Detecting change in dynamic fitness landscapes

Hendrik Richter

Abstract—Change detection enables an evolutionary algorithm operating in a dynamic environment to respond to undertaking necessary steps for maintaining its performance. We consider two major types of change detection, population-based and sensor-based. For population-based we show its relation to statistical hypothesis testing and analyze it using receiver-operating characteristics. For sensor-based the relationship between detection success and number of employed sensors is studied and the dimensionality problem is addressed. Finally, we discuss how both types of change detection compare to each other.

I. INTRODUCTION

Evolutionary algorithms (EAs) are considered to be a good choice for solving dynamic optimization problems (DOPs) [3], [15], [9]. Contrary to static problems, in dynamic optimization locations and values of the optima change with time, rendering the problem of optimum finding to optimum tracking. Likewise, the static fitness landscape consisting of fixed search space and fitness function becomes a dynamic fitness landscape with possibly both search space and fitness function being time-dependent. Studies revealed that certain modifications to the standard algorithmic structure are needed to make EAs fit to perform well in dynamic environments, particularly to undertake steps for maintaining and enhancing diversity, for instance hypermutation, random immigrants or different types of memory [14], [17], [18], [20], [19]. A considerable number of these modifications rely on detecting the points in time where a change in the fitness landscape occurs. In most cases, it was either assumed that the change points are known or that they can be easily deduced by simply monitoring the performance of (possibly only the best of) the population and noticing a sharp decline. However, if these assumptions do not hold, change detection in dynamic fitness landscapes requires additional discussion. As change detection is frequently treated as a topic that supports and enables the studied modification of the EA, it is of considerable importance and has been addressed, at least, in a multitude of research studies on solving DOPs, not only by EAs, (see e.g. [3], p. 28–29, [19], [18] and references cited there), but also in the related field of particle swarm optimization (PSO) [7], [8]. On the other hand, we are missing studies reviewing and comparing the different detection schemes, particularly for dynamic fitness landscapes that allow no easy detection. This paper intends to give such a study by considering different kinds of change detection schemes that are tested in dynamic fitness landscapes with adjustable detection difficulty.

Change detection makes use of information, most prominently (but possibly not only) the fitness values of points in the search space, extracted from the fitness landscape. In principle, there are two ways to do so. One is to use the fitness evaluations of the EA's population, which we will call population-based detection, the other is to use additional measurement of the landscape’s fitness on prescribed points [14], which is called sensor-based detection. Other strategies which have been suggested in the context of dynamic particle swarm optimization [7] and work by recomputing the fitness of the best-of-previous-generation can be regarded as a mix of both ideas as reevaluating previous best solutions can be seen as a kind of sensor-based change detection with the sensor locations dynamically allocated by previous bests.

The remainder of the paper is organized as follows. In Section II we review properties of dynamic fitness landscapes relevant to change detection and give a modification of the well-known “moving peak” benchmark that allows to adjust the detection difficulty. In Section III change detection methods are considered. It is shown how population-based detection is related to statistical hypothesis testing and can be tackled within that framework. Also, sensor-based detection is discussed in detail. Experimental results are presented in Section IV where we employ different test statistics and use receiver-operating characteristics (ROC) as an analyzing tool. Also, we study how sensor-based detection scales with landscapes’ dimension. We end up with comparing the schemes, drawing conclusions and pointing at further problems.

II. PROPERTIES OF DYNAMIC FITNESS LANDSCAPES

Fitness landscapes are an important tool for studying evolutionary computation from a theoretical point of view. A fitness landscape assigns fitness values to all the points in the search space, for which a distance metrics and neighborhood structure needs to be defined. If this assignment remains constant for the run–time of the EA, we have a static fitness landscape. For dynamic fitness landscapes, the fitness values of all or some points in the search space are changing with the run–time. To define a dynamic fitness landscape, we can employ a dynamic fitness function

\[ f(x, k), \quad k \geq 0 \]  

