Termination Criteria in Evolutionary Algorithms: A Survey

Seyyedeh Newsha Ghoreishi, Anders Clausen and Bo Noerregaard Joergensen
Center for Energy Informatics, The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense M, Denmark

Keywords: Evolutionary Computation, Evolutionary Algorithm, Termination Criterion, Stopping Criterion, Convergence, Performance Indicator, Progress Indicator.

Abstract: Over the last decades, evolutionary algorithms have been extensively used to solve multi-objective optimization problems. However, the number of required function evaluations is not determined by nature of these algorithms which is often seen as a drawback. Therefore, a robust and reliable termination criterion is needed to stop the algorithm. There is a huge amount of knowledge encapsulated in the studies targeting termination criteria in evolutionary algorithms, but an updated integrated overview of this knowledge is missing. For this reason, we aim to conduct a systematic research through a comprehensive literature study. We extended the basic categorization of termination criteria to a more advanced one that takes the most common used termination criteria into consideration based on their specifications and the way they have been evolved over time. The survey is concluded by suggesting a road-map for future research directions.

1 INTRODUCTION

Evolutionary Algorithms (EAs) introduce a class of probabilistic optimization algorithms inspired by the principles of biological evolution which is known as one of the main approaches in the computational intelligence field (Konar, 2005). In general, the optimization process starts with a randomly selected population and iterates over number of generations by modifying the start population to obtain a near-optimal or possibly optimal solution. The optimization process runs until a given termination criterion triggers. Inherently, EAs are not able to decide about terminate of the optimization process and the ability of automatic termination is not designed in the original versions of EA. Therefore prespecified termination criteria should be considered either by the end-user or the application programmer.

Since the termination criteria behave differently for various EA, it is not possible to formulate a general rule for optimal use of a termination criterion (Jain et al., 2001). Therefore an automated termination is desired in real-world applications for practical reasons. The first reason is that the software end-users are not often aware of the behavior of the EA. The second reason is the necessity to find a suitable value for the parameters of the termination criteria to obtain a proper ending of the EA. In most termination criteria, suitable parameters can only be determined by a trial-and-error method (Jain et al., 2001).

Recently, the rapid growth of many complex applications in science, engineering and economics highlights the need of developing practical versions of EA. The original versions of EA suffer from two main weaknesses: unlearned termination and slow convergence. Much of the current literature pays particular attention to overcome these two issues by identifying and evaluating a new termination criterion or by modifying the standard versions of EA to accelerate the convergence. Even though termination criteria is widely used and investigated, but there are few theoretical guidelines for determining a suitable and proper point of time to terminate the search algorithm (Bhandari et al., 2012).

Up to now, two prominent studies have paid particular attention to integrate the known approaches used as termination criteria in EA (Jain et al., 2001) and (Wagner et al., 2011). For the first time in 2001, a comprehensive study done by Jain to classify termination criteria into three main categories: direct, derived and cluster-based termination criteria. Moreover, a new cluster-based termination criteria proposed and evaluated by comparing with other termination criteria in terms of reliability and performance (Jain et al., 2001). Finally, guidelines for the application of termination criteria provided (Jain et al., 2001). The

Ghoreishi S., Clausen A. and Joergensen B.
Termination Criteria in Evolutionary Algorithms: A Survey.
DOI: 10.5220/0006577903730384
In *Proceedings of the 9th International Joint Conference on Computational Intelligence (IJCCI 2017)*, pages 373-384
ISBN: 978-989-758-274-5

proposed taxonomy of termination criteria is used to establish formal guidelines and later has been considered as a basis for MATLAB toolbox implementation to provide a framework for analysis and evaluation of existing and new approaches.

The studies presented thus far provide evidence that selecting an appropriate termination criterion is a challenging task. Obviously, the new termination criteria are not necessarily fit to the basic three categories introduced by (Jain et al., 2001). Also, the main contributions of the work done by Wagner have no direct focus on providing an overview on the current existing approaches (Wagner et al., 2011). In this survey, we conducted a systematic research to identify the most commonly used termination criteria and grouped them into seven categories by extending the basic categories proposed by (Jain et al., 2001). Furthermore, we applied the proposed taxonomy of termination criteria to discuss more sophisticated approaches which are proposed so far.

The survey is organized in the following way. In next section, the study method used for writing the state-of-the-art is explained. Later, termination criteria are categorized based on different aspects and discussed in details in section 3. For each category, an overview on commonly used termination criteria is presented. In section 4, pros and cons of using each category of termination criteria are discussed in details. Section 5 concludes the key important notes investigated in the survey and propose a road-map for future research directions.

2 THE STUDY METHOD

The study method we used to review termination criteria is based on the systematic mapping studies and literature reviews suggested by(Kitchenham and Charters, 2007) and (Petersen et al., 2008). As it is guided in (Petersen et al., 2008) we performed five main steps in order to find the most relevant works targeting termination criteria in EA. These are: identifying the research questions, conduct search, screening of the papers, key-wording, data extraction and mapping process. In the following, we discuss each process step briefly.

2.1 Research Questions

The main goals of doing research on termination criteria are reflected by numbers of research questions. In our case, we try to answer the main question: *How can we organize an structure based on the existing contributions of termination criteria in EAs?* More research questions are raised to answer the main research question in different aspects.

- What are the most frequently investigated termination criteria and what are their main characteristics?
- How can we categorize them?
- What are the most frequently applied termination criteria in evolutionary algorithms and how they have been evolved over time?

2.2 Conduct Search

Two search strategies are suggested for performing a systematic research, manual and automatic (Kitchenham and Charters, 2007) and (Petersen et al., 2008). We chose to only use manual search to find the papers in the most relevant journals and proceedings of the related conferences in field of Evolutionary Computations (ECs) and Evolutionary Algorithms (EAs).

