Convergence Measurement in Evolutionary Computation using Price’s Theorem

Hassan A. Bashir
School of Computer Science
University of Manchester
Manchester, UK
Email: bashirh@cs.man.ac.uk

Richard S. Neville
School of Computer Science
University of Manchester
Manchester, UK
Email: r.neville@manchester.ac.uk

Abstract—Evolutionary computations are naturally inspired stochastic algorithms that are capable of running perpetually. When deployed as optimization tools, it is imperative to prescribe a set of definitive stopping criteria that if satisfied, the evolutionary process could be brought to a halt. User specified limits on maximum evaluations or generations are the common measures used to stop the evolution due to resource constraints that might directly/indirectly be imposed on the system. Conversely, we propose a novel convergence detection mechanism that monitors the contribution of the genetic operators on the fitness progress and the diversity profile of the population via the $\pm \sigma$ crossover envelope. This adaptively terminates the evolution as convergence sets in. Extended Price’s theorem is utilized to estimate the dynamical contributions of the individual genetic operators. Experimental results show that under standard parameter settings with binary tournament selection, the proposed technique is robust and could be a promising alternative to the conventional similarity measure-based methods for convergence detection.

I. INTRODUCTION

Evolution is a process that originated from the biologically inspired neo-Darwinian paradigm [1] (i.e. the principle of survival of the fittest). It is believed to be a collection of stochastic processes that act on and within populations of species. These processes include reproduction, mutation, competition and selection. The far reaching impact of the idea of evolution has gone beyond the classical boundaries of biological thoughts with the process of evolution being used as an optimization tool in solving complex engineering problems. This applied simulated evolution is known as evolutionary computation (EC). For extensive review on EC, see [2].

Contrary to the traditional optimization methods, a genetic algorithm (GA) [3] is an evolutionary iterative process (cf. Fig. 1) that starts with an initial fixed set of candidate solutions called population $P(t)$ (line 3). A solution point is called an individual and represents a possible solution to the problem under consideration. An individual is represented by a computational data structure called a chromosome. Usually, a chromosome is encoded as a string of symbols of finite-length called genes. The possible values a gene can take correspond to the allele. Typically, a chromosome is represented as a binary bit string.

The chromosomes in the initial population are usually created randomly or via a simple heuristic construction. During each iteration step $t$, called a generation (lines 5-12), a stochastic selection process is applied on the initial population to choose better solutions following an evaluation that is based on some measures of fitness. Chromosomes that survive through the selection process constitute a new set $Q_s(t)$ called parents and are qualified to take part in the remaining stages of the evolution process.

In order to explore other areas of the search space, the parent chromosomes undergo recombination and/or mutation operations and respectively generate a new set of chromosomes $Q_r(t)$ and $Q_m(t)$ called offspring populations. The recombination entails exchange of characteristics by merging two parent chromosomes using a crossover operator, while mutation operation is a genetic alteration of a randomly chosen parent chromosome by a mutation operator.

A new generation $P(t+1)$ of chromosomes is then formed by selecting from either the combined pool of parents and offspring or the offspring pool based on a prescribed fitness measure. Fitter chromosomes have higher chances of being selected and the average fitness of the population is expected to grow with successive generations. The process continues until a termination criterion is met or it ultimately converges to the best chromosome which hopefully represents the optimum or suboptimal solution to the problem.

Consequently, it is crucial to carefully assess the long-term behaviour of these algorithms. Typically, at some stage of evolution, the search progress slows and the quality of the candidate solutions barely improve, this is a signal for convergence and majority of the convergence measures are designed to detect such conditions.

In section II some commonly used stopping and convergence measures are presented. The use of the Price’s theorem to investigate the effect of genetic operators on evolution progress is examined in section III. Empirical experiments using extended Price’s equation to compare and analyse the long-term behaviour of the standard EC parameters are presented and discussed in section IV. In section V, a new convergence measure is proposed and conclusions are given in section VI.

II. CONVENTIONAL MEASURES OF EC CONVERGENCE

Beside stopping evolution based on some user prescribed limits on evaluations, generations or execution time for a run, more sophisticated convergence measures that are mainly
hybridization approaches play a vital role to improving
genetic algorithm to converge to an optimal solution can be
duced by incorporating some information about the gradients
of the problem’s variables. Thus, they propose hybridizing GA with a steepest descent algorithm and argue that this
can positively influence the overall convergence rate. Similar
proposals were made in [7], [8].

