Research Article

An Analytical Framework for Runtime of a Class of Continuous Evolutionary Algorithms

Yushan Zhang and Guiwu Hu

School of Mathematics and Statistics, Guangdong University of Finance and Economics, Guangzhou 510320, China

Correspondence should be addressed to Yushan Zhang; scuthill@163.com

Received 26 May 2015; Accepted 26 July 2015

Academic Editor: Manuel Graña

Copyright © 2015 Y. Zhang and G. Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Although there have been many studies on the runtime of evolutionary algorithms in discrete optimization, relatively few theoretical results have been proposed on continuous optimization, such as evolutionary programming (EP). This paper proposes an analysis of the runtime of two EP algorithms based on Gaussian and Cauchy mutations, using an absorbing Markov chain. Given a constant variation, we calculate the runtime upper bound of special Gaussian mutation EP and Cauchy mutation EP. Our analysis reveals that the upper bounds are impacted by individual number, problem dimension number n , searching range, and the Lebesgue measure of the optimal neighborhood. Furthermore, we provide conditions whereby the average runtime of the considered EP can be no more than a polynomial of n . The condition is that the Lebesgue measure of the optimal neighborhood is larger than a combinatorial calculation of an exponential and the given polynomial of n .

1. Introduction

The running time to optimum is a key factor in determining the success of an evolutionary programming (EP) approach. Ideally, an implementation of an EP approach should run for a sufficient number of generations when the probability of achieving an optimum is greater than some desired value. However, few results on the running time of EP approaches can be found in the current literature.

As a technique of finite state machine, EP was first proposed for continuous optimization [1]. It has since been widely adopted as a powerful optimizing framework for solving continuous optimization problems [2–4]. More recently, EP research has mainly concentrated on the function of its parameters and its improvement by adaptive strategies. As a result, several EP variants have been proposed, distinguishing from each other mainly with different mutation schemes based on different probability distributions.

Arguably, the first EP algorithm that was widely considered successful was the one with Gauss mutation that has been termed as *classical evolutionary programming* (CEP) [2]. CEP has been intensively analyzed by Fogel [3, 4], Bäck and

Schwefel [5], and Schwefel [5, 6]. Subsequently, Yao et al. proposed a *fast evolutionary programming algorithm* (FEP) with a new mutation scheme based on the Cauchy distribution [7]. Computational experiments showed that FEP is superior to CEP when tackling the optimization problems of multimodal and dispersed peak functions. Another EP variant [8] was later proposed by using the Lévy distribution based mutation, which we call *Lévy evolutionary programming* (LEP) in this paper. Empirical analyses show that, overall, LEP exceeds CEP and FEP when solving the benchmark problems of multimodal and highly dispersed peak functions.

CEP [2], FEP [7], and LEP [8] can be considered as classical evolutionary programming algorithms. Several modified types of EP have since been designed based on these three basic approaches [2, 7, 8].

The performances of EP approaches such as CEP, FEP, and LEP have oftentimes been verified experimentally rather than theoretically. The theoretical foundations of EP have been an open problem ever since it was first put forward [1, 2]. In particular, Bäck and Schwefel [5] and Schwefel [6] suggested the convergence analysis of EP algorithms as a

research topic in their surveys of evolutionary algorithms. Fogel [2–4] presented an initial proof of EP convergence on the basis of the discrete solution space. Rudolph [9–11] then showed that CEP and FEP can converge with an arbitrary initialization; this result is more general since it applies to a continuous solution space.

Previous convergence studies only considered whether an EP algorithm is able to find an optimum within infinite iteration but did not mention the speed of convergence, that is, lacking running time analysis. To date, running time analyses have mainly focused on Boolean individual EAs like $(1 + 1)$ EA [12], $(N + N)$ EA [13], multiobjective EA [14–16], and so forth. Alternative theoretical measures for evaluating the running time of Boolean-individual EA approaches have also been proposed [17, 18]. Recently, the impact of particular components of EA on runtime has also been studied, on mutation, selection [19], and population size [18, 20, 21]. In addition to these studies on EAs solving Boolean functions, some results of runtime analysis have been obtained on some combinatorial optimization problems [22–24]. Lehre and Yao [25] completed the runtime analysis of the $(1 + 1)$ EA on computing unique input output sequences. Zhou et al. presented a series of EA analysis results for discrete optimization like the minimum label spanning tree problem [26], the multiprocessor scheduling problem [27], and the maximum cut problem [28]. Other proposed studies are on the topics of tight bounds on the running time of EA and randomized search heuristic [29–31].