2.3 Screening of Papers

All of the papers found in section 2.2 are analyzed according to several selection criteria such as: index of the conference or journal, number of citations, long or short papers, technical or commercial papers and whether they are part of a book or conference proceedings. In this case, we included the materials from book chapters, the papers with high number of citations and the technical papers published by high index conferences and journals in the field of EC and EA. Screening process continues by two iterations: studying the abstracts and conclusion and a more detailed review by reading the introduction and methodology sections of the papers.

2.4 Keywording

In order to search for the right papers, we had to consider most relevant keywords. Three categories of keywords are used for this purpose. The first category is related to the keywords used to describe the word termination such as *termination criterion*, *stopping criterion* and *convergence*. The second category refers to performance evaluation of an EA such as *performance indicator* and *progress indicator*. The third category refers to the problem domain in which we are looking for termination criteria. They are mainly *evolutionary computation, computational intelligent, evolutionary algorithm, multi-objective optimization algorithms, multi-objective evolutionary algorithms and genetic algorithms*.

2.5 Data Extraction and Mapping Process

Data extraction is the last process of the study method to capture the relevant information from the selected papers to address the main research questions from section 2.1. In this step, the bibliography of the selected paper, the contribution, the methodology and application of the termination criteria are extracted. Later, extracted data is used to maintain the categorization of termination criteria.

3 TERMINATION CRITERIA IN EVOLUTIONARY ALGORITHMS

It is worthwhile to note that the concept of convergence of an EA is different from termination criterion even though they overlap each other. (Bhandari et al., 2012) emphasized that the proof of convergence of an algorithm to an optimal solution is very important as it guarantees the utility of an algorithm to reach to the optimal solution in infinite iterations. Different works have proved the convergence of genetic algorithm to an optimal solution after running for infinite iterations (Rudolph, 1994), (Suzuki, 1995), (Bhandari et al., 1996), and (Murthy et al., 1998).

More concisely, once the convergence of an algorithm is assured, then the focus turns into the determination of stopping criteria of the algorithm. For that reason, some of the termination criteria are defined using convergence phenomena. In this case, termination criterion is fulfilled if the algorithm is converged to an optimal or near-optimal solution(s). The closeness of the found solution and the optimal solution depends on the accuracy that the user desires. More details are provided for each termination criterion further in this section.

Thus far, one of the major challenges in the implementation of genetic algorithm is to define a proper termination criterion or criteria to stop the algorithm while no a-prior information regarding the objective function is provided (Bhandari et al., 2012).

In the following of this section, a number of termination criteria are presented and classified into seven main categories. For each criterion, description, main properties and a termination condition is given. For simplicity, notations and terminologies are introduced stepwise where needed together with the respective references.

3.1 Direct Termination Criteria

Direct termination criteria refer to the class of termination criteria which stop the algorithm if a predefined condition is satisfied without considering any underlying data from the evolutionary search process (Jain et al., 2001).

Maximal Time Budget. The maximal time budget criterion is fulfilled if the given time budget is consumed. The maximal time budget can be measured as an absolute time or the CPU-time (Jain et al., 2001). In this case, the algorithm runs for a predetermined execution time and returns the final solution.

Maximal Number of Generations. The maximal number of generations/iterations criterion is fulfilled if the algorithm has been running for a given maximum number of generations/iterations (Jain et al., 2001) and (Bhandari et al., 2012).

Maximal Number of Objective Function Evaluations. In (Jain et al., 2001), the maximal number of objective function evaluations is defined in the same way similar to 3.1 where the algorithm stops after reaching to a specific number of objective function evaluations.

Hitting a Bound. The hitting a bound termination criterion is fulfilled if the best value for the objective function is obtained for a given bound for an objective function (Jain et al., 2001). In this case, the found solution is supposed to be close enough or equal to the known global optimum (Hansen and Kern, 2004), (Zhong et al., 2004), (Tsai et al., 2004) and (Ong et al., 2006).

***K*-iterations.** The K-iterations termination criterion is fulfilled if there is no improvement in the best fitness values through a K number of consecutive iterations. The user has to select a proper value for K by assuming that it is impossible to obtain better result after K consecutive iterations (Leung and Wang, 2001) and (Bhandari et al., 2012).

3.2 Derived Termination Criteria

Contrary to direct termination criteria, derived termination criteria calculate auxiliary values using underlying data obtained from the current generation of the evolutionary search process (Jain et al., 2001). They have also been used as a measure of the state of convergence (Jain et al., 2001).

Running Mean. The running mean termination criterion is fulfilled if the difference between the best objective value of the current generation and the average of the best objective values of the last t_{last} generations is equal to or less than a given threshold $\varepsilon \geq 0$ (Jain et al., 2001).

Standard Deviation. The standard deviation termination criterion is fulfilled if the standard deviation of all objective values of the current generation is equal to or less than a given threshold $\varepsilon \geq 0$ (Jain et al., 2001).

Best-Worst. The best-worst termination criterion is fulfilled if the difference between the best and the worst objective value of the current generation is equal to or less than a given threshold $\varepsilon \geq 0$ (Jain et al., 2001).

Phi. The Phi termination criterion is fulfilled if the quotient of the best objective value and the mean of all objective values of the current generation is equal to or greater than a given threshold 1 - ε with $1 \gg \varepsilon \geq 0$ (Jain et al., 2001).

Kappa. The Kappa termination criterion is fulfilled if the quotient of sum of all normalized distances between all individuals of the current population and $k_{max} = \frac{\mu^2-\mu}{2}$ is equal to or less than a given threshold $\varepsilon \geq 0$ where μ is the number of individuals known as population size (Jain et al., 2001).

POP-Var. The POP-Var termination criterion is fulfilled if the variance of fitness values of all the individuals in the current population is equal to or less than a given threshold $1 \gg \varepsilon \geq 0$ (Bhandari et al., 2012).

ε-Variance. The ε-Variance termination criterion is an extended version of a termination criterion similar to the *K*-iterations and POP-Var termination criteria (Bhandari et al., 2012). This criterion considers the concept of elitism by preserving the fittest individuals over the generations. The ε-Variance termination criterion is fulfilled if the variance of the best solutions over generations is equal to or less than a given threshold ε with $1 \gg \varepsilon \geq 0$ (Bhandari et al., 2012).