Virtually all the aforementioned approaches share their
merits and demerits. For instance, use of Hamming distance
can allow effective analysis of the diversity profile in a given
population; as convergence sets in, the search space reduces
to a Hamming landscape [9] of bit-flip neighbours. However,
we observe that this similarity measure works in the genotype
space, but lacks any knowledge of the dynamics in the solution
(i.e., phenotype) space. Thus, depending on the mapping
technique, closely related individuals in the genotype space
may differ significantly in the phenotype space. We further
observe that estimating GA convergence by marryng the two
diversity measures that monitor similarity in both the encoding
and solution spaces might not totally eliminate dissension
between the methods and could trigger false alarms for con-
vergence. Moreover, the combination can severely increase the
computational cost of the convergence detection process.

We therefore argue that since progress in evolutionary
search is internally governed by the actions of genetic oper-
ators (selection and reproduction operators like crossover and
mutation), building a convergence measure that assesses the
dynamical roles of these operators could be a promising
alternative to tackling this dilemma.

III. MONITORING THE EFFECT OF EC OPERATORS WITH
PRICE’S THEOREM

In order to examine the independent effect of EC operators
while in interaction, Price [10] formulated a theorem that
permits decomposition of the evolutionary process to separate
the genetic effect (or contribution) of the selection operator
from that of other reproduction operators (i.e. crossover and
mutation). Although Price’s work was mainly in the field of
evolutionary genetics, Price’s equation provides greater insight
into general selection processes (1). Price’s theorem states that:

\[
\Delta Q = \frac{\text{Cov}(z, q)}{\bar{z}} + \frac{\sum_{i=1}^{N} z_i \Delta q_i}{N \bar{z}}
\]

where \(\Delta Q = Q_2 - Q_1\) is the change in the measured
characteristics \(Q\) which could be fitness, \(N\) is the number of
individuals in the parent population (i.e., population size), \(z_i\)
is the number of offspring of parent \(i\), and \(\bar{z} = \sum z_i / N\) is the
average number of the offspring produced. Also, \(\Delta q_i = q_i' - q_i\),
where \(q_i\) is the fitness of parent \(i\) and \(q_i'\) is the average fitness of
the offspring of parent \(i\). And finally, \(z\) and \(q\) are vectors of
\(z_i\) and \(q_i\) respectively.

The two terms in the Price’s theorem (1) represent the
ctribution of different operators as the mean of the character-
istic being measured. Since we deploy the Price’s equation
to analyse convergence in EC, the measured characteristic
\(Q\) in this case will be fitness. The first term represents
the contribution of selection operator while the second term

Based on population diversity are used to automatically stop
the evolution. In this respect, population diversity could relate
to the genotype, phenotype or even the average fitness of the
population directly. Usually, diversity is measured by
assessing the similarity among the candidate solutions in the
genotype space using a distance measure [4] such as the
Hamming distance. This is because evolution converges when
the candidate solutions in the population become identical.

While the convergence characteristics of all evolutionary
algorithms rely on several factors, the most important factors
are the types of selection mechanism, the reproduction opera-
tors and the population size. Being an exploitation process, the
selection operation favours the fitter individuals at the
expense of the weaker ones and is seen as the major derive
to convergence in EC. Thierens et al. [5] use the normal fitness
distribution method to model the convergence characteristics
of various selection methods; via simple yet elegant ordinary
differential equations models that estimate the true conver-
gence behaviours of various selection methods. They argue
that fitness proportionate selection (FPS) methods have the
longest convergence time and can further slow down as the
search approaches optimal solution.

Reproduction operators influence GA convergence, and as
such, it is important to realize that besides the types of the
operators, the frequency of their application also play a
major role. A new diversity measure that estimates the
average Hamming distance in the population was proposed
by [4]. They noted that while mutation and its probability
of application can severely influence the convergence rate,
vast majority of the traditional crossover operators such as
\(n\)-point, uniform and punctuated crossover have little effect
on GA convergence. They analytically prove that the new
diversity measure (i.e., average Hamming distance) is not
affected by crossover operators but mostly influenced by the
selection mechanism. The measure was then used to predict
the average Hamming distance at convergence and thus define
an upper bound limit on the evolution run time.

