PRIFYSGOL
ABERYSTWYTH
UNIVERSITY

Aberystwyth University

An analytic expression of relative approximation error for a class of evolutionary algorithms

He, Jun

An Analytic Expression of Relative Approximation Error for a Class of Evolutionary Algorithms

Jun He

Abstract—An important question in evolutionary computation is how good solutions evolutionary algorithms can produce. This paper aims to provide an analytic analysis of solution quality in terms of the relative approximation error, which is defined by the error between 1 and the approximation ratio of the solution found by an evolutionary algorithm. Since evolutionary algorithms are iterative methods, the relative approximation error is a function of generations. With the help of matrix analysis, it is possible to obtain an exact expression of such a function. In this paper, an analytic expression for calculating the relative approximation error is presented for a class of evolutionary algorithms, that is, (1+1) strictly elitist evolution algorithms. Furthermore, analytic expressions of the fitness value and the average convergence rate in each generation are also derived for this class of evolutionary algorithms. The approach is promising, and it can be extended to non-elitist or population-based algorithms too.

I. INTRODUCTION

Evolutionary algorithms (EAs) have been widely used to find good solutions to hard optimization problems. Many experimental results claim that EAs can obtain good quality solutions quickly. Nevertheless, from the viewpoint of the NP-hard theory, no efficient algorithm exists for solving NP-hard combinatorial optimization problems at the present and possibly for ever. Therefore it is unlikely that EAs are efficient in solving hard combinatorial optimization problems too. Instead of searching for the exact solution to hard optimization problems, it is more reasonable to expect that EAs are able to find some good approximate solutions efficiently.

It is necessary to answer the question of how good solutions EAs can produce to hard optimization problems in terms of the approximation ratio. Current work focuses on the approximation ratio of the solution found by an EA within polynomial time. The research has attracted a lot of interests in recent years. Various combinatorial optimization problems have been investigated, including the minimum vertex cover problem [1], [2], the partition problem [3], the set cover problems [4], the minimum label spanning tree problem [5], and many others.

This paper studies the approximation ratio of EAs from a different viewpoint. It aims to estimate the relative approximation error of the best solution found by an EA in each generation, but without considering whether the EA is an approximation algorithm or not. The problem in this paper is described as follows: Given an EA for maximizing a fitness function $f(x)$, let f_{opt} be the optimal fitness and F_t the expected fitness value of the best solution found in the tth

Jun He is with Department of Computer Science, Aberystwyth University, Aberystwyth SY23 3DB, United Kingdom, Email: jun.he@aber.ac.uk

generation. The approximation ratio of the tth generation solution is F_t/f_{opt}. The approximation ratio of the optimal solution is 1. The relative approximation error is

$$E_t = 1 - \frac{F_t}{f_{\mathrm{opt}}}. \qquad (1)$$

In [6], E_t is called the performance ratio. In order to avoid confusion with the approximation ratio, it is renamed the relative approximation error. The relative approximation error E_t is a function of t. Our main research question is to find an upper bound $\beta(t)$ on the error E_t.

The perfect answer is to obtain a function $\beta(t)$ in a closed form such that $E_t = \beta(t)$. For (1+1) strictly elitist EAs, such an analytic expression has been constructed in this paper using matrix analysis. To the best of our knowledge, this is the first result of expressing the relative approximation error (also the fitness value and the average convergence rate) in a closed form for a class of EAs.

The paper is arranged as follows: Section II reviews the links to related work. Section III defines the relative approximation error. Sections IV and VII conduct a case study. Section V introduces Markov modelling. Section VI makes a theoretical analysis. Section IX summarizes the paper.

II. LINKS TO RELATED WORK

The relative approximation error belongs to the convergence rate study of EAs, which can be traced back to 1990s [7]–[9]. This paper only investigates EAs for discrete optimisation, although the converegnce rate of EAs for continuous optimization [10]–[12] is also important. EAs belong to iterative methods. A fundamental question in iterative methods is the convergence rate, which can be formalised as follow [9]. Since the tth generation solution is a random variable, we let $\mathbf{p}_t$ be a vector representing its probability distribution over the search space, $\boldsymbol{\pi}$ a vector such that $\pi(\mathcal{S}_{\mathrm{opt}}) = 1$ for the optimal solution set $\mathcal{S}_{\mathrm{opt}}$ and $\pi(\mathcal{S}_{\mathrm{non}}) = 0$ for the non-optimal solution set $\mathcal{S}_{\mathrm{non}}$. The convergence rate problem asks the question how fast $\mathbf{p}_t$ converges to $\boldsymbol{\pi}$. The goal is to obtain a bound $\beta(t)$ such that $\| \mathbf{p}_t - \boldsymbol{\pi} \| \leq \beta(t)$, where $\| \cdot \|$ is a norm. There are various ways to assign the norm. For example, if the tth generation solution is a binary string $x_1 \cdots x_n$ and the optimal solution is $1 \cdots 1$, the norm is set to be the Hamming distance:

$$\| \mathbf{p}_t - \boldsymbol{\pi} \| = \sum_i |1 - x_i|. \qquad (2)$$

In the current paper, the norm is set to be the relative approximation error $\| \mathbf{p}_t - \boldsymbol{\pi} \| = 1 - F_t/f_{\mathrm{opt}}$.

