note that reversing the order of the criteria would essentially lead to a conventional single-objective EA.) These selections have been defined as SV4 and SV7, respectively, in [2], and we adopt these names in the following. Details and pseudocode also can be found in [2].

## 3 Initial Investigations

For the analysis of the algorithms' behavior, we use the following experimental setup. A budget of $10^5$ objective function evaluations is allocated for optimization of the multimodal test problems described in [2] with a $(100+100)$-EA. The test problems are generated by taking the minimum of random samples of unimodal functions (so-called peaks, although pointing downwards). These samples can exhibit different global structures, which will be called *random*, *linear*, and *funnel* in the following. For further details we refer to [2]. For variation, we are using an isotropic mutation operator with gaussian random numbers and initial

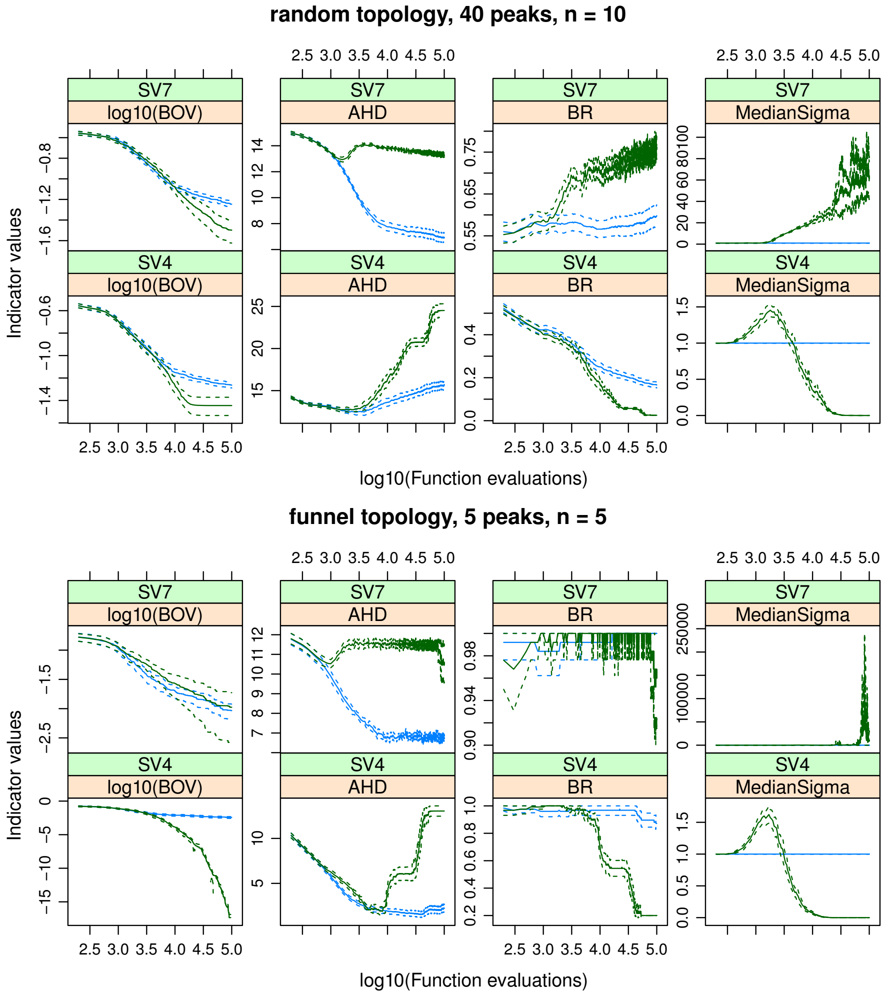


**Fig. 2.** EAs with fixed (blue) and self-adaptive (dark green) $\sigma$. Solid lines show mean values of 25 repeats, dashed lines are 95% pointwise confidence bands. BOV denotes best objective value and MedianSigma the median mutation strength in a population.

step size $\sigma_{\text{init}} = 1$. This operator can be used with either fixed or self-adaptive step sizes. In the latter case, the learning parameter is $\tau = 1/\sqrt{2n}$, according to recommendations of [6] for multimodal problems. Recombination is disabled.