Hybridization approaches play a vital role to improving
GA convergence rate. As reported by Miura et al. [6], for
most nonlinear optimization problems, the time required by a

```plaintext
1: begin
2: \(t \leftarrow 0;\)
3: initialize \(P(t)\);
4: evaluate \(P(t)\);
5: while not termination do
6: \(Q_s(t) \leftarrow \text{select } P(t)\);
7: \(Q_r(t) \leftarrow \text{recombine } Q_s(t)\);
8: \(Q_m(t) \leftarrow \text{mutate } Q_r(t)\);
9: evaluate \(Q_m(t)\);
10: \(P(t+1) \leftarrow \text{select } Q_m(t) \cup P(t)\);
11: \(t \leftarrow t + 1;\)
12: end while
13: end
```

Fig. 1. A Canonical Model of Genetic Algorithm - after [2].

\[
\text{Hybridization approaches play a vital role to improving}
\text{GA convergence rate. As reported by Miura et al. [6], for}
\text{most nonlinear optimization problems, the time required by a}
\text{genetic algorithm to converge to an optimal solution can be}
\text{reduced by incorporating some information about the gradients}
\text{of the problem’s variables. Thus, they propose hybridizing GA}
\text{with a steepest descent algorithm and argue that this can}
\text{positively influence the overall convergence rate. Similar}
\text{proposals were made in [7], [8].}
\text{Virtually all the aforementioned approaches share their}
\text{merits and demerits. For instance, use of Hamming distance}
\text{can allow effective analysis of the diversity profile in a given}
\text{population; as convergence sets in, the search space reduces}
\text{to a Hamming landscape [9] of bit-flip neighbours. However,}
\text{we observe that this similarity measure works in the genotype}
\text{space, but lacks any knowledge of the dynamics in the solution}
\text{(i.e., phenotype) space. Thus, depending on the mapping}
\text{technique, closely related individuals in the genotype space}
\text{may differ significantly in the phenotype space. We further}
\text{observe that estimating GA convergence by marrying the two}
diversity measures that monitor similarity in both the encoding
and solution spaces might not totally eliminate dissension
between the methods and could trigger false alarms for con-
vergence. Moreover, the combination can severely increase the}
\text{computational cost of the convergence detection process.}
\text{We therefore argue that since progress in evolutionary}
\text{search is internally governed by the actions of genetic oper-
ators (selection and reproduction operators like crossover and
mutation), building a convergence measure that assesses the}
dynamical roles of these operators could be a promising
alternative to tackling this dilemma.}

\text{III. MONITORING THE EFFECT OF EC OPERATORS WITH}
\text{PRICE’S THEOREM}

\text{In order to examine the independent effect of EC operators}
\text{while in interaction, Price [10] formulated a theorem that}
\text{permits decomposition of the evolutionary process to separate}
\text{the genetic effect (or contribution) of the selection operator}
\text{from that of other reproduction operators (i.e. crossover and
mutation). Although Price’s work was mainly in the field of}
evolutionary genetics, Price’s equation provides greater insight
into general selection processes (1). Price’s theorem states that:}
\text{\[
\Delta Q = \frac{\text{Cov}(z, q)}{\bar{z}} + \frac{\sum_{i=1}^{N} z_i \Delta q_i}{N \bar{z}}
\]}
\text{where \(\Delta Q = Q_2 - Q_1\) is the change in the measured
characteristics \(Q\) which could be fitness, \(N\) is the number of
individuals in the parent population (i.e., population size), \(z_i\)
is the number of offspring of parent \(i\), and \(\bar{z} = \sum z_i / N\) is the
average number of the offspring produced. Also, \(\Delta q_i = q_i' - q_i\),
where \(q_i\) is the fitness of parent \(i\) and \(q_i'\) is the average fitness of
the offspring of parent \(i\). And finally, \(z\) and \(q\) are vectors of
\(z_i\) and \(q_i\) respectively.}

\text{The two terms in the Price’s theorem (1) represent the}
ctribution of different operators as the mean of the character-
istic being measured. Since we deploy the Price’s equation
to analyse convergence in EC, the measured characteristic
\(Q\) in this case will be fitness. The first term represents
the contribution of selection operator while the second term
gives an aggregated contribution of the reproduction operators involved. Notice that the effect due to selection is modelled in terms of the covariance between the individuals \( z \) and their fitness \( q \). This conforms to the Fisher’s fundamental theory of natural selection that relates the change in the mean fitness in a population to the population’s fitness variance. It also agrees with the argument presented by [11] that sees the covariance between the phenotypic values of individuals and their fitness as the cause of differential productivity that leads to the change in phenotype.

The following lemmas elaborate fitness decomposition in evolution using Price’s theorem.

**Lemma 1:**

Supposing the fitness \( q_i \in \mathbb{R} \) of each member of the parent population in (1) is represented by a vector \( q \) such that:

\[
q = [q_1, q_2, ..., q_N]^T
\]

Suppose also the number of offspring \( z_i \in \mathbb{Z} \) produced by each of the \( N \) parents is represented by a vector \( z \) such that:

\[
z = [z_1, z_2, ..., z_N]^T
\]

Then, the following expansion of the two terms in (1) will demonstrate that it is only the first term that represents the contribution of the selection operator.