3.3 Cluster-based Termination Criteria

Cluster-based termination criteria use clustering techniques to examine the distribution of individuals in the search space at a given generation. Individuals usually form clusters in the search space after a few generations of stagnation (Jain et al., 2001). The search process is terminated when the clusters show the convergence of the evolutionary search. In this case, the fittest individuals are concentrated in few small regions of the search space.

ClusTerm. The ClusTerm termination criterion is fulfilled if the change of the average of the aggregate size of elitist clusters averaged over the last t_{last} generations is equal to or less than a given threshold ε with $1 \gg \varepsilon \geq 0$ where t_{last} is the maximal number of the last generations to be considered (Jain et al., 2001).

3.4 Operator-based Termination Criteria

In EA, fine-tuning of parameters is normally done in an empirical way, and usually has a significant impact on their performance (Zapotecas Martínez et al., 2011).

Over the past two decades, there has been an increasing amount of literature on emphasizing the effects of genetic operators on the stopping time. One of the first systematic study was reported by Murthy et al. in 1998. The authors introduced a new termination criterion known as ε-Optimal Stopping Time. Based on this research, investigations are necessary to judge theoretically the effect of selection, mutation and crossover operators on the stopping time. Even though the obtained stopping times are valid for Elitism GA (EGA) with selection, crossover and mutation operators (Murthy et al., 1998). In this work, behavior of GA is studied with different values for the probability of mutation by calculating the hamming distance between the found solution and the optimal solution. The ε-Optimal Stopping Time termination criterion fulfills when the distance between the found solution and the optimal solution is equal to or less than a given threshold ε. Nevertheless, the concept of elitism helps the algorithm not to suffer from finding the sub-optimal solutions like ε-Variance termination criterion discussed earlier in section 3.2.

Later in 2000, Greenhalgh and Marshall discussed about the convergence properties for genetic algorithms by looking at the effect of mutation on convergence (Greenhalgh and Marshall, 2000). They showed that by running the genetic algorithm for a sufficiently long time, convergence to a global optimum with any specified level of confidence, is guaranteed. Experimental results showed that the algorithm is able to obtain an upper bound for the number of iterations necessary to ensure convergence, which improved the previous results.

More attention has focused on improving the evolutionary operators which go further than modifying their genetics to use the estimation of distribution algorithms in generating the offspring (Lee and Yao, 2004), (Lozano et al., 2006) and (Hedar et al., 2007). A recent study involved domain specific operators to enhance state-of-the-art MOEAs (Ghoreishi et al., 2015). This work presents an evolutionary algorithm CONTROLEUM-GA that applies domain specific variables and operators to solve a real dynamic MOEA for a greenhouse climate control problem. By that, the domain specific operators only encode existing knowledge about the environment. In CONTROLEUM-GA, domain specific operators are combined with two other direct termination criteria such Maximal Time budget and Maximal Number of Generations to terminate the search process. Experimental results shows improvements in convergence time without compromising the quality of the final solutions compared to other state-of-art algorithms.

3.5 Performance Indicator Termination Criteria

The criterion will be a local termination criterion if it is calculated by using properties of the solutions belong to the current generation, which we discussed earlier. Local criteria require the prior knowledge of optimal solutions to some extent, which may not be always available. On the other hand, global criteria are those that compute the progress of evolution through numbers of consecutive generations (Wagner et al., 2011). Based on the definition, short-term characteristics of the evolution can be drawn from local termination criteria, whereas global termination criteria show the long-term nature of the evolution. To design a better termination criteria, it is preferable to apply a global criteria that will not only terminate the evolution at the appropriate time, but also guarantee the quality of the solutions (Bhuvana and Aravindan, 2016b).

To measure the performance of the MOEAs, a standard procedure introduced by Coello using performance indicators (Coello et al., 2006). The most commonly used performance indicators are Hypervolume, Generational Distance, Inverted Generational Distance, Spacing, Additive-ε Indicator, Contribution or Set Coverage Metric, R1-Indicator, R2-Indicator, R3-Indicator and Maximum Pareto Front Error (Zitzler et al., 2003) and (Coello et al., 2006). Performance indicators use the concept of Pareto optimality to evaluate the performance of MOEA; the same measures can be used to check the convergence of the population. In this case, when the population is converged, evolution can be terminated. In the following, two categories of performance indicator termination criteria are discussed.

3.5.1 Single Performance Indicator Termination Criteria

In general, to stop the evolutionary search process, it is needed to predefine a desired value for a performance indicator (Wagner et al., 2011). Termination criterion is fulfilled when the predefined value is triggered. In this section, the performance indicators which have been used as termination criteria are discussed in a simplified way.

Hypervolume Metric. Hypervolume metric calculates the volume of the approximation set at generation t with respect to nadir point or the reference set (Coello et al., 2006). Hypervolume metric is one of the most widely used performance indicator. Hypervolume has been used as one of the metrics to determine the stopping generation in (Trautmann et al., 2009), (Guerrero et al., 2009) and (Guerrero et al., 2010).

Additive Epsilon Indicator. Additive epsilon indicator is known as one of quality indicators to measure the performance MOEAs. This indicator uses the concept of domination to calculate the additive minimum distance in which the solution set covers every solution towards the reference set (Zitzler et al., 2003). (Trautmann et al., 2009), (Guerrero et al., 2009) and (Guerrero et al., 2010) used additive epsilon indicator as termination criteria in their respective works.

Mutual Domination Rate Metric. Set coverage metric or contribution percentage gives a percentage with respect to the number of solutions from the approximation set at generation t which belong to the reference set (Zitzler et al., 2000). Mutual domination rate metric is a customized version of set coverage metric to show the number of non-dominated solutions of generation t which can be dominated by at least one solution belongs to non-dominated solutions of generation $t-1$. Mutual domination rate metric is used as a termination criteria in a work done by (Guerrero et al., 2009)

3.5.2 Multiple Performance Indicator Termination Criteria

Regarding to the characteristics of performance indicators, the quality of approximation set cannot be evaluated using only one performance indicator. Therefore, it is recommended to apply multiple performance indicators to measure the convergence of a MOEA in terms of diversity and coverage (Hadka and Reed, 2012) and (Ghoreishi et al., 2015).