where \( x \in M \) are elements of a fixed bounded search space \( M \subset \mathbb{R}^n \) and \( k \in \mathbb{N}_0 \) is the discrete time variable. The result of combining this search space with the dynamic fitness function can be regarded as dynamic fitness landscape. The task
of finding the maximal \( x_S \in M \) with \( f(x_S, k) \geq f(x, k) \), \( \forall x \in M \) for every \( k \geq 0 \) defines the DOP
\[
f_S(k) = \max_{x \in M} f(x, k), \quad k \geq 0
\] (2)

and its solution trajectory \( x_S(k) = \arg \max_{x \in M} f(x, k), \quad k \geq 0 \). Solving the DOP means to find the solution trajectory \( x_S(k) \) and the sequence of the (temporarily) highest point in the landscape \( f_S(k) \), respectively, by using an EA. Based on this, we can define a change point \( k_{cp} \in \mathbb{N}_0 \) in the dynamic fitness landscape by \( \exists x \in M \) for which
\[
f(x, k_{cp}) \neq f(x, k_{cp} + 1).
\] (3)

As changes in the fitness landscape (1) are usually measured relatively to the run–time of the EA, there is a strong connection between the problem (2) and the EA, which we therefore need to define to complete the overall description. The EA we consider here has a real number representation and \( \lambda \) individuals \( x_j \in \mathbb{R}^n, \ j = 1, 2, \ldots, \lambda \), which build the population \( P \in \mathbb{R}^{n \times \lambda} \). Its dynamics is described by the generation transition function \( \psi : \mathbb{R}^{n \times \lambda} \rightarrow \mathbb{R}^{n \times \lambda} \), see e.g. [1], p. 64–65, which can be interpreted as a nonlinear probabilistic dynamical system that maps \( P(t) \) onto \( P(t+1) \) and hence transforms a population at generation \( t \in \mathbb{N}_0 \) into a population at generation \( t + 1 \),
\[
P(t + 1) = \psi(P(t)), \quad t \geq 0.
\] (4)

Starting from an initial population \( P(0) \), eq. (4) describes the population dynamics in the search space. The time scales \( t \) and \( k \) are related by the change frequency \( \gamma \in \mathbb{N} \) as \( t = \gamma k \) with \( \gamma \) being constant.

Our interest now is how changes in the fitness landscape (1) can be detected. Generally, we need to assume that the change pattern is not analytically known and we have to rely on information extracted numerically from eq. (1) on a sample base. Obviously, a decline of the population’s performance, measured over some time window, is a straightforward indicator for a change in the dynamic fitness landscape. However, there is some subtlety. A basic assumption for this indicator to work is that the population’s fitness is affected by the change and moreover shows a decrease. This may be given in a fitness landscape that is dynamically isometric, that is the sum over all optima in the search space is constant over time. In other words these fitness landscapes are a kind of dynamic “zero–sum game”, where a change in fitness that is gained in one area has to be lost in another. For such dynamic fitness landscapes finding all optima in between changes by the EA implies successful change detection. Dynamic isometry in particular includes the case where the optimum at time \( k \) has a lower fitness at \( k + 1 \), that is \( f(x_S(k), k) > f(x_S(k), k + 1) \). That dynamic isometry is not necessarily a guarantee for easy change detection using the population’s fitness can be seen by the example of small needles on an otherwise constant plain that grow and shrink alternatively and/or move slightly. Only if the static “needle–in–the–haystack” problem has been solved, that is if at least one individual found the needle, a change using fitness information from the population can be detected. A similar line of thought yields that the changes in the dynamic fitness landscape have to exceed a certain severity so that they can be detected. Clearly, in a practical context we are mainly interested in detecting changes that are in principle detectable, i.e. the fitness landscapes are dynamically distinguishable after each change, meaning that the number of point \( x \) for which condition (3) applies must be sufficiently large.

In the following, we will consider dynamic fitness landscapes for which hardness of change detection is between these two extremes, that is very easy by simply observing the decline of the population’s performance and very complicated as some hard–to–fulfil preconditions must be given. Therefore, we will use and modify a frequently used dynamic fitness function, an \( n \)–dimensional “field of moving cones on a zero plane”, where \( N \) cones’ coordinates \( c_i(k), \ i = 1, 2, \ldots, N \) may change with time and have randomly chosen initial coordinates \( c_i(0) \), heights \( h_i \) and slopes \( s_i \). So, the dynamic fitness function is
\[
f(x, k) = \max \left\{ 0, \ \max_{1 \leq i \leq N} \left[ h_i - s_i \|x - c_i(k)\| \right] \right\}.
\] (5)

A dynamic fitness landscape constructed in this way separates the topology of the fitness landscape from its dynamics. The dynamics is given by the sequence of \( c(k) \) and has to be specified by an external source. That might be a precalculated route, for instance a circle or track or realizations of a random process or a trajectory of a (possible chaotic) dynamical system.