First term:

\[
\frac{\operatorname{Cov}(z, q)}{\bar{z}} = \frac{\text{mean} \left[ \begin{array}{c}
(z_1 - \mu_{z_1})(q_1 - \mu_{q_1}) \\
(z_2 - \mu_{z_2})(q_2 - \mu_{q_2}) \\
\vdots \\
(z_N - \mu_{z_N})(q_N - \mu_{q_N})
\end{array} \right]}{\bar{z}}
\]

(4)

where \( \mu_{z_i} \) is the mean of the number of offspring of parent \( i \) and \( \mu_{q_i} \) is the mean fitness of the parents.

Second term:

\[
\frac{\sum_{i=1}^{N} z_i \Delta q_i}{N\bar{z}} = \frac{z_1(q_1' - q_1) + \cdots + z_N(q_N' - q_N)}{N\bar{z}}
\]

(5)

Recall that following any typical evolutionary selection process, the resulting offspring have the same fitness as their parents (i.e. selection process adds no new solutions to the population), therefore,

\[
\Delta q_i = q_i' - q_i = 0 : \text{for all } i,
\]

(6)

and the value of (5) sums to zero. Therefore, the contribution of selection operator is only in the first term of (1).

**Lemma 2:**

To investigate the contribution of the reproduction operators, assume any traditional \( n \)-point crossover operator is employed; crossing any two parents yields two offspring. Therefore, if all parents undergo the crossover operation (i.e., when crossover probability \( P_c = 1 \)) and the population size \( N \) is even, then, the number of offspring produced will always be equal to the number of their parents, i.e., \( \forall i : \mu_{z_i} = z_i \). Hence, the first term (4) of the Price’s equation will be equal to zero. This is also true for any bit-flip mutation operation where a parent chromosome yields a single offspring after mutation.

An exception worth noting is that in a situation where not all chromosomes undergo the crossover operation (i.e., when \( P_c < 1 \) or the population size \( N \) is odd), then from (4), \( \exists i : \mu_{z_i} \neq z_i \) and thus:

\[
\frac{\operatorname{Cov}(z, q)}{\bar{z}} \neq 0.
\]

Therefore, unlike the traditional belief [12] where it is often assumed that the contribution of the reproduction operators in the first term of Price’s equation is always zero. This investigation reveals the contrary if taking into account the special cases mentioned above; when \( P_c < 1 \).

Now, expanding the second term as in (5) reveals that the summation is non-zero since the fitness \( q_i \) of any parent \( i \) is often different from that of all its offspring\(^1\), i.e., for most \( i \), \( q_i' \neq q_i \). Hence, the reproduction operators mainly contribute via the second term of Price’s equation, though, they might contribute via the first term as proven above.

\[\square\]

### A. Extension of Price’s Equation

Bassett et al. [13] extend the second term (5) of Price’s equation to allow monitoring the contribution of reproduction operators so as to ascertain which among them is more effective at various stages of the evolution process. It was argued that utilizing fitness averages alone will not guarantee accurate conclusions on the convergence of the evolution since the average individuals are not the major driving force for the evolution. Thus, it becomes essential to dynamically analyse the fitness variance throughout the population paying attention to the exceptional (best and worst) individuals mainly created by the reproduction operators.

The extended Price’s equation can be defined as:

\[
\Delta Q = \frac{\operatorname{Cov}(z, q)}{\bar{z}} + \sum_{j=1}^{k} \frac{\sum_{i=1}^{N} z_i \Delta q_{ij}}{N\bar{z}}, k = 1, 2, ...
\]

(7)

where \( k \) is the number of genetic operators; \( q_{ij} \) is the average value of the fitness of all the offspring of parent \( i \) after the application of operator \( j \); \( \Delta q_{ij} = q_{ij}' - q_{ij} \) is the difference between the average fitness \( q \) of the offspring of parent \( i \) measured before and after the application of operator \( j \).