According to [13], there are two approaches to analyse the convergence rate of EAs for discrete optimization. The first

approach is based on the eigenvalues of the transition submatrix associated with an EA. Suzuki [7] derived a lower bound of convergence rate for simple genetic algorithms through analysing eigenvalues of the transition matrix. Schmitt and Rothlauf [14] found that the convergence rate is determined by the second largest eigenvalue of the transition matrix. The approach used in the current paper is the same as that in [7], [14]. All are based on analysing the powers and eigenvalues of the transition matrix. The other approach is based on Doeblin's condition [9], [15]. Using the minorisation condition in Markov chain theory, He and Kang [9] proved that for the EAs with time-invariant genetic operators, the convergence rate can be upper-bounded by ϵ^t where $\epsilon \in (0, 1)$.

The research in this paper is also linked to fixed budge analysis. Jansen and Zarges [16], [17] proposed fixed budget analysis. It aims to find lower and upper bounds $\beta_{low}(b)$ and $\beta_{up}(b)$ such that $\beta_{low}(b) \le f_b \le \beta_{up}(b)$ usually for a fixed budget b. They investigated two algorithms: random local search and the $(1+1)$ EA and obtained such bounds on several pseudo-Boolean optimization problems.

The convergence rate study can implement the same task as fixed budget analysis does. Provided that $1 - f_b/f_{opt} \le \beta(b)$, it is trivial to derive $f_b \ge f_{opt}(1 - \beta(b))$. Nevertheless, there are significant differences between the convergence rate study and fixed budget analysis.

- The convergence rate study has existed in EAs for more than two decades [7]–[9]. Fixed budget analysis was recently proposed by Jansen and Zarges [17].
- The convergence rate study focuses on estimating the error $\| \mathbf{p}_t - \boldsymbol{\pi} \|$, where the norm $\| \cdot \|$ can be chosen as the absolute error $f_{opt} - F_t$, relative error $1 - F_t/f_{opt}$ or Hamming distance. Fixed budget analysis aims at bounding f_b for a fixed budget b (that is a fixed number of generations) [17].
- In the convergence rate study, the upper bound $\beta(t)$ on $\| \mathbf{p}_t - \boldsymbol{\pi} \|$ usually is an exponential function or combination of linear functions of t [9]. In fixed budget analysis, the bound $\beta(b)$ on f_b may not be an exponential function of b [17].
- In the convergence rate, the bound on $\| \mathbf{p}_t - \boldsymbol{\pi} \|$ holds for all t. But in fixed budget analysis, the bound on f_b often is estimated for a fixed budget b [17].
- Matrix analysis is widely used in the convergence rate study [7], [14], but it is not used in fixed budget analysis.

III. RELATIVE APPROXIMATION ERROR, FITNESS VALUE AND AVERAGE CONVERGENCE RATE

Consider a maximization problem, that is, $\max\{f(x); x \in \mathcal{S}\}$ where $\mathcal{S}$ is a finite set, and $f_{opt} > f(x) \ge 0$. For the sake of analysis, let $\mathcal{S} = \{0, 1, \cdots, L\}$ denote the set of all solutions. We assume that

$$f_{\max} = f(0) > f(1) \ge \cdots \ge f(L) = f_{\min}.$$

$\mathcal{S}$ is split into two subsets: the optimal solution set $\mathcal{S}_{opt} = \{0\}$ and the set of non-optimal solutions $\mathcal{S}_{non} = \{1, \cdots, L\}$.

An EA for solving the above problem is regarded as an iterative procedure: initially construct a population of solutions Φ_0; then given the tth generation population Φ_t, generate a new population Φ_{t+1} in a probabilistic way. This procedure is repeated until an optimal solution is found. This paper investigates a class of $(1+1)$ elitist EAs which are described in Algorithm 1. This kind of EAs is very popular in the theoretical analysis of EAs.

Algorithm 1 A (1+1) Strictly Elitist EA

1: set $t \leftarrow 0$;
2: $\Phi_0 \leftarrow$ choose a solution from $\mathcal{S} = \{0, 1, \cdots, L\}$;
3: **while** Φ_t is not an optimal solution **do**
4:     $\Psi_t \leftarrow$ mutate Φ_t;
5:     **if** $f(\Psi_t) > f(\Phi_t)$ **then**
6:         $\Phi_{t+1} \leftarrow$ select Ψ_t;
7:     **else**
8:         $\Phi_{t+1} \leftarrow$ select Φ_t;
9:     **end if**
10:    $t \leftarrow t + 1$;
11: **end while**

The expression $\Phi_t = x$ means the tth generation individual Φ_t at state x where Φ_t is a random variable and x its value taken from $\mathcal{S}$. The fitness of Φ_t is denoted by $f(\Phi_t)$. Since $f(\Phi_t)$ is a random variable, we consider its expectation $F_t \overset{\text{def}}{=} \mathrm{E}[f(\Phi_t)]$. The approximation ratio of the tth generation individual is F_t/f_{opt}. The approximation ratio of the optimal solution is 1.

Definition 1: The relative approximation error of the tth generation individual is defined by

$$E_t = 1 - \frac{F_t}{f_{opt}}. \tag{3}$$

There is a link between the relative approximation error and the fitness value. From the definition of the relative approximation error, we know that the fitness value in the tth generation equals to

$$F_t = f_{opt}(1 - E_t). \tag{4}$$

There is a link between the relative approximation error and average convergence rate [18]. From the definition of the (geometric) average of the convergence rate of an EA for t generations, we get

$$R_t \overset{\text{def}}{=} 1 - \left(\left| \frac{f_{opt} - F_t}{f_{opt} - f_0} \right| \right)^{1/t} = 1 - \left(\frac{E_t}{E_0} \right)^{1/t}. \tag{5}$$

IV. EXAMPLE

Given F_t, E_t and R_t, which is the best option to measure the performance of an EA? We use a simple experiment to show their advantage and disadvantage. Consider the problem of maximizing a pseudo-Boolean function $f(x)$ where $x = x_1 \cdots x_n$ is a binary string. Three test functions are used in the experiment.

$$\text{OneMax function } f_{one}(x) = |x|,$$
$$\text{square function } f_{squ}(x) = |x|^2,$$
$$\text{logarithmic function } f_{log}(x) = \ln(|x| + 1),$$

where $|x| = x_1 + \cdots + x_n$. A (1+1) EA is used for solving the optimisation problem. This EA is also called randomised local search.