Based on these settings, we modify eq. (5) to include a varying degree of hardness to the change detection process. It is generally understood and intuitively plausible that hardness in change detection scales with the number of points in the landscape whose fitness is affected by the change and the diversity of the probe that is taken, see. e.g. [15], p. 15–17. Whereas the diversity of the probe is an issue of designing the change detection test and will be considered later, the number of affected points is a property of the dynamic fitness landscape. Here, we model this property by the number of cones out of the maximal \( N \) that actually move. We define a moving rate \( \chi \in \mathbb{R}, \) with \( 0 < \chi \leq 1 \) and set
\[
c_i(k) = c_i(0), \quad 1 \leq i \leq N - \lceil \chi N \rceil, \quad \forall k.
\] (6)

So, for \( \chi \rightarrow 0 \), we define that no cone moves (in that we solve a static problem for which change detection is futile, of course), for \( \chi = 1 \), all \( N \) cones move. For search spaces with fixed bounds, all cones lying within those bounds for all times \( k \) and the heights \( h_i \) and slopes \( s_i \) being realizations of a uniformly distributed random variable, the moving rate at least approximately reflects the number of points affected by changes.

III. CHANGE DETECTION METHODS

After defining the dynamic fitness landscape we now approach the problem of the change detection, which as has been mentioned before is based on numerical samples from
the dynamic fitness function \( f(x,k) \) for the population. With these samples we create a data set \( S(t) \) in every generation \( t \). By comparing the data sets \( S(t) \) and \( S(t+1) \), we intend to decide if a change has occurred in the fitness landscape at \( k_{cp} = \lfloor \gamma^{-1}(t+1) \rfloor \). We next consider how to create and evaluate the data set \( S(t) \) in order to do the detection test. In the given setting of solving a DOP by an EA the data set, in principle, can come from two sources: fitness function evaluations and metrics on the population, e.g. diversity. It is immediately clear that the values of the dynamic fitness function carry most of the information about changes, while information about the population can have auxiliary meaning at best. This given, fitness function evaluations, in turn, can now be carried out for either the individuals of the population or for arbitrary points, called sensors, in the search space. Both alternatives are considered in detail in the following.

Finally, a word on which fitness values to use in terms of generational time. If the full fitness distribution is used in change detection tests, it is not realistic to assume that it should be stored for a larger number of generations as usually there is a need to budget computational resources, particularly storage space. Obviously, these restrictions could be loosened if only a fraction is stored. However, this would imply that the question of which fitness values should be selected has been answered satisfactorily. If the answer is easily given by the best-in-generation, change detection can be done by the straightforward methods considered so far. As the dynamic fitness landscapes considered in this paper do not guarantee these characteristics, the usage of fractions and generational overlaps of fitness values for change detection proposes is not considered here. Our requirements on the test is that it works with a data stream, that is we want to use only the fitness values from two consecutive generations.

A. Population-based detection: Statistical hypothesis testing

In solving the DOP (2) by an EA, the individuals \( x_j(t) \) of the population \( P(t) \) change their position from generation to generation, hence move in the dynamic fitness landscape (1), follow the changing optima and so somehow model the dynamics of the change. For population-based change detection, this is used as we evaluate the fitness of these individuals. Hence, the data set \( S(t) \) consists of the items \( s_j(t) = f(x_j(t),k) \) for \( k \in \lfloor \gamma^{-1}t \rfloor \).