For the proposed EC algorithm, we deploy only the crossover and mutation operators, thus, \( k = 2 \). Therefore, the following extension (8) to the Price’s equation contains only three terms; a term for the selection, crossover and mutation respectively. Each of these terms estimates the changes in the mean of the population’s fitness (\( \Delta Q \)) due to one of the three genetic operators.

\[
\Delta Q = \frac{\operatorname{Cov}(z, q)}{\bar{z}} + \frac{\sum_{i=1}^{N} z_i \Delta q_{i \text{Crossover}}}{N\bar{z}} + \frac{\sum_{i=1}^{N} z_i \Delta q_{i \text{Mutation}}}{N\bar{z}}
\]

(8)

\(^1\)Although it is possible for any or all the offspring resulting from crossover to have equal fitness as their parent, this is rare and usually only occur when the population’s diversity collapses. Thus it signals convergence.
TABLE I
GA Parameter Settings for Investigating the Proposed Extended Price’s Equation.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Symbol</th>
<th>Typical Values/Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Size</td>
<td>N</td>
<td>50 – 100</td>
</tr>
<tr>
<td>Representation</td>
<td>–</td>
<td>Binary encoding</td>
</tr>
<tr>
<td>Selection Scheme</td>
<td>–</td>
<td>RWS and a BTS method without replacement</td>
</tr>
<tr>
<td>Crossover Probability</td>
<td>$P_c$</td>
<td>1.0 or 100%</td>
</tr>
<tr>
<td>Mutation Probability</td>
<td>$P_m$</td>
<td>1/len where len = string length</td>
</tr>
<tr>
<td>Replacement Scheme</td>
<td>–</td>
<td>An overlapping elitist strategy</td>
</tr>
<tr>
<td>Generation gap</td>
<td>$G$</td>
<td>5% of the population size $N$</td>
</tr>
<tr>
<td>Termination Criteria</td>
<td>MaxGen</td>
<td>Maximum generations limit 100</td>
</tr>
</tbody>
</table>

IV. Analyzing Evolution Progress with Extended Price’s Equation

In order to analyse the effect of the proposed extension (8) to the Price’s theorem, we conduct a number of experiments with a GA having mostly standard parameters [14] as shown in table I. Various global optimization test problems are used for these experiments. However, due to space limitations, the results reported here\(^2\) are those obtained on simulating the solution process of (i) a 2-dimensional Schwefel function (9):

$$f(x) = \sum_{i=1}^{n} x_i \sin \left( \sqrt{|x_i|} \right); \quad -5.0 \leq x_i \leq 5.0; \quad n = 2; \quad (9)$$

and (ii) a 100 dimensions of Rastrigin function defined as:

$$f(x) = 10 \cdot n + \sum_{i=1}^{n} \left( x_i^2 - 10 \cdot \cos(2\pi x_i) \right) \quad (10)$$

where $-5.12 \leq x_i \leq 5.12; \quad n = 100$. Functions (9) and (10) are both multimodal and highly dispersed [16], [17] maximization benchmark problems.

The objectives for these experiments are:

- To observe and analyse the growth in the fitness of the population of candidate solutions as the evolution progresses; and
- To investigate the change in fitness that can be attributed to the selection operator, and changes due to the crossover and the mutation operator respectively.

Ultimately, the outcome of this investigation is aimed at providing new insight through observation of the manner in which the selection and reproduction operators collectively move the evolutionary search forward. It will also give insight on their effect on ensuring sustainable balance for the exploitation and exploration of problem search space, making GAs effective and promising global search mechanisms.

Experiment 1: Visualizing Evolution with Roulette Wheel Selection (RWS)

In this experiment, RWS is used to investigate the independent effects of the EC operators over the entire period of evolution. Table II shows the parameter sets for this experiment.

Experiment 2: Visualizing Evolution with Binary Tournament Selection (BTS)

Here a BTS method without replacement is used to investigate the effects of the EC operators during the evolution, (see [18] for detailed comparison of various selection methods). Similarly, parameters in table II are used.

Notice from table II that there are six parameter sets for testing each of the two selection methods. The first half will be tested with a population size of 50 while for the other half, a population size of 100 will be used. Though with the parameter sets in tables I and II, a population size of 50 is fairly sufficient for the proposed GA to converge to the optimum of the 2-dimensional Schwefel problem (9), the motive here is to investigate whether changes in population size has any significant effect on the effect of the EC operators during evolution. Note that maximum generation limit was 100 and all results are averages of 100 runs.

A. Results Analysis: Comparing RWS and BTS methods

The results obtained for the above experiments 1 and 2 are summarized in table III. It compares RWS and BTS methods from two perspectives. First, on their performance under two different population sizes $N$. Second, on their performance against three different settings for $P_c$ and $P_m$.

Notice from table III that for the two different population sizes, the results obtained with the BTS clearly outperform

\(^2\) Similar results are obtained with many other global optimization benchmark problems of various dimensions (cf. [15]), thus, omitted.
and for further exploration of the problem space. This however may increase wasteful function evaluations.