Historically, the idea of using performance indicators over the run of MOEA was first presented in 2002 in a joint work by Deb and Jain (Deb and Jain, 2002). They applied two performance indicators to compute the convergence and the diversity of the approximation set at generation t.

Later, in another work performed by Trautmann et al., an offline convergence analysis has been presented (Trautmann et al., 2008). Three performance indicators generational distance, spread and hypervolume are calculated in each generation. Then Kolmogorow Smirnov test is applied on the performance indicators of the current and past five consecutive generations. The termination criterion is fulfilled if the distribution of indicators value has no change, otherwise the evolution is allowed to continue (Trautmann et al., 2008).

3.6 Progress Indicator Termination Criteria

As the previous sections describe, termination of MOEA is often decided based on heuristic stopping criteria, such as the maximum number of evaluations or a desired value of a performance indicator. Whereas the termination criteria are suitable for defined benchmark problems, where the optimal indicator value is known, their applicability to the real-world problems is still questionable (Wagner et al., 2011). The challenge raises in cases where the evaluation budget or the desired indicator level is inappropriately specified, so the MOEA can either waste computational resources or can be stopped although the approximation still shows a significant improvement. For that reason, heuristic stopping criteria for the online detection of the generation, where the expected improvement in the approximation quality does not justify the costs of additional evaluations, provide an important contribution to the efficiency of MOEA (Wagner et al., 2011).

Performance indicators described in section 3.6 are known as unary performance indicators computed by comparing the approximation set at generation t with the reference set. Unary performance indicators except spacing, require the knowledge of known optimal solutions which is not always available. In order to use performance indicators independent of the reference set, a customized version of the indicator is calculated considering two consecutive generations t and $t-1$. These indicators are known as binary performance indicators or progress indicators which measure the performance improvement of MOEA.

In recent years, research on sophisticated heuristic Online Stopping Criteria (OSC) has became extensively popular (Martí et al., 2009), (Trautmann et al., 2009), (Wagner and Trautmann, 2010), (Goel and Stander, 2010) and (Wagner et al., 2009). OSC compute the progression of single or multiple Progress Indicators (PI) during the run of the MOEA. At convergence time, the expected improvement for considered indicators seems to be lower than a specified threshold and the MOEA is terminated in order to avoid needless computations. OSC are supposed to detect that further improvements are unlikely, or are expected to be too small even if no formal convergence is obtained (Wagner et al., 2011).

In general, the procedure of OSC can be divided into two main steps: 1. The progress improvement of the MOEA is evaluated using progress indicators. 2. Given a predefined threshold and the value obtained from PIs, a decision about termination of the MOEA is made (Wagner et al., 2011). In the following, single PI termination criteria and aggregated PI termination criteria are presented in more details.

3.6.1 Single PI Termination Criteria

Single performance Indicator termination criteria refer to a class of termination criteria in which one of the binary performance indicators are used for calculation of termination time. In the following, single progress indicator termination criteria are described in details.

Hypervolume Metric. Customized version of hypervolume metric as a binary performance indicator is calculated by using non-dominated solution set obtained at two consecutive generations t and $t-1$ (Guerrero et al., 2009).

Additive Epsilon Metric. Similar to hypervolume metric, Additive Epsilon Metric can be customized to work upon solution sets of two consecutive generations, t and $t-1$. It has been used as termination criterion in two works done by (Guerrero et al., 2009) and (Trautmann et al., 2009).

Mutual Domination Rate Metric. This metric is derived from set coverage metric which gives either 1 or 0 to check the number of non-dominated solutions at generation t which dominate the non-dominated solutions at generation $t-1$ (Guerrero et al., 2009).

MDR equals to 1 indicates that the entire approximation set at generation t is better than its predecessor. For MDR equals to 0, no substantial progress has been achieved. MDR<0 indicates a deterioration of the current population.

Dominance-based Quality of Pareto Front. Bui et al. introduced the dominance-based quality of Pareto front metric (DQP) for an approximation set at generation t (Bui et al., 2009). For each solution in the approximation set at generation t, the ratio of dominating individuals in the neighborhood of this solution is calculated by performing Monte Carlo sampling simulation with 500 evaluations. DQP equals to 0 indicates that no improved solutions can be found in the neighborhood of the current solutions in the approximation set at generation t.

Consolidation Ratio. The Consolidation Ratio (CR) is introduced by (Goel and Stander, 2010) as a dominance-based convergence metric. To calculate this metric, an external archive of all non-dominated solutions found during the run of the MOEA is needed. Given the archive, CR is defined as the relative amount of the archive members in generation t_{mem} which are still contained in the archive of the current generation t.

3.6.2 Aggregated PI Termination Criteria

In single PI termination criteria, the value of the PI calculated for generation t and used directly to decide on termination time. However, a single PI evaluation usually cannot provide enough information for a robust conclusion. Because of the non-deterministic nature of EAs, it can be advantageous to define an evidence gathering process (EGP) to incorporate different values for PI (Wagner et al., 2011). EGP is designed to store the calculations of more than one PI for previous generations. Descriptive statistics justifies the need for aggregating different PIs ie., the standard deviation of PI values can evaluate the variability within the PI values(Rudenko and Schoenauer, 2004), (Wagner et al., 2009) and (Wagner and Trautmann, 2010). In most known aggregated PI approaches, MOEA terminates when the current value of the EGP triggers a threshold or a given probability limit (Rudenko and Schoenauer, 2004), (Bui et al., 2009) and (Goel and Stander, 2010).

In the following of this section, the most known and common approaches for aggregated PI are presented in a chronological order. For simplicity, we describe the approaches by presenting the structure of EGP and the final decision making for termination of the MOEA without the mathematical representation of each approach.