Survivor selection is done by the two alternatives described in §2. Figure 1 shows snapshots of self-adaptive SV4- and SV7-EAs after 5000 function evaluations on a two-dimensional problem with 20 peaks and funnel topology. It can be seen that solutions close to optima possess small mutation strengths, while other solutions exhibit diverging step sizes. From Fig. 2 it is evident that SV7-EA constantly explores the search space, while SV4-EA will sooner or later converge to a single optimum. The figure shows the average performance on two selected problem classes over time. For SV4, step size adaptation first leads to a slight increase of $\sigma$ before the conventional convergence to zero starts. Therefore, the self-adaptive SV4-EA is the best suited for global optimization among the tested algorithms, as it does converge to one single optimum at some point, but does this later than a conventional single-objective EA. Note that SV4-EA has the same structure as [7], but is expected to yield better global optimization performance due to step-size adaptation and use of $d_{\mathrm{nb}}$. In this case, also the existing stopping criteria for single-objective optimization (see §4) can be applied.

SV7, on the other hand, exhibits a permanent tendency towards larger $\sigma$, which is sometimes beneficial but often leads to a deterioration of quality in the late stages of the run. Additionally, the algorithms' performance also depends on problem features as the search space dimension and the number of local optima. Here, low-dimensional, weakly multimodal problems favor SV4, while SV7 seems more adequate in the opposite case. Thus, if a diverse set of good solutions is sought as a result, special stopping criteria for multimodal optimization should be employed in any case.

## 4 Stopping Criteria

So far, research on stopping criteria within the field of EAs concentrated on assessing the convergence behavior of the respective algorithms. Formal analysis of convergence behavior is difficult and only possible for special and usually simplified cases. As optimality criteria such as the Karush-Kuhn-Tucker conditions usually cannot be applied in the black-box scenario due to the lack of sufficient gradient information, heuristic approaches were introduced to check whether the expected improvement in convergence is worth the additional amount of function evaluations which has to be spent. So far, to the best of our knowledge, no specific stopping criteria for multimodal optimization have been introduced, which have to be designed to focus on tracking the trade-off between maintaining diversity and ensuring sufficient convergence.

An overview about recent approaches for multiobjective optimization is provided in [8]. As most of the methods rely on analyzing (single-objective) performance indicators, the approaches in principle could be transferred to single-objective optimization tasks as well. However, none of these criteria allows for adequately terminating an EA in the multimodal situation in which the population is desired to converge while simultaneously maintaining diversity. In single-objective optimization the same problem exists. In [9] existing termination conditions in the single-objective case are discussed which consist of either theoretically motivated approaches [10,11], movement criteria [12], or qualified runtime distributions [13]. To our knowledge, criteria based on statistical hypothesis testing are surprisingly uncommon. It shall also be stated that contrary to intuition, deriving criteria for single-objective algorithms is *not* necessarily simpler than for multiobjective ones. As, e. g., the list of criteria in [14] demonstrates, there is much more information available for the latter case, so that many