Therefore, setup (i) is the best candidate for achieving the proposed objective. The fitness comparison plots in Fig. 2(a and c) compare the best, worst and the average fitness. The Price’s plots (Fig. 2(b and d)) on the other hand depict the changes in the averages of the population’s fitness due to each of the three genetic operators according to the definitions in (8), i.e., they show the average contributions of the selection, crossover and mutation operators to the fitness progress over the entire generations. All results are averages of 100 independent runs.

Discussion and Analysis of Results for Experiment 1: RWS:
A quick glimpse at the curves in the Price’s plot in Fig. 2(b) for the \( \Delta q \) contribution of operators barely allows any meaningful conclusion. One might naively infer that the idea of decomposing the fitness progress to investigate the effect of the genetic operators provides no vital information on the evolution dynamics. However, we would like to make the following observations:

1) It can be noticed from Fig. 2(b) that the curves for all the three operators fluctuate above and below zero on the fitness axis in a somewhat random manner. Thus, the plot provides no vital information as to which among the three operators ensures continuous growth in fitness (i.e. exploitation) and which is responsible for maintenance of diversity in the population by exploring other areas of the search space.

2) A careful look at the initial part of the curve for the selection operator in Fig. 2(b) reveals a huge selection pressure. This corresponds to the steep gradient in the best and average fitness curves (Fig. 2(a)) during the early generations. However, the pressure suddenly drops after around the first 10 generations and thereafter, it continues to fluctuate within a somewhat uniform range. Consequently, the roles of the crossover and selection operator throughout this period remain quite undifferentiated. It is therefore hard to identify which among the two is responsible for exploitation or exploration and at what stage of the evolution does it contribute most.

3) Another crucial observation is that because all the curves for the three operators in Fig. 2(b) continuously effect changes in the population’s fitness, the population continues to slowly evolve (see Fig. 2(a)) thereby growing the average fitness (although at a slower rate) till the evolution is terminated. This unwanted phenomenon thwarts any possibility of detecting the precise moment at which the evolution converges.

Nevertheless, the above behaviour of the selection operator agrees with the fact that RWS is an FPS method and it suffers inherent stochastic sampling errors\(^3\) which increases as the population average fitness grows. Thus, it is this phenomenon that led to the sudden collapse in the selection pressure just after the very early generations.

Moreover, as can be seen from the plot in Fig. 2(b), the contribution of the selection operator to the fitness progress

\(^3\)It benefits from the high mutation rate of 0.05, which increase the chances for further exploration of the problem space. This however may increase wasteful function evaluations.

\(^4\)Some selection methods like FPS due to being stochastic may lose the best member of the population by simply not picking it for any contest.
becomes quite similar to that of the crossover and mutation operators. In other words, all the three operators play the role of exploration and exploitation of the search space. This effect may not be suitable for optimization purpose where the ultimate aim is to explore and as much as possible converge to the optimum solution point.

Discussion and Analysis of Results for Experiment 2: BTS: The plots shown in Fig. 2(c and d) convey similar information to those in experiment 1 above but now BTS is employed. Yet, observations of these plots lead to the following deductions:

1) A simple observation of the fitness curves in Fig. 2(c) reveals that both the average and the best fit individuals in the population rapidly grow with a steep gradient to their peak values within the very early generations of the evolution. Notice how at the same time useful diversity is maintained in the population by retaining a reasonable amount of the worst fit individuals throughout the evolution. This is a clear manifestation of achieving proper separation of roles among the genetic operators as will be analysed shortly. Hence, while the BTS operator is undertaking its role of exploitation, mutation is ensuring effective exploration and the crossover operator serves as a regulator by performing both roles.

2) Notice from Fig. 2(d) that the curve for the selection operator is always above zero on the fitness axis. Thus, selection can now be seen as a biased process that primarily guides the search towards the promising regions of the search space thereby increasing the population’s average fitness. We therefore acknowledged that this behaviour conforms to the characteristics of a typical BTS method [19], and thus, it is as expected.

3) Notice also from the same figure that the curve for the crossover operator swings above and below zero on the fitness axis. This is indicative of the fact that the operator has both exploration and exploitation effects, i.e., while crossover improves the population’s average fitness via exploitation, it also lowers it during exploration. It can be observed that the effect of crossover lessens over generations, and eventually, neither its exploitation nor its exploration effect seem to influence the fitness growth as convergence sets in.

4) Contrary to the previous operators, notice that the curve for the mutation operator always lie beneath zero on the fitness axis (see Fig. 2(d)). This is because mutation operator improves population’s diversity by exploring other parts of the search space. But in so doing, it tends to drastically lower the fitness growth in the population. Moreover, the manner in which mutation affects the fitness growth is unaffected by convergence. Thus, the effect of mutation on the fitness progress remains fairly constant over the entire run.