Onebit Mutation. Given a binary string, chose one bit at random and then flip it.

Elitist Selection. Choose the best from the parent and child as the next parent.

Onebit mutation is chosen for the sake of demonstrating that the average convergence rate R_t may equal to a constant, according to the theory of the average convergence rate [18]. The three functions are the easiest to the (1+1) EA among all pseudo-Boolean functions whose optimum is unique at $1 \cdots 1$ according to the theory of the easiest and hardest functions [19].

In the experiment, we set $n = 4$. This small value is chosen for the sake of displaying matrices in Section VII in one column. The initial solution is set to 0000. We run the EA 10^8 times. The EA stops after 35 generations for each run. Fig. 1 demonstrates the fitness value F_t which is averaged over 10^8 runs. The figure shows that F_t converges to 4 on f_{one}, 16 on f_{squ} and $\ln 5$ on f_{log}. But it is not clear how F_t is close to f_{opt}, and how fast F_t converges to f_{opt}.
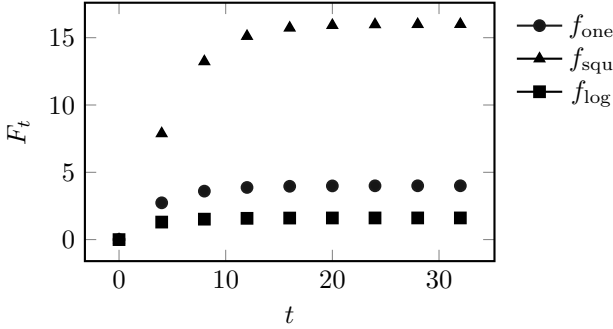

Fig. 1. Fitness value F_t.

Fig. 2 presents the relative approximation error E_t, which converges to 0. From the figure, we observe that for any t, E_t on f_{log} is smaller than that on f_{one}, then smaller than that on f_{squ}.
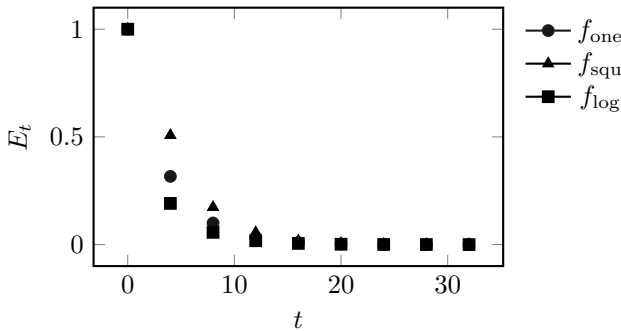

Fig. 2. Relative approximation error E_t.

Fig. 3 illustrates the average convergence rates R_t, which converges to 0.25. From the figure, we see the difference of the average convergence rate on the three functions.

- $R_t = 0.25$ on f_{one}. The EA converges as fast as an exponential decay: $E_t = 0.75^t E_0$.
- R_t converges to 0.25 on f_{squ} but its value is larger than 0.25 . The EA converges faster than the exponential decay: $E_t \leq 0.75^t E_0$.
- R_t converges to 0.25 on f_{log} but its value is smaller than 0.25 . The EA converges slower than the exponential decay: $E_t \geq 0.75^t E_0$.
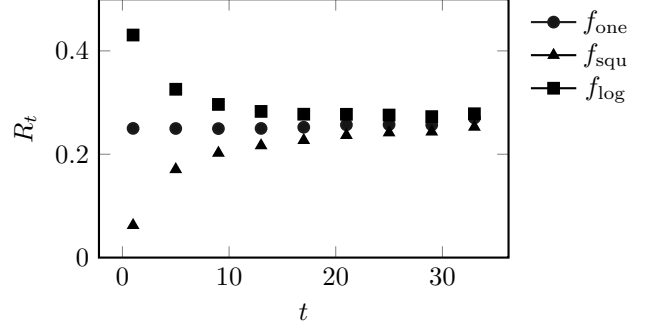

Fig. 3. Average convergence rate R_t.

V. MARKOV CHAIN MODELLING FOR (1+1) STRICTLY ELITIST EAS

This section introduces Markov chain modelling for (1+1) strictly elitist EAs. It follows the Markov chain framework described in [18], [20].