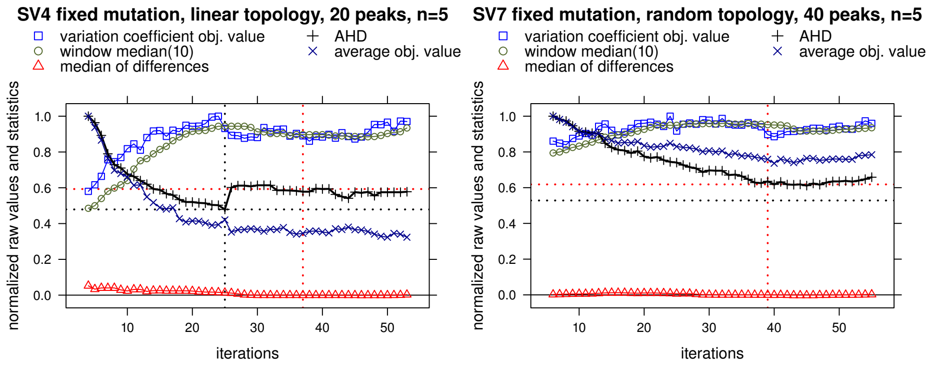


**Fig. 3.** Exemplary runs of SV4- and SV7-EAs with fixed $\sigma$. In both cases, the variation coefficient of the objective value is a suitable signal for stopping. The optimal stopping point is marked with a black dotted line, the actual stopping point with a red one.

more generic criteria may be established. Algorithm internal criteria, as, e. g., integrated into CMA-ES, focus on a concentrated population and the convergence to a single optimum, which is reflected by a fading mutation strength. Thus, the challenges of multimodal optimization are not properly reflected within the existing criteria which focus on stagnation related to desired convergence.

*Stopping Criteria for Multimodal Optimization (SMMO):* In contrast to the requirements for single- and multiobjective stopping criteria, we cannot directly observe the measures we are actually interested in (there: best objective value and hypervolume, respectively). In order to find a signal that can be exploited as stopping criterion due to its correlation to the course of the AHD indicator, we have investigated a large number of time series by visual inspection, e. g., the best and average objective values, the standard deviation of the objective values, the average mutation strength, its standard deviation, the *coefficients of variation* (standard deviation divided by mean value, CV) of objective values and mutation strengths, and diversity indicators [3].

We found two signals that appear to be useful. For the self-adaptive case, the mutation strength on average (Fig. 2) experiences a peak when the AHD reaches its minimum. In most cases, this behavior can also be found when looking at individual runs. A slightly less obvious correspondence that may be used for the fixed mutation strength algorithms exists between the CV of the objective values and the AHD. In many cases, the CV starts to decline when the AHD passes its minimum, as displayed in Fig. 3. As the raw signal shows fluctuations in both cases, we employ the window median $\tilde{x}_w(t) = \text{median}(x_{(t-w+1):t})$, where $x_t$ is a time series and $t$ runs from 1 to $t_{\max}$, over the median mutation strength and the CV of the objective values, respectively. After some initial experimentation, we chose $w = 10$ as window size. From that, we compute the window median $\tilde{x}'_w(t)$ of the forward differences of $\tilde{x}_w(t)$ in order to find the point in time when the original value is decreasing considerably. We stop as soon as the median of

**Table 1.** Factors for the experiment in §5

| Factor | Type | Symbol | Levels |
|---|---|---|---|
| Problem topologies | environmental | | {random, linear, funnel} |
| Number of variables | environmental | $n$ | $\{2, 3, 5, 10\}$ |
| Number of peaks | environmental | $N$ | $\{5, 20, 40\}$ |
| Selection variants | control | | {SV4, SV7} |
| Mutation strength | control | | {fixed, self-adaptive} |

**Table 2.** Differences between the optimal (w.r.t. AHD) stop generation and the suggested one as well as the percentages of generations after *StopGen* with higher AHD

| | | | StopGenAHD − StopGen | | | % higher AHD after | | |
|---|---|---|---|---|---|---|---|---|
| | Strategy | Criterion | LQ | Median | UQ | LQ | Median | UQ |
| SV4 | SA | MutStrength | −6 | 1 | 21 | 97.1 | 99.4 | 99.9 |
| | NonSA | VarCoeffObj | −38 | −11 | 95 | 67.9 | 95.1 | 98.9 |
| SV7 | SA | MutStrength | −57 | 351 | 680.5 | 14.5 | 35.1 | 61.2 |
| | NonSA | VarCoeffObj | 304.8 | 568 | 784.2 | 0.9 | 12.1 | 41.2 |

differences, $\tilde{x}'_w(t) = \tilde{x}_w(\tilde{x}_w(t) - \tilde{x}_w(t-1))$, gets negative for the first time (the first $w$ time steps are ignored).