Following the above observations, it is thought that among all the three genetic operators, it is the behaviour of the crossover operator with regard to the fitness growth in the population that is most crucial to effective detection of convergence in the evolution. This is because both the mutation and selection operators continue to evolve throughout the evolution. Moreover, as highlighted in item (2) above, the selection operator is a mere biased process that favours highly fit solutions. It adds no new solutions into the population and is therefore not suitable for convergence detection.

Finally, for the same set of parameters (test setup (i)), it is interesting to notice how the best and average fitness curves in the fitness comparison plot in Fig. 2(c) (where BTS is used) clearly outperform those in Fig. 2(a) (where RWS is used) by converging rapidly towards the global optimum solution of 838. In the following section an investigation on how the crossover can be used to detect convergence will be presented.

V. USING EXTENDED PRICE’S EQUATION TO MEASURE CONVERGENCE

Recall from (8) that $\Delta Q$ is only the collective change in the average fitness due to the three genetic operators. As argued by [13] observing changes in averages alone does not convey sufficient information on the true effect of an operator. This is evident from the plot of the crossover operator in Fig. 2(d) where it is hard to appreciate its effect on the fitness even during the early generations of the evolution. Therefore monitoring the spread in fitness in the population by exploring the best and worst individuals produced by an operator is a viable alternative that could yield better insight. The spread can be measured by evaluating the change in fitness variance in the population due to a given genetic operator.

From (8), let the change in fitness due to a genetic operator $j$ (crossover in this case) be:

$$ \Delta Q_{ij} = \frac{\sum_{i=1}^{N} \Delta q_{ij}}{N}.$$  

Then, since $\Delta q_{ij} = q'_{ij} - q_{ij}$ is the random variable of interest, the expectation and variance of the crossover term can be obtained by respectively taking the first and second moment of (11) with respect to the $\Delta q_{ij}$ values, such that:

$$ E[\Delta q_{ij}] = \frac{\sum_{i=1}^{N} \sum_{z_i} \Delta q_{ij}}{N \bar{z}} = \frac{\sum_{i=1}^{N} \sum_{z_i} \Delta q_{ij}}{N \bar{z}} $$

$$ Var[\Delta q_{ij}] = E[\Delta q_{ij}^2] - [E(\Delta q_{ij})]^2 $$

where $\sum_{z_i}$ is the sum over all the offspring of parent $i$. Typically, there will be two offspring for each parent when $P_c = 1.0$. All other parameters are as previously defined in section III. The variance, $Var[\Delta q_{ij}]$ is:

$$ Var[\Delta q_{ij}] = E[\Delta q_{ij}^2] - [E(\Delta q_{ij})]^2 $$

and the standard deviation is:

$$ \sigma[\Delta q_{ij}] = \sqrt{Var[\Delta q_{ij}]}.$$

Now, overlaying the curves for the one standard deviation interval ($\pm \sigma$) on the plot of the $\Delta Q$ changes in the average fitness due to the crossover operator yield the envelope of the shaded areas shown in Fig. 3(a and c). The plots give a better impression of the distribution of the effect of this operator on
the fitness progress over generations. A similar technique was used to plot the effect of mutation operator (Fig. 3(b and d)).

Fig. 3(a and b) show the results for experiment 1 where RWS is used. The shaded area above the curve for the change in the average fitness due to crossover in Fig. 3(a) reveal that crossover operator does indeed contribute to the fitness growth, but at the same time, the shaded area underneath it which lie under zero on the fitness axis indicates how much the operator contributed to producing low fit individuals. Notice how the ±σ (standard deviation) envelopes for both the crossover and mutation operators (Fig. 3(a and b)) remain uniformly constant throughout the period of the evolution. Hence, when RWS is utilized, the effects of both crossover and mutation on fitness growth continue till evolution is terminated.

Similarly, Fig. 3(c and d) show the results for experiment 2 where BTS is utilized. From the crossover plot (c) it is interesting to note how the ±σ crossover envelope shrinks towards zero (see the point labelled A) as the curve for the change in the average fitness settles around zero on the fitness axis. This indicates that beyond point A, the crossover operator barely contribute to the fitness progress. Hence, it shows how the use of BTS method facilitates monitoring the effect of crossover to effectively detect loss of diversity in the population. It also justifies our previous supposition that monitoring crossover operator can allow sufficient convergence detection in EC.