Genetic operators in EAs can be either time-invariant or time-variant [9], [21]. This paper only considers time-invariant operators. Such an EA can be modelled by a homogeneous Markov chain with transition probabilities

$$r_{i,j} \overset{\text{def}}{=} \Pr(\Phi_{t+1} = i \mid \Phi_t = j), \quad i, j \in \mathcal{S}.$$

According to the strictly elitist selection, transition probabilities satisfy

$$r_{i,j} = \begin{cases} \geq 0, & \text{if } f(i) > f(j), \\ \geq 0, & \text{if } i = j, \\ = 0, & \text{otherwise.} \end{cases} \quad (6)$$

Let $\mathbf{R}$ denote the transition submatrix which represents transition probabilities among non-optimal states $\{1, \cdots, L\}$. It is a $L \times L$ matrix, given as follows:

$$\mathbf{R} = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,L-1} & r_{1,L} \\ 0 & r_{2,2} & r_{2,3} & \cdots & r_{2,L-1} & r_{2,L} \\ 0 & 0 & r_{3,3} & \cdots & r_{3,L-1} & r_{3,L} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & r_{L,L} \end{pmatrix}. \quad (7)$$

Let $p_t(i) = \Pr(\Phi_t = i)$ denote the probability of Φ_t at state i and the vector

$$\mathbf{q}_t \overset{\text{def}}{=} (p_t(1), p_t(2), \cdots, p_t(L))^T.$$

Here notation $\mathbf{v}$ is a column vector and $\mathbf{v}^T$ the row with the transpose operation.

For any $t \geq 1$, the probability $p_t(i)$ (where $i \in \mathcal{S}_{\mathrm{non}}$) equals to

$$\Pr(\Phi_t = i) = \sum_{j \in S_{\mathrm{non}}} \Pr(\Phi_t = i \mid \Phi_t = j) \Pr(\Phi_{t-1} = j)$$
$$= \sum_{i \in S_{\mathrm{non}}} p_{t-1}(i) r_{i,j}.$$

It can be represented by matrix iteration

$$\mathbf{q}_t = \mathbf{R}\mathbf{q}_{t-1} = \mathbf{R}^t \mathbf{q}_0. \tag{8}$$

Let $e(i) = f_{\mathrm{opt}} - f(i)$ denote the fitness error between the optimal solution and each non-optimal solution and the vector

$$\mathbf{e}^T \stackrel{\text{def}}{=} (e(1), e(2), \cdots, e(L)).$$

Then the relative approximation error E_t can be represented by

$$E_t = \frac{\mathbf{e}^T \mathbf{q}_t}{f_{\mathrm{opt}}} = \frac{\mathbf{e}^T \mathbf{R}^t \mathbf{q}_0}{f_{\mathrm{opt}}}. \tag{9}$$

From formula (9), we see that E_t is determined by the initial distribution $\mathbf{q}_0$, matrix power $\mathbf{R}^t$, fitness error $\mathbf{e}^T$ and optimal fitness value f_{opt}. Only $\mathbf{R}^t$ is a function of t, so it plays the most important role in determining the relative approximation error.

VI. AN ANALYTIC EXPRESSION OF RELATIVE APPROXIMATION ERROR

This section gives an analytic expression of the relative approximation error for (1+1) strictly elitist EAs. The analysis is based on an existing result in matrix analysis [22], [23].

From (9), we see that calculating E_t becomes a mathematical problem of expressing the matrix power $\mathbf{R}^t$ once the initial probability distribution $\mathbf{q}_0$ and the fitness error $\mathbf{e}^T$ are known. For (1+1) strictly elitist EAs, the matrix $\mathbf{R}$ is an upper triangular, and then it is feasible to express matrix $\mathbf{R}^t$ explicitly in terms of its entries in a closed form [22].

For the sake of simplicity, matrix $\mathbf{R}$ is assumed to satisfy the following condition:

- **Unique condition:** transition probabilities $r_{i,i} \neq r_{j,j}$ if $i \neq j$.

If transition probabilities $r_{i,i} = r_{j,j}$ for some $i \neq j$, a similar discussion can be conducted but will be given in a separate paper.

Definition 2: The power factors of $\mathbf{R}$, $[p_{i,j,k}]$ (where $i, j, k = 1, \cdots, L$), are recursively defined as follows:

$$p_{j,j,j} = r_{j,j}, \tag{10}$$
$$p_{i,j,k} = 0, \qquad k < i \text{ or } k > j, \tag{11}$$
$$p_{i,j,k} = \frac{\sum_{l=k}^{j-1} p_{i,l,k} r_{l,j}}{r_{k,k} - r_{j,j}}, \qquad i \leq k < j, \tag{12}$$
$$p_{i,j,j} = r_{i,j} - \sum_{l=i}^{j-1} p_{i,j,l}, \qquad i < j. \tag{13}$$

Lemmas 1 and 2 show how to calculate the matrix power $\mathbf{R}^t$. For the sake of completeness, their proofs [22] are given here.

Lemma 1 (Lemma 1.2 in [22]): Let $\mathbf{R} = [r_{i,j}]$ be a non-singular upper triangular matrix with unique diagonal entries. Denote the entries of the matrix power $\mathbf{R}^t$ by $[r_{i,j|t}]$. For any $t \geq 1$, if $r_{i,j|t} = \sum_{k=i}^{j} p_{i,j,k}(r_{k,k})^{t-1}$, then $r_{i,j|t+1} = \sum_{k=i}^{j} p_{i,j,k}(r_{k,k})^t$.