As summarized above, the majority of runtime analyses are limited to discrete search space; analyses for continuous search space require a more sophisticated modeling and remain relatively few, which is unsatisfactory from a theoretical point of view. Jägersküpper conducted a rigorous runtime analysis on $(1 + 1)$ ES, $(1 + \lambda)$ ES minimizing the Sphere function [32, 33]. Agapie et al. modeled the continuous EA as a renewal process under some strong assumption and analyzed the first hitting time of $(1 + 1)$ EA on inclined plane and Sphere function [34, 35]. Chen et al. [36] proposed general drift conditions to estimate the upper bound of the first hitting time for EAs to find ϵ -approximation solutions. Inspired by the studies above, especially the estimating approach in [17] for discrete situation, this paper presents an analytical framework for the running time of evolutionary programming. We also discuss whether the running time of CEP and FEP is influenced by individual number, problem dimension number, searching range, and the Lebesgue measure of the optimal neighborhood in the target problem. Furthermore, we give the approximate conditions under which the EP can converge in a timespan less than a polynomial of n .

2. EP Algorithms and the Associated Markov Chain

2.1. Introduction to EP Algorithms. This section introduces the two EP algorithms CEP [2] and FEP [7] that are studied in this paper. The skeleton of EP algorithms analyzed in this paper is given in Algorithm 1; the sole difference among CEP, FEP, and LEP lies in their treatments of Step (3).

Algorithm 1 (framework of EP algorithms).

- (1) Generate individuals randomly.
- (2) Calculate the fitness of each individual.
- (3) Generate offspring by mutation.
- (4) Evaluate the offspring.
- (5) Select individuals by a tournament rule.
- (6) If the terminal condition is satisfied, output the best-so-far solution found and exit; otherwise, go back to Step (3).

In Step (1), the generated k individuals are denoted by a couple of real vectors $\mathbf{v}_i = (\mathbf{x}_i, \boldsymbol{\sigma}_i)$, $i = 1, 2, \dots, k$, where the elements $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$ are the variables to be optimized, and the elements $\boldsymbol{\sigma}_i = (\sigma_{i1}, \sigma_{i2}, \dots, \sigma_{in})$ are *variation variables* that affect the generation of offspring. We set the iteration number $t = 0$ and the initial $\sigma_{ij} \leq 2$ ($j = 1, \dots, n$) as proposed in [8]. The fitness values in Steps (2) and (4) are calculated according to the objective function of the target problem.

In Step (3), a single $\bar{\mathbf{v}}_i = (\bar{\mathbf{x}}_i, \bar{\boldsymbol{\sigma}}_i)$ is generated for each individual $\mathbf{v}_i$, where $i = 1, \dots, k$. For $j = 1, \dots, n$,

$$\bar{\sigma}_{ij} = V(\sigma_{ij}), \quad (1)$$

where $V(\sigma_{ij})$ denotes a renewing function of variation variable σ_{ij} . The renewing function $\bar{\sigma}_{ij} = V(\sigma_{ij})$ may have various forms, and a representative implementation of it can be found in [7]. However, for ease of theoretical analysis, we will consider a kind of simple renewing function, that is, constant function, in our case studies (see Section 4):

$$V(\sigma_{ij}) = \sigma > 0. \quad (2)$$

The three EP algorithms differ in how $\bar{\mathbf{x}}_i$ is derived, which can be explained as follows (in this paper, we only consider CEP and FEP):

$$\text{CEP: } \bar{x}_{ij} = x_{ij} + \bar{\sigma}_{ij} N_j(0, 1), \quad (3)$$

$$\text{FEP: } \bar{x}_{ij} = x_{ij} + \bar{\sigma}_{ij} \delta_j, \quad (4)$$

where $N_j(0, 1)$ is a newly generated random variable by Gaussian distribution for each j and δ_j is a Cauchy random number produced anew for j th dimension, whose density function is

$$C_{\phi=1}(y) = \pi^{-1} \cdot (1 + y^2)^{-1} \quad -\infty < y < +\infty. \quad (5)$$

In Step (5), for each individual in the set of parents and offsprings, q different *opponent* individuals are uniformly and randomly selected to be compared. If the selected individual's fitness value is more than the opponent's, the individual obtains a “win.” The top k individuals with the most “wins” are selected to be the parents in the next iteration, breaking ties by greater fitness.

2.2. Target Problem of EP Algorithms. Without loss of generality, we assume that the EP approaches analyzed in this study aim to solve a minimization problem in a continuous search space, defined as follows.