Running Metrics. In a joint work done by Deb and Jain performance indicators are used to evaluate convergence and diversity of the approximation set at generation t known as convergence metric (CM) and diversity metric (DVM). By this, all objective vectors of the approximation set at generation t are projected onto a dimensional hyperplane presented as discrete grid cells (Deb and Jain, 2002). The convergence metric (CM) calculates the average of the smallest normalized euclidean distance from each individual in the approximation set at generation t to a known reference set. Whereas DVM tracks the number of attained grid cells and computes the distribution by assigning different scores for predefined neighborhood patterns. In this approach, EGP is a visual investigation of the progression of CM and DVM done by user which leads to the final decision for termination of the MOEA (Deb and Jain, 2002).

Stability Measure. According to the work done by Rudenko and Schoenauer in 2004, the stagnation of the maximum crowding distance (maxCD) within the approximation set at generation t can be used as an indicator to detect convergence of NSGA-II algorithm (Rudenko and Schoenauer, 2004). After each generation, maxCD and standard deviation (STD) over approximation set is computed. In this approach, EGP stores STD of the values of maxCD. The stability measure termination criterion is fulfilled once SDT falls below a user-defined threshold ε.

MBGM Termination Criterion. MBGM Termination Criterion (according to the authors' last names) uses combination of the mutual domination rate (MDR) and a simplified version of Kalman filter (Martí et al., 2007). For each generation, MDR indicator is applied to the approximation set obtained from two last consecutive generations t and $t-1$. EGP stores all MDR values. After that Kalman filter is applied on EGP and the corresponding estimated error is calculated. MBGM Termination Criterion is fulfilled when the confidence interval of the a-posteriori estimation falls below the predefined threshold ε (Martí et al., 2009).

Classic On-line Convergence Detection. In 2009, Naujoks and Trautmann presented a new approach called On-line Convergence Detection (OCD) to solve a multi-objective optimization in an aerodynamic application. This approach is known as classic OCD because it is considered as original version of OCD for further studies. In OCD approach, PI is incorporated three performance indicators hypervolume, R2-Indicator and additive ε-indicator (Naujoks and Trautmann, 2009), (Wagner et al., 2009) and (Wagner and Trautmann, 2010). At each generation, the approximation set at generation t is considered as the reference set to update PI values stored in EGP. Later for each PI, different variance tests, corresponding p-values and standard errors are computed and stored in EGP. Moreover, after standardizing the values stored in EGP individually, a least-squares fit of a linear model with slope parameter β is performed. Finally, the termination decision is made when the p-values of two consecutive generations are below the confidence level α for one of the variance tests where $\alpha = 0.05$.

ODC-Hypervolume. Later in 2010, Wagner and Trautmann proposed a new version of classic ODC for Multi-objective selection based on dominated hypervolume to create PI. Since the hypervolume is a unary indicator, only the absolute values for hypervolume have to be stored in EGP. Compared to ODC, the complexity of OCD-Hypervolume is reduced by concentrating on the variance test for one specific PI. The termination decision is similar to the classic OCD approach.

Least Squares Stopping Criterion. In 2010, Guerrero et al. presented a light version of classic ODC which simplifies PI computation, EGP, and termination criterion. Similar to classic OCD, generation t as the reference set is considered to update PI values and EGP stores a regression analysis performed on PI values. In contrast, in Least Squares Stopping Criterion (LSSC), only one PI is considered while the variance tests stored in EGP to make the termination decision are omitted (Guerrero et al., 2010). In addition, in LSSC, PI values are not standardized in order to be able to observe the expected improvement by means of the slope β. Consequently, the analyses performed in OCD and LSSC are different. The termination decision is made when β falls below the predefined threshold ε. Interested audiences can read more about the details of implementation of classic ODC and LSSC in (Naujoks and Trautmann, 2009) and (Guerrero et al., 2010), respectively.

Non-dominance-based Convergence Metric. Another work is done in 2010 by Goel and Standerto present a non-dominance-based on-line termination criterion for MOEAs. This approach uses a dominance-based PI based on an external archive of non-dominated solutions which is updated in each generation. In addition to that, utility is defined as a parameter to compute the difference in value of CR between the approximation set at generations t and the external archive of non-dominated solutions. In this approach, EGP is designed as a moving average as utility U_t to increase the robustness of the approach. The termination decision is made when the utility falls below an adaptively computed threshold called $\varepsilon_{adaptive}$.

3.7 Termination Criteria in Hybrid MOEA

Research on termination criteria in MOEAs has highlighted several approaches to identify the termination of evolution process. However previous studies have not dealt with hybrid MOEAs and it is not certain whether the existing termination criteria are suitable for hybrid MOEAs and whether they are adequate enough to identify convergence in such algorithms (Bhuvana and Aravindan, 2016b). Previous research has established that incorporating additional knowledge during search process can improve the performance of evolutionary algorithms. The additional knowledge is gained by applying a local refinement procedure in the evolution process (Bhuvana and Aravindan, 2016b). The main advantage of hybrid MOEAs is to incorporate local refinement procedures to prevent premature convergence of the search process (El-mihoub et al., 2006).

In a worked done by Bhuvana and Aravindan in 2016, a hybrid MOEA has been proposed in which the termination happens respect to the maximum number of functional evaluations known as MAPLSAW (Bhuvana and Aravindan, 2016a). The experimental results show that fixing an upper bound for number of objective function evaluations or number of generations will not be an appropriate termination condition in all cases (Bhuvana and Aravindan, 2016a). Moreover, another recent work is done by Bhuvana and Aravindan in 2016, to develop a new termination criteria for hybrid MOEAs. They proposed a termination scheme including five termination criteria which two of them are new termination criteria proposed for a hybrid MOEA called MAPLS-AW (Bhuvana and Aravindan, 2016b).