Proof: Since $\mathbf{R}^{t+1} = \mathbf{R}^t \cdot \mathbf{R}$, $r_{i,l|t} = 0$ if $l < i$ (because $\mathbf{R}^t$ is upper triangular) and $r_{l,j} = 0$ if $l > j$ (because $\mathbf{R}^t$ is upper triangular), we have

$$r_{i,j|t+1} = \sum_{l=i}^{j} r_{i,l|t} r_{l,j}. \tag{14}$$

From the assumption: $r_{i,j|t} = \sum_{k=i}^{j} p_{i,j,k}(r_{k,k})^{t-1}$, and noting that $p_{i,l,k} = 0$ if $k > l$, we have

$$r_{i,j|t+1} = \sum_{l=i}^{j} r_{l,j} \sum_{k=i}^{l} p_{i,l,k}(r_{k,k})^{t-1}$$
$$= \sum_{k=i}^{j} (r_{k,k})^{t-1} \sum_{l=k}^{j} r_{l,j} p_{i,l,k}. \tag{15}$$

Notice that

$$\sum_{l=k}^{j} r_{l,j} p_{i,l,k} = \sum_{l=k}^{j-1} r_{l,j} p_{i,l,k} + r_{j,j} p_{i,j,k}. \tag{16}$$

Then substituting the sum in (16) by (12) in Definition 2, we have

$$\sum_{l=k}^{j} r_{l,j} p_{i,l,k} = p_{i,j,k}(r_{k,k} - r_{j,j}) + r_{j,j} p_{i,j,k}$$
$$= p_{i,j,k} r_{k,k}. \tag{17}$$

Finally (15) is simplified as $r_{i,j|t+1} = \sum_{k=i}^{j} p_{i,j,k}(r_{k,k})^t$. This is the required conclusion. ∎

Lemma 2 (Theorem 1.3 in [22]): Let $\mathbf{R} = [r_{i,j}]$ be a non-singular upper triangular matrix with unique diagonal entries. For any $t \geq 0$,

$$r_{i,j|t+1} = \sum_{k=i}^{j} p_{i,j,k}(r_{k,k})^{t-1} = \sum_{k=i}^{j} p_{i,j,k}(r_{k,k})^t. \tag{18}$$

Proof: According to (10), (11) and (13) in Definition 2, we see that (18) is true for $t = 1$. Then by induction, (18) is true for all $t > 1$ from Lemma 1. ∎

The above lemma gives an analytic expression of the matrix power $\mathbf{R}^t$. Given a $L \times L$ matrix $\mathbf{R}$, the time complexity of calculating $\mathbf{R}^t$ is $2\binom{L+2}{3} + L(t-3)$ in terms of the number of multiplication and divisions [22].

For the sake of notation, $e(i)$ is denoted by e_i and $q_0(i)$ by q_i. Define coefficients

$$c_k \stackrel{\text{def}}{=} \frac{\sum_{i=1}^{L} \sum_{j=i}^{L} e_i p_{i,j,k} q_j}{f_{\mathrm{opt}}}, \qquad k = 1, \cdots, L, \tag{19}$$

where c_k is independent of t.

Theorem 1: If $\mathbf{R} = [r_{i,j}]$ is a non-singular upper triangular matrix with unique diagonal entries, then for any $t \geq 1$, the relative approximation error E_t is expressed by

$$E_t = \sum_{k=1}^{L} c_k \lambda_k^{t-1}, \tag{20}$$

where $\lambda_k = r_{k,k}$ are eigenvalues of matrix $\mathbf{R}$.

Proof: From (9), we know

$$E_t = \frac{\mathbf{e}^T \mathbf{R}^t \mathbf{q}_0}{f_{\text{opt}}} = \frac{\mathbf{e}^T \mathbf{R}^t \mathbf{q}_0}{f_{\text{opt}}}. \tag{21}$$

Using (18), we get

$$\mathbf{e}^T \mathbf{R}^t \mathbf{q}_0 = \sum_{i=1}^{L} \sum_{j=1}^{L} \sum_{k=i}^{j} e_i p_{i,j,k} (r_{k,k})^{t-1} q_j. \tag{22}$$

According to Definition 2, $p_{i,j,k} = 0$ if $k < i$ or $k > j$ and $p_{i,j,k} = 0$ if $i > j$, then

$$\begin{aligned}
\mathbf{e}^T \mathbf{R}^t \mathbf{q}_0 &= \sum_{i=1}^{L} \sum_{j=1}^{L} \sum_{k=1}^{L} e_i p_{i,j,k} (r_{k,k})^{t-1} q_j \\
&= \sum_{k=1}^{L} (r_{k,k})^{t-1} \sum_{i=1}^{L} \sum_{j=i}^{L} e_i p_{i,j,k} q_j.
\end{aligned} \tag{23}$$

Using c_k, (21) is rewritten as

$$E_t = \sum_{k=1}^{L} c_k (r_{k,k})^{t-1}. \tag{24}$$

The conclusion then is proven. ∎

This theorem shows the relative approximation error is represented as a linear combination of exponential functions $(\lambda_k)^t$ (where $k = 1, \cdots, L$).

From the relationship between F_t and E_t and that between R_t and E_t, we get the following corollaries.

Corollary 1: The fitness value F_t equals to

$$F_t = f_{\text{opt}}(1 - \sum_{k=1}^{L} c_k (\lambda_k)^{t-1}). \tag{25}$$

Corollary 2: The average convergence rate R_t equals to

$$R_t = 1 - \left(\sum_{k=1}^{L} c_k (\lambda_k)^{t-1} \frac{f_{\text{opt}}}{f_{\text{opt}} - f_0} \right)^{1/t}. \tag{26}$$

In practice, the relative approximation error is calculated as follows:

1: given an initial probability distribution $\mathbf{p}_0$, the fitness error $\mathbf{e}$ and matrix $\mathbf{R}$;
2: calculate power factors $[p_{i,j,k}]$ where $i, j, k = 1, \cdots, L$ using Definition 2;
3: calculate coefficients $[c_k]$ (where $k = 1, \cdots, L$) using (19);
4: calculate the relative approximation error E_t using Theorem 1.