Statistically speaking, the considered data set \( S(t) \) can be regarded as coming from an unknown distribution \( P(t) \). This transforms the problem of change detection into the problem of testing whether the data sets \( S(t) \) and \( S(t+1) \) coming from different distributions, that is \( P(t) \neq P(t+1) \), which is known as statistical hypothesis testing. This connection is widely applied in solving change detection problems, e.g. [2], [11], [4]. The obvious question here is which test can tell us whether \( P(t) \) is different from \( P(t+1) \) and if this difference necessarily and sufficiently implies that a change has occurred. In the language of statistical hypothesis testing, the test should ideally show only true changes, that is have no false positives and indicate all of them, that is have no false negatives.

Based on these remarks suitable test statistics can be specified. The data set \( S(t) \) including the fitness value of an evolving population represents two types of interfering population dynamics. A first is that an evolving population changes its mean and standard deviation (ideally and in the best case) monotonically towards the optima. Such a convergence behavior of the EA which is desired and the intended working mode is again a statistical phenomenon. A second is the reaction of a change in the fitness landscape. The statistical test for change detection needs to distinguish between the both of them. Further, in a multimodal dynamic optimization problem as the one posed here there are several optima and the population might split into groups pursuing different ones. So, the fitness distribution in \( S(t) \) is most likely not in a certain parametric form (for instance a normal distribution). This fact rules out change detection schemes based on standard hypothesis testing such as CUSUM or SPRT as they require the distributions \( P(t) \) to have a parametric form, see e.g. [2], p. 35–40. In the numerical experiments, we will therefore use tests that make no assumptions on the form of the distribution. These are two non-parametric statistical tests, one the Kolmogorov–Smirnov test, another the Wilcoxon–Mann–Whitney test (see e.g. [10], p. 68-86 and also [11]) and in addition a test based on an entropic measure from information theory, the Jensen–Shannon distance [13], [6]. Population-based change detection has the advantage that it requires no additional fitness function evaluations. On the other hand, its success depends on how well the movement of individuals models the fitness landscape. That is the suitability of \( S(t) \) for change detection is influenced by the problem solving qualities of the used EA.

B. Sensor-based detection: Additional landscape measurements

Another method for change detection is to carry out additional measurements in the dynamic fitness landscape using so-called “fitness landscape sensors” [14], that are predefined, remain fixed during the landscape changes, and are placed either randomly or on a regular grid or elsewhere based on some placement strategy. If any of the sensors detects an altered fitness value, we know that change has occurred. There is no need for elaborated statistical analysis and there can be no false positives. However, we need to compute additional fitness function evaluations. The problem in this scheme is what principles should be employed in placing the sensors and how many of them are needed to reliably detect changes.

The answer to the question of whether a random or a regular grid or an in some other way designed placement of the sensors is to prefer depends highly on the topology and the change pattern of the dynamic fitness landscape. If the optima and their changes occur in all search space regions and for all times with equal probability, a uniform or regular grid-based distribution is sensible to chose. Is the spatial or temporal likelihood for optima change significantly higher for a subset of the search space, a significantly higher
number of sensors should be situated there. However, in the absence of any further information, no preference can be justified. Therefore, in accordance with the fitness landscape considered here, normally spatial distribution of sensors will be employed. The questions of how many are needed has not been studied so far. Here, not only the dependency between sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for the given landscape. Here, not only the dependency between sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for a given landscape is of interest, but also the dimensionality of the sensor number, detection difficulty and detection success for.

IV. EXPERIMENTAL RESULTS

In the following we report numerical experiments with the change detection schemes described above. In the experiments, we use a dynamic fitness landscape \( S(t) \), for which the detection difficulty can be adjusted by the moving rate \( \chi \) defined by (6); the number of cones is \( N = 50 \), for population–based detection we use a landscape with dimension \( n = 2 \), for sensor–based detection we will vary the dimension. The dynamics of the moving sequence for the cones’ coordinates is normally random, that is each \( c_i(k) \) for each \( k \) is an independent realization of a normally distributed random variable. Also for population–based detection, we employ an EA with a fixed number of \( \lambda = 48 \) individuals that uses tournament selection of tournament size \( 2 \), a fitness–related intermediate sexual recombination (which is operated \( \lambda \) times and works by choosing two individuals randomly to produce an offspring that is the fitness–weighted arithmetic mean of both parents) and a standard mutation with the mutation rate 0.1. Note that the choice of the EA is a secondary concern as long as it solves the DOP with some success.