The termination scheme relies on three performance indicators, hypervolume, mutual domination metric and additive-ε indicator together with the concept of maintaining the elites in the population over generations. In the proposed schema, two new termination criteria derived based on the features of MAPLS-AW to detect the convergence of the population, named Preferential Local Search (PLS) and Elite

Average Fitness (EAF). In each generation, PLS and EAF are computed and compared with the values obtained in the previous generations. In the following of this section, PLS and EAF are explained in more details.

Preferential Local Search. In hybrid MOEA presented by In MAPLS-AW implementation presented in Bhuvana and Aravindan, the local refinement procedure relies on elite solutions identified at every generation (Bhuvana and Aravindan, 2016b). Over generations, PLS deepens on the elites iteratively until such elite solutions become locally optimal. Locally optimized elites are considered as the measure of convergence to calculate the stopping criteria, SC_{PLS}.

In order to calculate SC_{PLS}, a local termination criterion, L_{PLS} is needed. L_{PLS} is computed using the current optimized elites and the total number of elites at every generation and compared with its previous generation to derive SC_{PLS}. Termination decision is made when TC_{PLS} equal to 0 which means the value of L_{PLS} is not changed over last two generations, t and $t-1$. SC_{PLS} equal to -1 shows that no improvement have been observed from generation t to generation $t-1$. SC_{PLS} is equal to 1 indicates that the population is still evolving and the current generation has progress toward optimal solutions (Bhuvana and Aravindan, 2016b).

Elite Average Fitness. Similar to PLS, Elite Average Fitness stopping criterion SC_{EAF} is proposed in the same work by Bhuvana and Aravindan for the hybrid algorithm, MAPLS-AW (Bhuvana and Aravindan, 2016b). EAF incorporates elite's average fitness value at every generation using a local measure called L_{EAF}. Calculating I_{EAF} is a challenging task in a multi-objective problem when multiple objectives are not combined into a single objective. For dealing with this challenge, MAPLS-AW has proposed an adaptive weight (AW) objectives by considering their positions in the objective space. I_{EAF} of two consecutive generations t and $t-1$ are used to compute SC_{EAF}. Termination decision is made if no improvement have been observed by SC_{EAF}.

4 DISCUSSION

Recalling the research questions in section 2.1, the most frequent termination criteria and their main characteristics explained in details in seven different categories. We distinguished the categories by considering functionality and applicability of the termination criteria and the way they have been evolved over time. Applying all aforementioned termination criteria in different test cases is out of the scope of this work. But nevertheless, we discuss on strengths and weaknesses of termination criterion belong to each category in the following section.

In overall, even though direct termination criteria are easy to implement but determining the fixed value for termination condition is again a challenge. Moreover, finding optimal values requires a good and a-priori knowledge about the behavior of the GA for the specific problem and the global optimal solution which are not always available (Bhandari et al., 2012).

In 2001, Jain et al. evaluated direct and derived termination criteria from two main aspects, reliability and performance (Jain et al., 2001). They concluded that all direct termination criteria except the hitting a bound, are reliable criteria which guarantee termination of EA within finite number of iterations or finite time. One major drawback for the hitting a bound termination criterion is that it does not terminate the algorithm if it converges to a sub-optimal solution because in this case the objective value of the sub-optimal solution is always worst than the required bound (Jain et al., 2001). In order to guarantee termination, it is suggested to apply the hitting a bound condition in combination of any other direct termination criteria (Jain et al., 2001). Other finding of the same work shows selecting suitable values for the parameters of the corresponding criteria has direct influence on performance of direct termination criteria which can be found by trial-and-error method while setting a default value is not recommended (Jain et al., 2001).

The reliability and performance of derived termination criteria has been analyzed and evaluated in a research done by (Jain et al., 2001). Among all, running means termination criterion is a reliable one which assures the termination of the algorithm after finite number of iterations. In contrast, four termination criteria presented in sections 3.2-3.2 can only terminate the search process if the objective values of all individuals at the current generation are sufficiently similar. This means that termination is not guaranteed and can be postponed if few outliers are existed in the population. To avoid prevention of termination, it is suggested to combine these criteria with other direct termination criteria discussed earlier in section 3.1.

Bhandari et al. proposed ε-Variance termination criterion which is similar to both K-Variance and POP-Var termination criteria in some extents (Bhandari et al., 2012). The main difference between the ε-Variance and K-Variance termination criteria is that the user does not need to predefine the number of iterations in ε-variance termination criterion. By predefining K, user assumes that it is impossible to obtain a better solution after K consecutive iterations. But the main challenge for using K-Variance termination criterion is that, it is proven that given a finite value for K, there is always a positive probability of obtaining K consecutive equal sub-optimal solutions (Bhandari et al., 2012).

On the other hand, in POP-Var termination criterion, the variance is calculated using the individuals in the current population while in ε-Variance termination criterion, the variance of the best solutions over generations is used as termination condition. However, ε-Variance still has two main limitations. The first one is the necessity to have a prior knowledge about the global optimal solution which makes it inapplicable when the optimal solution is unknown or not available. And the second is to choose the optimal value for ε. A challenge to apply this criterion is that no automatic way of choosing the value ε is suggested (Bhandari et al., 2012). Experimental results show that the value of ε varies in different problems with the same size of search space to obtain global optimal solution (Bhandari et al., 2012). Smaller number of ε close to 0 leads to higher accuracy. This means that the choice of ε is dependent on the level of accuracy that the user desires.

In a similar way, in CLusTerm termination criterion, selecting an appropriate value for ε is important to obtain a proper termination. The parameter ε controls the duration of stagnation until termination. Hence, smaller values of ε prolongs stagnation phase without termination and vice versa. In practical use, it is recommended to employ ClusTerm or running mean termination criterion presented in 3.2, together with one of the four approaches provided in sections 3.1 and 3.2-3.2, to prevent needless computations (Jain et al., 2001).

It is worthy to mention that operator-based approaches may not focus directly on proposing a new termination criterion but they are designed mostly to accelerate the convergence of the EA in an automatic way to avoid unnecessary computations (Ghoreishi et al., 2015). Operator-based approaches are varies based on the implementation specifications. For instance, as discussed earlier, CONTROLEUM-GA applies domain specific genetic operators to avoid generating new solutions that are not valid based on the existing knowledge about the problem and environmental measurements (Ghoreishi et al., 2015). The strength point about domain specific approach is the automatic way of accelerating the search process.