VII. EXAMPLE (CONTINUED)

This section applies Theorem 1 to the example in Section IV. The example is chosen for the sake of illustration. Nevertheless Theorem 1 covers all $(1 + 1)$ strictly elitist EAs on any function under the unique condition.

We consider the OneMax function $f_{\text{one}}(x)$ first. The set $\{0, 1\}^4$ is split into 5 subsets

$$\mathcal{S}_i = \{x; |x| = i\}, \qquad i = 0, 1, \cdots, 4. \tag{27}$$

Each subset $\mathcal{S}_i$ is regarded as a state i.

Transition probabilities $r_{i,j} = \Pr(\Phi_t \in \mathcal{S}_i \mid \Phi_{t-1} \in \mathcal{S}_j)$ are given by

$$r_{i,j} = \begin{cases} \frac{j}{4}, & \text{if } j = i + 1, \\ 1 - \frac{j}{4}, & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

Matrix $\mathbf{R}$ is

$$\begin{pmatrix} 0.750 & 0.500 & 0.000 & 0.000 \\ 0.000 & 0.500 & 0.750 & 0.000 \\ 0.000 & 0.000 & 0.250 & 1.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \end{pmatrix} \tag{29}$$

The fitness error $e_i = i$ for $i = 1, \cdots, 4$. The fitness error vector is

$$\mathbf{e}^T = (1, 2, 3, 4).$$

Choose the initial probability distribution in the non-optimal set to be

$$\mathbf{q}_0 = (0, 0, 0, 1)^T.$$

Using Definition 2, we calculate matrix $[p_{i,j,k}]$ which is given by

$$[p_{1,j,k}] = \begin{pmatrix} 0.750 & 1.500 & 2.250 & 3.000 \\ 0.000 & -1.000 & -3.000 & -6.000 \\ 0.000 & 0.000 & 0.750 & 3.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \end{pmatrix},$$

$$[p_{2,j,k}] = \begin{pmatrix} 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.500 & 1.500 & 3.000 \\ 0.000 & 0.000 & -0.750 & -3.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \end{pmatrix},$$

$$[p_{3,j,k}] = \begin{pmatrix} 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.250 & 1.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \end{pmatrix},$$

$$[p_{4,j,k}] = \begin{pmatrix} 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \end{pmatrix}.$$

Using (19), we calculate coefficients c_k (where $k = 1, \cdots, 4$), given by

$$(0.750, 0.000, 0.000, 0.000).$$

Recall transition probabilities $r_{k,k}$ (where $k = 1, \cdots, 4$) are

$$(0.750, 0.500, 0.250, 0.000).$$

Using (20), we calculate the relative approximation error E_t, given by

$$E_t = 0.75^t. \tag{30}$$

Furthermore, using Corollary 1, we calculate the fitness value F_t, given by

$$F_t = 4(1 - 0.75^t). \tag{31}$$

And using Corollary 2, we calculate the average convergence rate R_t, given by

$$R_t = 1 - (0.75^t)^{1/t} = 0.25. \tag{32}$$

This means that E_t decays as fast as an exponential function: $E_t = 0.75^t E_0$.

The analysis of the quadratic function $f_{\text{squ}}(x)$ and logarithmic function $f_{\text{log}}(x)$ is almost the same as that of the OneMax function, except the fitness error vector $\mathbf{e}$. The results are summarised in Table I. Notice that the expressions for quadratic and logarithmic functions are more complex than that for the OneMax function.

Fig. 4 demonstrates the fitness value F_t. Fig. 5 presents the relative approximation error E_t. Fig. 6 illustrates the average convergence rates R_t. The theoretical predictions are consistent to the experimental results, labelled by f^*.
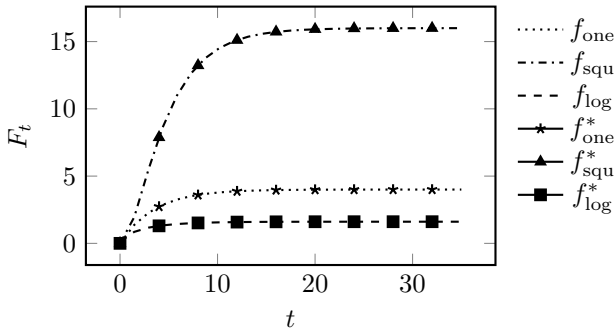

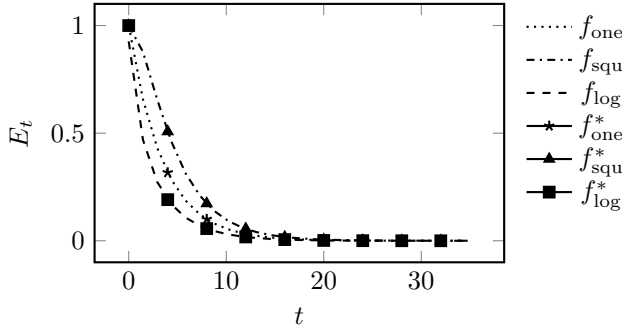Fig. 4. Fitness value F_t.

Fig. 5. Relative approximation error E_t.

VIII. EXTENSION

This section devotes to an extension from (1+1) strictly elitist EAs to non-elitist or population-based EAs.