A. Population–based detection

As discussed before, the change detection test follows the procedure of traditional statistical hypothesis testing, which is threefold, consisting of the null hypothesis, the test statistics and a critical region defined as lying above a threshold value. The null hypothesis reflects our assumption about the distribution \( P(t) \), the test statistic is a function computed from the samples \( S(t) \) and \( S(t + 1) \) producing a critical value. If this value exceeds the threshold value, it tells us to reject the null hypothesis, meaning that \( P(t) \neq P(t + 1) \) with some probability. In population–based change detection the null hypothesis corresponds to our assumption about the fitness distributions from generation to generation if there is no change in the landscape. Note that this implies some “natural change” in the distribution as described by the population dynamics (4) which the test statistics computed from samples of these fitness distributions have to be robust against. That is the test statistics must classify \( P(t) = P(t + 1) \) as long as it is a result of the convergence behavior of the EA, but \( P(t) \neq P(t + 1) \) if a change in the fitness landscape has occurred. The test statistics are the considered tests (Wilcoxon–Mann–Whitney, Kolmogorov–Smirnov and Jensen–Shannon) which also define their threshold values. The analysis presented here uses the techniques of receiver–operating characteristics (ROC) curves, which, for instance, is widely used in change detection in biomedicine, data mining and image processing [12], [16], [5]. ROC curves are a tool for organizing and visualizing classifications together with their performances. So, they can be used to analyze and depict the relative tradeoffs between benefits of the schemes (correctly identified instances according to the classification) and costs (incorrect identifications). That makes them particularly useful to assess change detection schemes. The classification here is between positive and negative change detections. Hence, we can define the following performance metrics. If we have a positive detection and a change in the fitness landscape has happened, it is counted as true positive \( (tp) \), if a change happened but is not detected, it is a negative positive \( (np) \). If, on the other hand, no change has happened and the detection is negative, it is a true negative \( (tn) \). A positive detection in this situation yields a false negative \( (fn) \). For this two–by–two change classification, we obtain as the elements of performance metrics: the \( tp \) rate

\[
tp \text{ rate} \approx \frac{\text{correctly identified changes}}{\text{total changes}}
\]

and \( fn \text{ rate} = 1 - tp \text{ rate} \) as well as the \( fp \) rate

\[
fp \text{ rate} \approx \frac{\text{incorrectly identified changes}}{\text{total non changes}}
\]

and \( tn \text{ rate} = 1 - fp \text{ rate} \). In the ROC plot, the \( tp \) rate is given (on the ordinate) versus the \( fp \) rate (on the abscissa). Hence, the \( tp \) and \( fp \) rates for the test statistics ( Wilcoxon–Mann–Whitney, Kolmogorov–Smirnov and Jensen–Shannon) for a given threshold value give a point in the ROC space, ROC curves are obtained by plotting the rates for varying the threshold values.

Figs. 1a,b and 2 show the ROC curves for different moving rates \( \chi \) and the three different test statistics. The \( tp \) and \( fp \) rates given here were calculated as the means over \( R = 200 \) runs for \( T = 150 \) generations; also the 95% confidence intervals are shown for the \( tp \) rate. For the \( fp \) rate the intervals are significantly smaller and therefore not depicted here. However, note that the calculated confidence intervals are to be taken with some care as strictly speaking they only apply to the statistical tests being used for independent samples, which the fitness distributions in between changes are probably not, but the fitness distributions before and after a change most likely are. From these figures, the following conclusions can be drawn by considering several important features within the graph. In general, the lower left point \((0,0)\) represents a change detection that never produces a positive decision. It makes neither a false positive error nor yields any true positives. Likewise but opposite, a detection represented by the upper right point \((1,1)\) only produces positive decisions with only true positives but also false positive errors in all cases. The line between these two points in the ROC space can be regarded as expressing a purely random guessing strategy to decide on whether or not a change has happened. For instance, such a random guessing might be right half of the cases where a change truly happened and also wrong in half of the cases where actually
no change happened, which yields the point \((0.5, 0.5)\) in the ROC space. By relaxing or tightening of the classification threshold, the number of true positives can be lowered or raised but at the cost of having adjusted the number of false positives in the same way. Hence, random classification of change detection yields points on the diagonal between \((0, 0)\) to \((1, 1)\) in ROC. Any classification that is represented by a point below that line is worse than random guessing, while classification above is better, the more so if one point is more to the north–west of another, with the point \((0, 1)\) expressing perfect classification. With these relationships in mind, the following interpretation of the experimental results can be given.