Definition 2 (minimization problem). Let $\mathbf{S} \subseteq \mathbf{R}^n$ be a finite subspace of the n -dimensional real domain $\mathbf{R}^n$, and let $f : \mathbf{S} \rightarrow \mathbf{R}$ be an n -dimensional real function. A minimization problem, denoted by the 2-tuple $(\mathbf{S}, f)$, is to find an n -dimensional vector $\mathbf{x}_{\min} \in \mathbf{S}$ such that, $\forall \mathbf{x} \in \mathbf{S}$, $f(\mathbf{x}_{\min}) \leq f(\mathbf{x})$.

Without loss of generality, we can assume that $\mathbf{S} = \prod_{i=1}^n [-b_i, b_i]$ where $b_i = b > 0$. The function $f : \mathbf{S} \rightarrow \mathbf{R}$ is called the *objective function* of the minimization problem. We do not require f to be continuous, but it must be bounded. Furthermore, we only consider the unconstrained minimization.

The following properties are assumed, and we will make use of them in our analyses:

- (1) The subset of optimal solutions in $\mathbf{S}$ is nonempty.
- (2) Let $f^* = \min\{f(x) \mid x \in \mathbf{S}\}$ be the fitness value of an optimal solution, and let $\mathbf{S}^*(\varepsilon) = \{x \in \mathbf{S} \mid f(x) < f^* + \varepsilon\}$ be the optimal epsilon neighborhood. Every element of $\mathbf{S}^*(\varepsilon)$ is an optimal solution of the minimization problem.
- (3) $\forall \varepsilon > 0$, $m(\mathbf{S}^*(\varepsilon)) > 0$, where we denote the Lebesgue measure of $\mathbf{S}^*(\varepsilon)$ as $m(\mathbf{S}^*(\varepsilon))$.

The first assumption describes the existence of optimal solutions to the problem. The second assumption presents a rigorous definition of optimal solution for continuous minimization optimization problems. The third assumption indicates that there are always solutions whose objective values are distributed continuously and arbitrarily close to the optimal, which makes the minimization problem solvable.

2.3. Markov Chain Associated with EP Algorithms. Our running time analyses are based on representing the EP algorithms as Markov chain. In this section, we explain the terminologies and notations used throughout the rest of this study.

Definition 3 (stochastic process of EP). The stochastic process of an evolutionary programming algorithm EP is denoted by $\{\xi_t^{EP}\}_{t=0}^{+\infty}$; $\xi_t^{EP} = (\mathbf{v}_1^{(t)}, \mathbf{v}_2^{(t)}, \dots, \mathbf{v}_k^{(t)})$, where $\mathbf{v}_i^{(t)} = (\mathbf{x}_i^{(t)}, \sigma_i^{(t)})$ is the i th individual at the t th iteration.

The stochastic status ξ_t^{EP} represents the individuals of the t th iteration population for the algorithm EP. All stochastic processes examined in this paper are discrete-time; that is, $t \in \mathbf{Z}^+$.

Definition 4 (status space). The status space of EP is $\Omega_{EP} = \{(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) \mid \mathbf{v}_i = (\mathbf{x}_i, \sigma_i), \mathbf{x}_i \in \mathbf{S}, \sigma_{ij} \leq 2, j = 1, \dots, n\}$.

Ω_{EP} is the set of all possible population statuses for EP. Intuitively, each element of Ω_{EP} is associated with a possible

population in the implementation of EP. Let $\mathbf{x}^* \in \mathbf{S}^*(\varepsilon)$ be an optimal solution. We define the optimal status space as follows.

Definition 5 (optimal status space). The optimal status space of EP is the subset $\Omega_{EP}^*(\varepsilon) \subseteq \Omega_{EP}$ such that, $\forall \mathbf{u}^* \in \Omega_{EP}^*(\varepsilon)$, $\exists (\mathbf{x}^*, \sigma) \in \mathbf{u}^*$, where $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ and $\sigma > 0$.

Hence, all members of $\Omega_{EP}^*(\varepsilon)$ contain at least one optimal solution represented by individual $\mathbf{x}^*$.

Definition 6 (Markov chain). A stochastic process $\{\xi_t\}_{t=0}^{+\infty}$ with status space Ω is a Markov chain if $\forall \bar{\Omega} \subseteq \Omega$, $P\{\xi_{t+1} \in \bar{\Omega} \mid \xi_0, \dots, \xi_t\} = P\{\xi_{t+1} \in \bar{\Omega} \mid \xi_t\}$.