Many non-elitist or population-based EAs can be modelled by homogeneous Markov chains but matrices $\mathbf{R}$ are not upper triangular. Given any matrix $\mathbf{R}$, according to Schur's triangularisation theorem (in textbook [24, p508]), there exists an upper triangular matrix $\tilde{\mathbf{R}}$ and unitary matrix $\mathbf{U}$ such that $\mathbf{R} = \mathbf{U}\tilde{\mathbf{R}}\mathbf{U}^*$. Then the matrix iteration (8) can be rewritten as follows,

$$\mathbf{q}_t = \mathbf{R}^t \mathbf{q}_0 = \mathbf{U}\tilde{\mathbf{R}}^t\mathbf{U}^*\mathbf{q}_0. \tag{33}$$

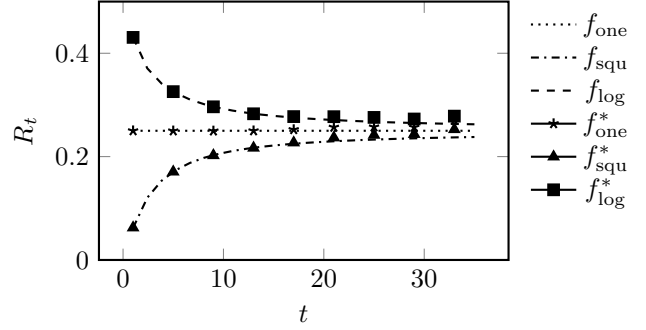Fig. 6. Average convergence rate R_t.

Then the relative approximation error equals to

$$E_t = \frac{\mathbf{e}^T \mathbf{U}\tilde{\mathbf{R}}^t\mathbf{U}^*\mathbf{q}_0}{f_{\text{opt}}}. \tag{34}$$

Let $\tilde{\mathbf{e}}^T = \mathbf{e}^T\mathbf{U}$ and $\tilde{\mathbf{q}}_0 = \mathbf{U}^*\mathbf{q}_0$, then the relative approximation error can be rewritten as follows,

$$E_t = \frac{\tilde{\mathbf{e}}^T \tilde{\mathbf{R}}^t \tilde{\mathbf{q}}_0}{f_{\text{opt}}}. \tag{35}$$

Since $\tilde{\mathbf{R}}$ is an upper triangular matrix, the analysis of E_t becomes the problem of expressing the matrix power $\tilde{\mathbf{R}}^t$. If $\tilde{\mathbf{R}}$ is an upper triangular matrix with unique diagonal entries, Theorem 1 can be applied directly. If this does not hold, a similar analysis can be conduced (but in a separate paper). Therefore, in theory it is feasible to apply the approach to non-elitist or population-based EAs too.

Furthermore, even if exact transition probabilities are unknown, it is still possible to apply the method to bounding the relative approximation error. The idea is simple. We construct an upper triangular matrix $\mathbf{S} = [s_{i,j}]$ so that the matrix iteration using $\mathbf{S}$ is slower than that using $\mathbf{S}$. That is $\mathbf{e}^T\mathbf{S}^t\mathbf{q}_0 \geq \mathbf{e}^T\mathbf{R}^t\mathbf{q}_0$. For example, the simplest matrix $\mathbf{S}$ is

$$s_{i,j} = \begin{cases} 1 - r_{j,j}, & \text{if } j = i+1, \\ r_{i,i}, & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

This issue will be discussed in a separate paper.

IX. CONCLUSIONS

In this paper, the solution quality of an EA is measured by the relative approximation error, that is

$$E_t = 1 - \frac{F_t}{f_{\text{opt}}}. \tag{37}$$

Then an analytic expression of the relative approximation error E_t is presented for any (1+1) strictly elitist EAs on any fitness function. Provided that transition probabilities $r_{i,i} \neq r_{j,j}$ for any $i \neq j$, the formula is given by

$$E_t = \sum_{k=1}^{L} c_k \lambda_k^{t-1}, \tag{38}$$

where $\lambda_k = r_{k,k}$ (where $k = 1, \cdots, L$) are eigenvalues of transition submatrix $\mathbf{R}$ and c_k are coefficients.

TABLE I
ANALYTIC EXPRESSIONS OF F_t, E_t AND R_t IN THE EXAMPLE.

function	F_t
$f_{\mathrm{one}} = \|x\|$	$4 \times (1 - 0.75 \times 0.75^{t-1})$
$f_{\mathrm{squ}} = \|x\|^2$	$16 \times (1 - 1.313 \times 0.75^{t-1} + 0.375 \times 0.5^{t-1})$
$f_{\mathrm{log}} = \ln(\|x\| + 1)$	$\ln(5) \times (1 - 0.416 \times 0.75^{t-1} - 0.120 \times 0.5^{t-1} - 0.033 \times 0.25^{t-1})$

	E_t
$f_{\mathrm{one}} = \|x\|$	$0.75 \times 0.75^{t-1}$
$f_{\mathrm{squ}} = \|x\|^2$	$1.313 \times 0.75^{t-1} - 0.375 \times 0.5^{t-1}$
$f_{\mathrm{log}} = \ln(\|x\| + 1)$	$0.416 \times 0.75^{t-1} + 0.120 \times 0.5^{t-1} + 0.033 \times 0.25^{t-1}$

	R_t
$f_{\mathrm{one}} = \|x\|$	0.25
$f_{\mathrm{squ}} = \|x\|^2$	$1 - (1.313 \times 0.75^{t-1} - 0.375 \times 0.5^{t-1})^{1/t}$
$f_{\mathrm{log}} = \ln(\|x\| + 1)$	$1 - (0.416 \times 0.75^{t-1} + 0.120 \times 0.5^{t-1} + 0.033 \times 0.25^{t-1})^{1/t}$

The above formula is also useful to fixed budget analysis. Since the exact expression of the fitness value F_t is

$$F_t = f_{\mathrm{opt}} \left(1 - \sum_{k=1}^{L} c_k(\lambda_k)^{t-1} \right), \qquad (39)$$

a good bound on F_t should be represented in the form of a combination of exponential functions of t.