On the other hand, an observation of Fig. 3(d) reveals that the curve for the change in the average fitness due to mutation operator lie and remain beneath zero on the fitness axis throughout the evolution. Also, larger portion of the shaded area for the ±σ mutation envelope is beneath zero. This is contrary to its behaviour when RWS is used (see Fig. 3(b)) where mutation lacks precise role. Nevertheless, the ±σ mutation envelopes in both Fig. 3(b and d) remain virtually uniform (without shrinking) throughout the evolution. This means the effect of mutation on the fitness of the population continues for the entire evolution run. Therefore, it cannot be deployed to conduct any convergence detection.

A. The proposed convergence threshold parameter

Having used one standard deviation interval to analyse the effect of genetic operators, then, the width of the ±σ envelope for the effect of an operator j (crossover in this case) on the fitness growth at every kth generation lies within the interval:

\[
[\Delta Q_{jk} - \sigma_{jk}, \Delta Q_{jk} + \sigma_{jk}]
\]

where \(\Delta Q_{jk}\) is the change in the average fitness of the population at iteration k due to operator j and \(\sigma_{jk}\) is the corresponding standard deviation. Let the width be represented by \(\sigma q_{jk}\), then it can be determined as follows:

\[
\sigma q_{jk} = (\Delta Q_{jk} + \sigma_{jk}) - (\Delta Q_{jk} - \sigma_{jk}) = 2 \sigma_{jk}.
\]

This means mutation operator particularly ensures maintenance of diversity in the population by producing lower fit individuals when BTS is used.
Consider the simulation for the solution of Rastrigin function (100-Dimensions) (10) using the setup in experiment 2 (for BTS) shown in Fig. 4. This simulation closely resembles that of the previously seen Schwefel function under the same test setup (see Fig. 2 and 3 (c and d)). The point labelled A on Fig. 4(a) directly corresponds to the point labelled A on Fig. 4(c) when read from the x-axis (i.e., the generations axis). It is fairly easy to notice the correlation between the ±σ crossover envelope and the gradient of the best and average fitness curves, i.e., the more the width of the ±σ crossover envelope shrinks to zero (Fig. 4(c)), the flatter the gradient of the best and average fitness curves in Fig. 4(c). Hence, the width σq₂jk in (16) indirectly represent the available diversity in the population, and the generation at which σq₂jk tends to zero signifies the beginning of convergence in the evolution.

Therefore, the proposed convergence measure is to prescribe a threshold value for the parameter σq₂jk such that whenever σq₂jk falls below this threshold, the search process is automatically terminated. We must remark here that this threshold parameter is user defined and its appropriate value is determined empirically, preliminary investigations reveal that a value of σq₂jk ≤ 0.01 is suitable for crossover and mutation probabilities of Pc = 1.0 and Pm = 1/l respectively (see [15] for details). Nonetheless, since determining optimum value for this threshold parameter relies on some other EC parameters, a thorough sensitivity analysis will be conducted in the future.

VI. CONCLUSION

This study provides additional insight on the relevance of convergence detection in evolutionary computation (EC) and sheds light on some conventional convergence measures. Most importantly, it provides insight on the interactions among the genetic operators during the evolutionary search process. A visual means for investigating the independent roles of genetic operators on fitness progress is developed using the extended Price’s equation.

The experimental results obtained after comparing the plots for the contribution of the genetic operators using roulette wheel selection (RWS) and binary tournament selection (BTS) reveal two interesting findings.

First, with RWS method, the proposed technique for visualizing the effect of genetic operators using extended Price’s equation have shown that all the three operators play random roles of exploiting and exploring the search space throughout the evolution period. This has made it infeasible to deploy their individual effect as a means to measure convergence.

Second, substituting the RWS with a BTS method reveals the contrary. The plots for the extended Price’s equation in this case demonstrate a clear separation of roles among the three genetic operators. The BTS operator takes the lead in exploiting the highly fit areas of the search space; the mutation operator handles the exploration aspect while the crossover operator serves as a moderator between the two. This has made it possible to monitor and use the crossover’s effect as an effective means of detecting convergence in the evolution.

Consequently, the insight gathered above led to the development of a novel convergence measure (section V) that can allow automatic convergence detection in EC. Because this investigation augments the previous works on termination criteria for evolutionary algorithms by analysing the processes that could lead to successful termination of the evolutionary search, it lays a platform that will facilitate hybridizing evolutionary algorithms with other algorithms such as gradient-based local search methods. Further work in this direction could be promising since the design of a future hybrid scheme would require no user imposed maximum generation limit on the evolution as the algorithm can heuristically detect convergence and decide when best to halt itself.

ACKNOWLEDGEMENT

H. Bashir is funded by the Petroleum Technology Development Fund (PTDF), Nigeria. The authors would like to thank the anonymous reviewers for their valuable and insightful comments.

REFERENCES