Other approaches such as ε-Optimal Stopping Time termination criterion, has still two main drawbacks. The first one is a prior knowledge requires about the optimal solution which is not always available for many real world problems. And the second one is that it is proven that with existing operations in GA, whatever value decided for the number of iterations, there is always a positive probability of not obtaining an optimal solution if ε is not small enough (Bhandari et al., 2012).

The main purpose of applying performance indicators is to measure the convergence and diversity of the solutions without compromising the quality. Unary performance indicators (both single and multiple termination criteria) except spacing are computed by comparing the approximation set at generation t towards the reference set which requires the knowledge of known optimal solutions. But on the contrary, binary performance indicators are independent of the reference set. Binary performance indicators are known as progress indicators (PIs). They are customized in a way that they consider two consecutive generations t and $t-1$ instead of using the approximation set and the reference set for the measurements.

Among all single PIs, DQP is the only PI that requires many additional evaluations of the objective functions for computing the ratio of dominating solutions in the neighborhood of a specific solution. Consequently, DQP is a very expensive, but powerful measure and it is only applicable if many additional evaluations of the objective functions can be performed. In contrast, MDR is capable of measuring the progress of the optimization with almost no additional computational cost by reusing the computations done to apply Pareto-optimality. Therefore it is suitable for solving large-scale or many-objective problems with large population sizes. In addition, CR is also not an efficient metric for problems with many objectives, due to the large archive size.

In single PI termination criteria, only the last value of PI is used to decide on termination time which cannot always provide enough information for a robust conclusion due to the non-deterministic nature of EAs (Wagner et al., 2011). For this reason, it can be beneficial to incorporate different values of PI for previous generations using an evidence gathering process (EGP) in aggregated PI approaches.

One of the important issues in designing aggregated PI approaches is concerning about using the statistical tests and confidence intervals which make the termination decision robust compared to monitoring random variables overtime. In addition, parameterization of aggregated PI approaches plays important role in obtaining the desired results with respect to the trade-off between runtime and approximation quality (Wagner et al., 2011). One of the limitations with aggregated PI approaches is that there is no clear guideline for parameterization and visualized analysis (Deb and Jain, 2002) and (Bui et al., 2009). One of the only guideline is proposed by Wagner and Trautmann for parameterization of ODC approach based on statistical analysis of experimental results (Wagner and Trautmann, 2010).

Discussion on termination criteria in hybrid approaches is a challenging task because most studies in the field of EA have focused on termination criteria in single MOEA. Nevertheless, the experimental results indicate that PLS and EAF together with three performance indicators are promising approaches for predicting the convergence of population by considering additional knowledge about the number of locally optimized elites in the population and their average fitness (Bhuvana and Aravindan, 2016b).

5 CONCLUSION

The study was conducted in the form of a survey to provide a concise categorization of prominent termination criteria in EA. This research will serve as a base for future studies targeting termination criteria in EA and makes several noteworthy contributions to the current existing literature. This work extends our knowledge about the existing termination criteria and can be used to improve the development of such approaches in terms of convergence speed, computation cost and assessment of the quality of approximation set with or without known reference set. In addition to principal theoretical implication of this study, strengths and weaknesses of the approaches are highlighted and discussed.

A key strength of the present study was to highlight the role of combination of direct termination criteria and threshold-based termination criteria in order to guarantee the convergence of EA in a reliable manner. Whilst the main focus of this study was not the evaluation of aforementioned termination criteria, rigorous and comprehensive set of experiments is still needed. Therefore, it would be interesting to assess the effects of selecting different termination criteria and evaluating their impacts on different problem domains with various number of objectives and also their applicability to use in real-world applications.

It is worthy to note that, the termination criteria in this survey are the most commonly used ones and the reader can investigate about the details of implementation and mathematical formulations in the respective references.

REFERENCES

Bhandari, D., Murthy, C. A., and Pal, S. K. (1996). Genetic algorithms with elitist model and its convergence. *International Journal of Pattern Recognition and Artificial Intelligence*, 10(6):731–747.

Bhandari, D., Murthy, C. A., and Pal, S. K. (2012). Variance as a stopping criterion for genetic algorithms with elitist model. *Fundam. Inf.*, 120(02):145–164.

Bhuvana, J. and Aravindan, C. (2016a). Memetic algorithm with preferential local search using adaptive weights for multi-objective optimization problems. *Soft Comput.*, 20(4):1365–1388.

Bhuvana, J. and Aravindan, C. (2016b). Stopping criteria for mapls-aw, a hybrid multi-objective evolutionary algorithm. *Soft Comput.*, 20(6):2409–2432.

Bui, L. T., Wesolkowski, S., Bender, A., Abbass, H. A., and Barlow, M. (2009). A dominance-based stability measure for multi-objective evolutionary algorithms. In *Proceedings of the Eleventh Conference on Congress on Evolutionary Computation*, CEC'09, pages 749–756, Piscataway, NJ, USA. IEEE Press.

Coello, C. A. C., Lamont, G. B., and Veldhuizen, D. A. V. (2006). *Evolutionary Algorithms for Solving Multi-Objective Problems (Genetic and Evolutionary Computation)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Deb, K. and Jain, S. (2002). Running performance metrics for evolutionary multi-objective optimization. Technical report.

El-mihoub, T. A., Hopgood, A. A., Nolle, L., and Battersby, A. (2006). Hybrid genetic algorithms: A review.

Ghoreishi, S. N., Sørensen, J. C., and Jørgensen, B. N. (2015). Enhancing state-of-the-art multi-objective optimization algorithms by applying domain specific operators. *Proceeding of 2015 IEEE Symposium Series on Computational Intelligence*.

Goel, T. and Stander, N. (2010). A non-dominance-based online stopping criterion for multi-objective evolutionary algorithms. *International Journal for Numerical Methods in Engineering*, 84(6):661–684.