## 5 Experimental Evaluation of SMMO

**Research Question:** Do the stopping criteria of §4 provide a reasonable performance?
**Pre-experimental Planning:** The stopping criteria in §4 were selected after a first visual inspection of several summary characteristics. After some preliminary investigations, we decided to test the CV-based criterion only with fixed $\sigma$ as the mutation strength criterion seemed superior (when available).
**Task:** The task of the stopping criteria is to abort the runs early with as few loss of performance as possible. The key criterion for us is the AHD indicator.
**Setup:** The bulk of the setup was already described in §3. Table 1 summarizes all the factors for this full-factorial experiment. For each configuration, five random test instances are drawn and five independent algorithm runs are carried out per problem instance, leading to a total of 25 repeats per configuration.
**Results:** Figures 4 and 5 show how much worse the obtained AHD values are for early stopping in comparison to the best value of the same algorithm run that would be obtained sometime during the full $10^5$ function evaluations. Table 2 contains another investigation of the same data, focusing on the differences of the actual stop generation and the respective one with minimum AHD value. Furthermore, the percentage of generations after the stop generation where the obtained AHD value was higher than the one in the stop generation is provided.

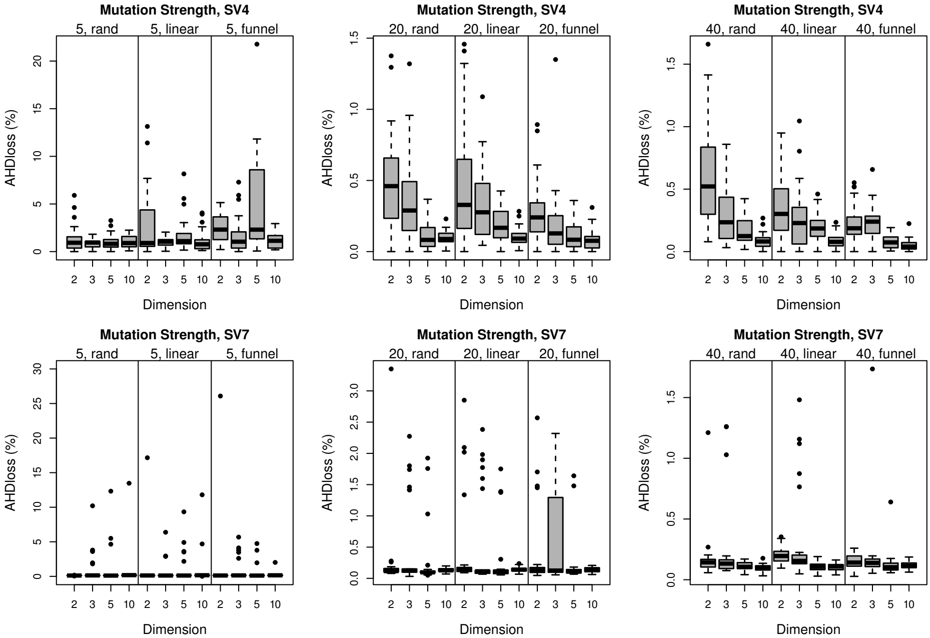


**Fig. 4.** Performance losses through the stopping criterion based on $\sigma$

**Observations:** For SV4, the loss of AHD performance is decreasing with increasing $n$ if $N$ is 20 or 40. SV7 shows the opposite behavior, especially with the CV-based criterion. If the problem contains only five peaks, the AHD loss is generally higher, especially with the $\sigma$-criterion (although the absolute values may still be better than with the CV). Table 2 reflects that the recommended stop generation does not differ much from the respective one with minimum AHD for SV4. Moreover, an almost neglectable percentage of obtained AHD values after stopping results in smaller AHD. SV7 shows a different behavior, the interquartile ranges of both statistics are relatively large and the median levels differ quite much from the respective ones of SV4.