The ROC curves for Wilcoxon–Mann–Whitney test statistic for different moving rates \(\chi\) (representing different detection difficulty) are shown for the change frequency \(\gamma = 10\) in Fig. 1a and for \(\gamma = 20\) in Fig. 1b. Likewise, the results for Jensen–Shannon are given in Fig. 2a,b and for Kolmogorov–Smirnov in Fig. 2c,d. Best detection results are achieved for the Wilcoxon–Mann–Whitney test for which graphs are obtained that climb from the point \((0, 0)\) vertically for a considerable amount of threshold values towards \((0, 1)\) before bending off to \((1, 1)\). The different moving rates \((\chi = 1, 0.8, 0.6, 0.4)\) give clearly distinctive curves, where the most difficult change detection \(\chi = 0.4\) has the lowest curves, meaning that the least best change detection is achieved. The results for the different change frequencies are slightly different only, with change detection a little better for \(\gamma\) being larger. The reason for this feature is that success in change detection mildly correlates with success in solving the DOP. As for \(\gamma\) becoming larger, the performance in solving the DOP is usually higher and so also change detection gets better. However, this relation is of secondary importance, as to be seen in Fig. 1c, where the mean fitness error (MFE) in \%\) (which quantifies performance in solving the DOP) is shown in a scatter plot versus the \(tp\) rate for \(\gamma = 10\). Note that we are using an EA that does not directly and outside–triggered react on changes in the dynamic fitness landscape. Therefore, the performance, in particular for small change frequencies, is considerably less high. As for the \(fp\) and \(tp\) rates, the results here are means over \(R = 200\) runs for \(T = 150\) generations. We see that for constant \(\chi\) good and bad detection results are obtained for approximately the same MFE, whereas for a different \(\chi\) we obtain clearly separated
data clouds. In other words, for constant $\gamma$, there is only a weak correlation between the EA’s performance in solving the DOP and performance in change detection. A somehow similar picture can be seen by analyzing the relationship between diversity of the fitness distribution and change detection success, see Fig. 1d which gives the diversity (as sum of moments of inertia, see e.g. [15], p. 31) of the fitness distribution in the generation before the change (and in this way the relevant for calculating the test statistics) versus the $tp$ rate, again as scatter plot. It can be seen that pre-change diversity and $tp$ rate are only weakly correlated and the data clouds overlap for different $\chi$. Similar results have been obtained for post-change diversity. The results for Jensen–Shannon test statistic (Fig. 2a,b) show similar features but are generally inferior. For Kolmogorov–Smirnov (Fig. 2c,d) we obtain considerably poorer results, where for larger detection difficulty ($\chi = 0.4$) the results are only slightly better than purely random classification. The reason for the poor performance of Kolmogorov–Smirnov might be that it has large difficulties in distinguishing between ‘natural’ alterations in the fitness distributions (caused by the evolving population pursuing to solve the DOP) and alterations by changes in the fitness landscape. The relationship between $tp$ rate, MFE and diversity for Jensen–Shannon and Kolmogorov–Smirnov mimic closely the results obtained for Wilcoxon–Mann–Whitney (and are therefore not depicted here).

In the numerical results no experiments addressing the dimensionality problem are reported. The reason is that dimensionality is not directly an issue here. Population–based change detection is based on evaluating a one–dimensional fitness distribution from one generation to the next, which makes solving the associated DOP a subject to the dimensionality problem but not the change detection test in itself.