Greenhalgh, D. and Marshall, S. (2000). Convergence criteria for genetic algorithms. *SIAM J. Comput.*, 30(1):269–282.

Guerrero, J. L., Garcia, J., Marti, L., Molina, J. M., and Berlanga, A. (2009). A stopping criterion based on kalman estimation techniques with several progress indicators. In *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, GECCO '09, pages 587–594, New York, NY, USA. ACM.

Guerrero, J. L., Martí, L., Berlanga, A., García, J., and Molina, J. M. (2010). Introducing a robust and efficient stopping criterion for multi-objective evolutionary algorithms. *IEEE Congress on Evolutionary Computation*.

Hadka, D. and Reed, P. (2012). Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. *Evol. Comput.*, 20(3):423–452.

Hansen, N. and Kern, S. (2004). Evaluating the cma evolution strategy on multimodal test functions. *Proceedings of Eighth International Conference on Parallel Problem Solving from Nature, PPSN VIII*, 3242:282–291.

Hedar, A.-R., Ong, B., and Fukushima, M. (2007). Genetic algorithms with automatic accelerated termination.

Jain, B., Pohlheim, H., and Wegener, J. (2001). On termination criteria of evolutionary algorithms. pages 768–768.

Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.

Konar, A. (2005). *Computational Intelligence: Principles, Techniques and Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Lee, C.-Y. and Yao, X. (2004). Evolutionary programming using mutations based on the levy probability distribution. *Trans. Evol. Comp*, 8(1):1–13.

Leung, Y. and Wang, Y. (2001). An orthogonal genetic algorithm with quantization for global numerical optimization. *Trans. Evol. Comp*, 5(1):41–53.

Lozano, J. A., Larrañaga, P., Inza, I. n., and Bengoetxea, E. (2006). *Towards a New Evolutionary Computation: Advances on Estimation of Distribution Algorithms (Studies in Fuzziness and Soft Computing)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Martí, L., García, J., Berlanga, A., and Molina, J. M. (2007). A cumulative evidential stopping criterion for multiobjective optimization evolutionary algorithms. In *Proceedings of the 9th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '07, pages 2835–2842, New York, NY, USA. ACM.

Martí, L., García, J., Berlanga, A., and Molina, J. M. (2009). An approach to stopping criteria for multi-objective optimization evolutionary algorithms: The mgbm criterion. *IEEE Congress on Evolutionary Computation*.

Murthy, C., Bhandari, D., and Pal, S. K. (1998). e-optimal stopping time for genetic algorithms. *Fundam. Inform.*, 35(1-4):91–111.

Naujoks, B. and Trautmann, H. (2009). Online convergence detection for multiobjective aerodynamic applications. In *Proceedings of the Eleventh Conference on Congress on Evolutionary Computation*, CEC'09, pages 332–339, Piscataway, NJ, USA. IEEE Press.

Ong, Y., Lim, M., Zhu, N., and Wong, K. (2006). Classification of adaptive memetic algorithms: A comparative study. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, 36(1):141–152.

Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. In *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*, EASE'08, pages 68–77, Swindon, UK. BCS Learning & Development Ltd.

Rudenko, O. and Schoenauer, M. (2004). A steady performance stopping criterion for pareto-based evolutionary algorithms. *The 6th International Multi-Objective Programming and Goal Programming Conference*.

Rudolph, G. (1994). Convergence analysis of canonical genetic algorithms. *Trans. Neur. Netw.*, 5(1):96–101.

Suzuki, J. (1995). A markov chain analysis on simple genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(4):655–659.

Trautmann, H., Ligges, U., Mehnen, J., and Preuss, M. (2008). A convergence criterion for multiobjective evolutionary algorithms based on systematic statistical testing. In *Proceedings of the 10th International Conference on Parallel Problem Solving from Nature: PPSN X*, pages 825–836, Berlin, Heidelberg. Springer-Verlag.

Trautmann, H., Wagner, T., Naujoks, B., Preuss, M., and Mehnen, J. (2009). Statistical methods for convergence detection of multi-objective evolutionary algorithms. *Evol. Comput.*, 17(4):493–509.

Tsai, J.-T., Liu, T.-K., and Chou, J.-H. (2004). Hybrid taguchi-genetic algorithm for global numerical optimization. *IEEE Transactions on Evolutionary Computation*, 8(4):365–377.

Wagner, T. and Trautmann, H. (2010). Online convergence detection for evolutionary multi-objective algorithms revisited. *IEEE Congress on Evolutionary Computation*.

Wagner, T., Trautmann, H., and Martí, L. (2011). A taxonomy of online stopping criteria for multi-objective evolutionary algorithms. In *Proceedings of the 6th International Conference on Evolutionary Multi-criterion Optimization*, EMO'11, pages 16–30, Berlin, Heidelberg. Springer-Verlag.

Wagner, T., Trautmann, H., and Naujoks, B. (2009). Ocd: Online convergence detection for evolutionary multi-objective algorithms based on statistical testing. In *Proceedings of the 5th International Conference on Evolutionary Multi-Criterion Optimization*, EMO '09, pages 198–215, Berlin, Heidelberg. Springer-Verlag.

Zapotecas Martínez, S., Yáñez Oropeza, E. G., and Coello Coello, C. A. (2011). Self-adaptation techniques applied to multi-objective evolutionary algorithms. In *Proceedings of the 5th International Conference on Learning and Intelligent Optimization*, LION'05, pages 567–581, Berlin, Heidelberg. Springer-Verlag.

Zhong, W., Liu, J., Xue, M., and Jiao, L. (2004). A multiagent genetic algorithm for global numerical optimization. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, 34(2):1128–1141.

Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of multiobjective evolutionary algorithms: Empirical results. *Evol. Comput.*, 8(2):173–195.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and da Fonseca, V. G. (2003). Performance assessment of multiobjective optimizers: An analysis and review. *Trans. Evol. Comp*, 7(2):117–132.