**Discussion:** The seemingly worse performance with five peaks may occur because these problems are relatively easy and therefore the obtained AHD values are close to zero. So, even small absolute deviations appear as high relative deviations. On SV7 the losses are smaller, which is probably because the AHD values are generally fluctuating less. For SV7, the statistics in Table 2 reflect that the AHD quality usually does not show an obvious decreasing tendency after StopGenAHD but rather a fluctuating behavior around the minimum AHD. Applied to SV4, the suggested stopping criteria successfully detect an adequate stopping generation in the vicinity of StopGenAHD.

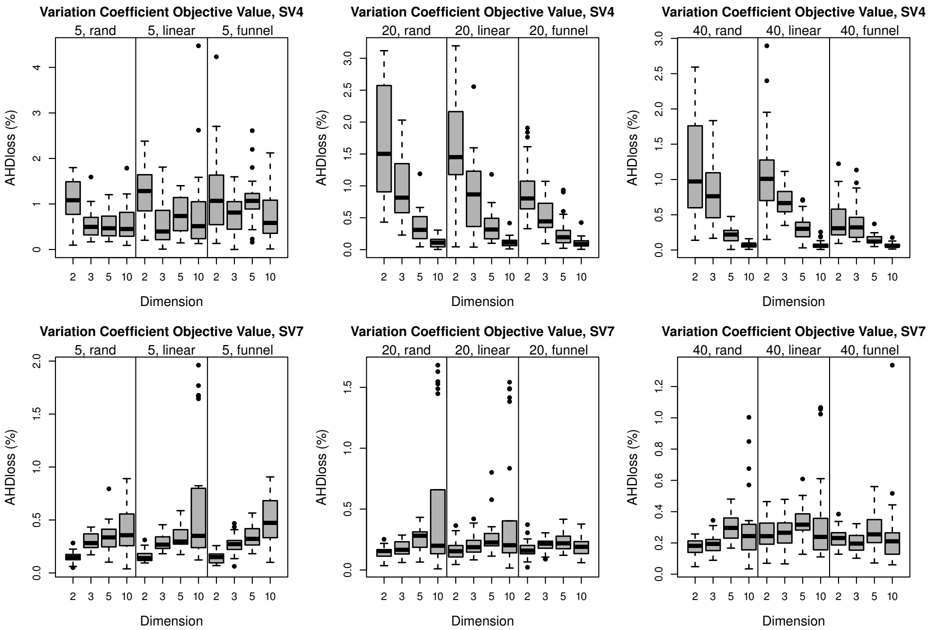


**Fig. 5.** Performance losses through the criterion based on the variation coefficient

## 6 Conclusions and Outlook

By means of systematic experiments we are able to show that transition phases between maintaining diversity and converging to single optima exist. While this is intuitive for classical evolution strategies, this effect also can be observed for strategies which explicitly address multimodality as SV4. Structural differences compared to SV7 are present, for which a large percentage of local optima are successfully approximated during the whole algorithm run due to extensive exploration of the search space.

Decreasing AHD between the set of local optima and the population coincides with increasing approximation quality in the multimodal setting. Indicators based on the mutation strength (self-adaptive strategies) or the variation coefficient of objective values (fixed step sizes) could be set up which appropriately reflect the AHD behavior over time which is naturally unknown within the actual algorithm run. The suggested stopping criteria, in most cases, recommended stopping generations which simultaneously ensure the coverage of the modes as well as sufficient proximity to the latter. However, they face greater challenges for decreasing number of modes but improve for increasing search space dimensionality for the mutation strength criterion.