The work is a further development of the average convergence rate [18]. The exact expression of the average convergence rate R_t is

$$R_t = 1 - \left(\sum_{k=1}^{L} c_k(\lambda_k)^{t-1} \frac{f_{\mathrm{opt}}}{f_{\mathrm{opt}} - f_0} \right)^{1/t}. \qquad (40)$$

The approach is promising. Using Schur's triangularization theorem, it is feasible to make a similar analysis for non-elitist or population-based EAs if they are modelled by homogeneous Markov chains.

Our next work is to present a closed form for (1+1) strictly elitist EAs whose transition matrices are upper triangular but diagonal entries are not unique.

Acknowledgement:

REFERENCES

[1] P. S. Oliveto, J. He, and X. Yao, "Analysis of the (1+1)-EA for finding approximate solutions to vertex cover problems," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 5, pp. 1006 –1029, 2009.

[2] T. Friedrich, J. He, N. Hebbinghaus, F. Neumann, and C. Witt, "Approximating covering problems by randomized search heuristics using multi-objective models," *Evolutionary Computation*, vol. 18, no. 4, pp. 617–633, 2010.

[3] C. Witt, "Worst-case and average-case approximations by simple randomized search heuristics," in *Proceedings of the 22nd Annual Conference on Theoretical Aspects of Computer Science*. Springer-Verlag, 2005, pp. 44–56.

[4] Y. Yu, X. Yao, and Z.-H. Zhou, "On the approximation ability of evolutionary optimization with application to minimum set cover," *Artificial Intelligence*, no. 180-181, pp. 20–33, 2012.

[5] X. Lai, Y. Zhou, J. He, and J. Zhang, "Performance analysis of evolutionary algorithms for the minimum label spanning tree problem," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 6, pp. 860–872, 2014.

[6] J. He and X. Yao, "An analysis of evolutionary algorithms for finding approximation solutions to hard optimisation problems," in *Proceedings of IEEE 2003 Congress on Evolutionary Computation*. IEEE Press, 2003, pp. 2004–2010.

[7] J. Suzuki, "A Markov chain analysis on simple genetic algorithms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 4, pp. 655–659, 1995.

[8] ——, "A further result on the markov chain model of genetic algorithms and its application to a simulated annealing-like strategy," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 28, no. 1, pp. 95–102, 1998.

[9] J. He and L. Kang, "On the convergence rate of genetic algorithms," *Theoretical Computer Science*, vol. 229, no. 1-2, pp. 23–39, 1999.

[10] G. Rudolph, "Local convergence rates of simple evolutionary algorithms with Cauchy mutations," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 4, pp. 249–258, 1997.

[11] ——, "Convergence rates of evolutionary algorithms for a class of convex objective functions," *Control and Cybernetics*, vol. 26, pp. 375–390, 1997.

[12] ——, "Convergence rates of evolutionary algorithms for quadratic convex functions with rank-deficient hessian," in *Adaptive and Natural Computing Algorithms*. Springer, 2013, pp. 151–160.

[13] L. Ming, Y. Wang, and Y.-M. Cheung, "On convergence rate of a class of genetic algorithms," in *Proceedings of 2006 World Automation Congress*. IEEE, 2006, pp. 1–6.

[14] F. Schmitt and F. Rothlauf, "On the importance of the second largest eigenvalue on the convergence rate of genetic algorithms," in *Proceedings of 2001 Genetic and Evolutionary Computation Conference*, H. Beyer, E. Cantu-Paz, D. Goldberg, Parmee, L. Spector, and D. Whitley, Eds. Morgan Kaufmann Publishers, 2001, pp. 559–564.

[15] L. Ding and L. Kang, "Convergence rates for a class of evolutionary algorithms with elitist strategy," *Acta Mathematica Scientia*, vol. 21, no. 4, pp. 531–540, 2001.

[16] T. Jansen and C. Zarges, "Fixed budget computations: A different perspective on run time analysis," in *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*. ACM, 2012, pp. 1325–1332.

[17] ——, "Performance analysis of randomised search heuristics operating with a fixed budget," *Theoretical Computer Science*, vol. 545, pp. 39–58, 2014.

[18] J. He and G. Lin, "Average convergence rate of evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 2, pp. 316–321, 2016.

[19] J. He, T. Chen, and X. Yao, "On the easiest and hardest fitness functions," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 2, pp. 295–305, 2015.

[20] J. He and X. Yao, "Towards an analytic framework for analysing the computation time of evolutionary algorithms," *Artificial Intelligence*, vol. 145, no. 1-2, pp. 59–97, 2003.

[21] G. Rudolph, "Finite Markov chain results in evolutionary computation: a tour d'horizon," *Fundamenta Informaticae*, vol. 35, no. 1, pp. 67–89, 1998.

[22] W. Shur, "A simple closed form for triangular matrix powers," *Electronic Journal of Linear Algebra*, vol. 22, pp. 1000–1003, 2011.

[23] C. Huang, "An efficient algorithm for computing powers of triangular matrices," in *Proceedings of the 1978 ACM Annual Conference-Volume 2*. ACM, 1978, pp. 954–957.

[24] C. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.