### B. Sensor–based detection

In sensor–based change detection additional fitness function evaluations at randomly selected but fixed–over–run–time positions in the search space are used. In the following we give experimental results on the question of how many sensors are needed for performing this task. Therefore, we define the number of sensors by $\# S$. If at least one of these sensors reports that its fitness value has become different from one generation to the next, change is detected and always correctly so. As the sensors are at fixed points in the search space a different fitness value guarantees a change in the landscape according to condition (3) and hence there can
be no false positives. On the other hand, a change may occur that none of the sensors becomes aware of. The percentage of the noticed changes defines the true positive rate.

In Fig. 3, the true positive rate $tp_{rate}$ expressing the probability of a correct change detection is given over the moving rate $\chi$ specifying the detection difficulty for different $#_S$. Again, $R = 200$ runs for $T = 150$ generations were taken into account; the 95% confidence intervals are given. The dynamic fitness landscape and its parameters are the same as in the experiments with population–based detection reported above. The locations of the sensors were set before each run using independent realizations of a random variable and remained fixed during the run. Fig. 3a shows the results for the landscape dimension $n = 2$. We see an exponential curve approaching the $tp_{rate} = 1$ line, indicating that all changes in the landscape are detected correctly. This line is met for all numbers of sensors $#_S$ as early as $\chi = 0.8$. Clearly, the largest numbers of sensors $#_S$ gives the best results. This remains essentially true for landscape dimension $n = 4$, see Fig. 3b, but the 100% correct detection is not reached for a number of sensors as small as $#_S = 8$; detection success deteriorates rapidly with increasing dimension.

To verify this observation experiments were carried out to obtain the $tp$ rate versus landscape dimension $n$, see Fig. 4, which shows the relationship on semi–logarithmic scale. We notice a curve that is parallel to the $tp_{rate} = 1$ line for a certain number of dimensions $n$ before falling linearly with dimension in the semi–logarithmic plot. By increasing the number of sensors $#_S$ the critical point of the bending off can be postponed, but the decline follows at similar rates. So, these data suggest a power law $tp_{rate} \propto #_S \cdot c^{-n}$, with $c > 0$, for dimensions above the critical value, which experimentally confirms the theoretical results given in [15], p. 16. For higher dimensions, the different moving rates $\chi$ have only little influence (mainly how steep the $tp$ rate falls shortly after the critical bending–off–value) as the dynamic fitness landscape’s dimension is a much more prominent factor in detection success.

V. DISCUSSION AND CONCLUSIONS

With population–based and sensor–based two major types of scheme for detecting change in dynamic fitness landscapes have been considered. The advantage of population–based is that no additional fitness function evaluations are required, sensor–based can forgo elaborated statistical procedures. A comparison of the methods for change detection relies heavily on comparing the computational effort for doing the statistical tests with doing fitness function evaluations. Clearly, there cannot be an unambiguous answer. However, the numerical results suggest the conjecture that the more difficult the change detection is, the more favorable is sensor–based detection. Besides, only sensor–based detection allows robust 100% detection, if a sufficiently high number of sensors is employed.

For the change detection schemes extensions are possible. One way is combining both methods, maybe doing additional sensor evaluations if the population–based test was inconclusive or untrustworthy. Another improvement might be to compare the combined fitness distributions of several past generations to the present one. The test statistics considered here allow to compare samples of different size. That might be particularly useful to tell abrupt changes, for which severity is high from one generation to the next, from gradual changes, where severity has to accumulate over a certain time to be noticeable.

A further research topic in change detection is to use the schemes analyzed here to drive and trigger modification in the standard EA structure such as hyper–mutation, random immigrants or different types of memory. So, the significance of change detection for the performance of the EA can be studied. It should finally be noted that the results for population–based change detection have been obtained with an EA, but apply, at least in principle, to other population–based optimization schemes as PSO.

Fig. 3. Sensor–based change detection: True positive rate $tp$ over moving rate $\chi$ for different numbers of sensors $#_S$ and landscape dimensions (a) $n = 2$, (b) $n = 4$. 

2009 IEEE Congress on Evolutionary Computation (CEC 2009) 1619
Fig. 4. Sensor–based change detection: Dependency of the true positive rate $tp$ on landscape dimension $n$ in a semi–logarithmic plot (a) $\chi = 1.0$, (b) $\chi = 0.8$, (c) $\chi = 0.6$, (d) $\chi = 0.4$.

REFERENCES