In future work, we will explicitly analyze the influence of self-adaption of the mutation strength on algorithm performance. Moreover, the suggested stopping criteria will be included into more sophisticated MMO algorithms.

**Acknowledgments.** Heike Trautmann and Mike Preuss acknowledge support by the European Research Center for Information Systems (ERCIS).

## References

1. Das, S., Maity, S., Qu, B.Y., Suganthan, P.N.: Real-parameter evolutionary multimodal optimization – a survey of the state-of-the-art. Swarm and Evolutionary Computation 1(2), 71–88 (2011)
2. Wessing, S., Preuss, M., Rudolph, G.: Niching by multiobjectivization with neighbor information: Trade-offs and benefits. In: IEEE Congress on Evolutionary Computation (CEC), pp. 103–110 (2013)
3. Preuss, M., Wessing, S.: Measuring multimodal optimization solution sets with a view to multiobjective techniques. In: Emmerich, M., Deutz, A., Schütze, O., Bäck, T., Tantar, E., Tantar, A.A., Moral, P.D., Legrand, P., Bouvry, P., Coello, C.A. (eds.) EVOLVE – A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation IV. AISC, vol. 227, pp. 123–137. Springer, Heidelberg (2013)
4. Schütze, O., Esquivel, X., Lara, A., Coello Coello, C.A.: Using the averaged Hausdorff distance as a performance measure in evolutionary multiobjective optimization. IEEE Transactions on Evolutionary Computation 16(4), 504–522 (2012)
5. Stoean, C., Preuss, M., Stoean, R., Dumitrescu, D.: Multimodal optimization by means of a topological species conservation algorithm. IEEE Transactions on Evolutionary Computation 14(6), 842–864 (2010)
6. Beyer, H.G., Schwefel, H.P.: Evolution strategies – a comprehensive introduction. Natural Computing 1(1), 3–52 (2002)
7. Tran, T.D., Brockhoff, D., Derbel, B.: Multiobjectivization with NSGA-II on the noiseless BBOB testbed. In: Proceeding of the Fifteenth Annual Conference companion on Genetic and Evolutionary Computation Conference Companion, GECCO 2013 Companion, pp. 1217–1224. ACM (2013)
8. Wagner, T., Trautmann, H., Martí, L.: A taxonomy of online stopping criteria for multi-objective evolutionary algorithms. In: Takahashi, R.H.C., Deb, K., Wanner, E.F., Greco, S. (eds.) EMO 2011. LNCS, vol. 6576, pp. 16–30. Springer, Heidelberg (2011)
9. Trautmann, H., Wagner, T., Naujoks, B., Preuss, M., Mehnen, J.: Statistical methods for convergence detection of multi-objective evolutionary algorithms. Evolutionary Computation Journal 17(4), 493–509 (2009)
10. Hernandez, G., Wilder, K., Nino, F., Garcia, J.: Towards a self-stopping evolutionary algorithm using coupling from the past. In: GECCO 2005: Proceedings of the 2005 Conference on Genetic and Evolutionary Computation, pp. 615–620. ACM (2005)
11. Safe, M., Carballido, J.A., Ponzoni, I., Brignole, N.B.: On stopping criteria for genetic algorithms. In: Bazzan, A.L.C., Labidi, S. (eds.) SBIA 2004. LNCS (LNAI), vol. 3171, pp. 405–413. Springer, Heidelberg (2004)
12. Zielinski, K., Laur, R.: Stopping criteria for a constrained single-objective particle swarm optimization algorithm. Informatica 31(1), 51–59 (2007)
13. Hoos, H.H., Stützle, T.: Stochastic Local Search – Foundations and Applications. Morgan Kaufmann, San Francisco (2004)
14. Sastry, K.: Single and multiobjective genetic algorithm toolbox for Matlab in C++. Technical Report 2007017, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign (2007)