Lemma 7. The stochastic process $\{\xi_t^{EP}\}_{t=0}^{+\infty}$ of EP is a Markov chain.

Proof. The proof is given in the appendix. $\square$

We now show that the stochastic process $\{\xi_t^{EP}\}_{t=0}^{+\infty}$ of EP is an absorbing Markov chain, defined as follows.

Definition 8 (absorbing Markov chain to optimal status space). A Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ is an absorbing Markov chain to Ω^* if $\exists \Omega^* \subset \Omega$, such that $P\{\xi_{t+1} \notin \Omega^* \mid \xi_t \in \Omega^*\} = 0$ for $t = 0, 1, \dots$.

Lemma 9. The stochastic process $\{\xi_t^{EP}\}_{t=0}^{+\infty}$ of EP is an absorbing Markov chain to $\Omega_{EP}^*(\varepsilon)$.

Proof. The proof is given in the appendix. $\square$

This property implies that once an EP algorithm attains an optimal state, it will never leave optimality.

3. General Runtime Analysis Framework for EP Algorithms

The analysis of EP algorithms has usually been approximated using a simpler measure known as the first hitting time [13, 17], which is employed in this study.

Definition 10 (first hitting time of EP). μ_{EP} is the first hitting time of EP if $\mu_{EP} = \min\{t \geq 0 : \xi_t^{EP} \in \Omega_{EP}^*(\varepsilon)\}$.

If an EP algorithm is modeled as an absorbing Markov chain, the running time of the EP can be measured by its first hitting time μ_{EP} . We denote its expected value by $E\mu_{EP}$ which can be calculated by Theorem 11.

Let $\lambda_t^{EP} = P\{\xi_t^{EP} \in \Omega_{EP}^*(\varepsilon)\} = P\{\mu_{EP} \leq t\}$ be the probability that EP has attained an optimal state by time t .

Theorem 11. If $\lim_{t \rightarrow +\infty} \lambda_t^{EP} = 1$, the expected first hitting time $E\mu_{EP} = \sum_{i=0}^{+\infty} (1 - \lambda_i^{EP})$.

Corollary 12. The expected first hitting time can also be expressed as $E\mu_{EP} = (1 - \lambda_0^{EP}) \sum_{t=0}^{+\infty} \prod_{i=1}^t (1 - P\{\xi_i^{EP} \in \Omega_{EP}^*(\varepsilon) \mid \xi_{i-1}^{EP} \notin \Omega_{EP}^*(\varepsilon)\})$.

Proof. The proof is given in the appendix. $\square$

The conclusions of Theorem 11 and Corollary 12 are a direct approach to calculating the first hitting time of EP. However, Corollary 12 is more practical than Theorem 11 for this purpose since $p_i = P\{\xi_i^{EP} \in \Omega_{EP}^*(\varepsilon) \mid \xi_{i-1}^{EP} \notin \Omega_{EP}^*(\varepsilon)\}$, which is the probability that the process first finds an optimal solution at time i , is dependent on the mutation and selection techniques of the EP. Hence, the framework of Corollary 12 is similar to the one in [17]. However, the estimating method [17] is based on a probability q_i in discrete optimization $q_i = \sum_{y \notin \Omega_{EP}^*(\varepsilon)} P\{\xi_i^{EP} \in \Omega_{EP}^*(\varepsilon) \mid \xi_{i-1}^{EP} = y\} P\{\xi_{i-1}^{EP} = y\}$. p_i is a probability for continuous status while q_i is for the discrete one. In general, the exact value of p_i is difficult to calculate. Alternatively, first hitting time can also be analyzed in terms of an upper and a lower bound of p_i , as shown by the following Corollaries 13 and 15.

Corollary 13. If $\alpha_t \leq P\{\xi_t^{EP} \in \Omega_{EP}^*(\varepsilon) \mid \xi_{t-1}^{EP} \notin \Omega_{EP}^*(\varepsilon)\} \leq \beta_t$,

$$(1) \sum_{t=0}^{+\infty} [(1 - \lambda_0^{EP}) \prod_{i=1}^t (1 - \beta_i)] \leq E\mu_{EP} \leq \sum_{t=0}^{+\infty} [(1 - \lambda_0^{EP}) \prod_{i=1}^t (1 - \alpha_i)] \text{ where } 0 < \alpha_t \text{ and } \beta_t < 1;$$

$$(2) \beta^{-1}(1 - \lambda_0^{EP}) \leq E\mu_{EP} \leq \alpha^{-1}(1 - \lambda_0^{EP}) \text{ when } \alpha_t = \alpha \text{ and } \beta_t = \beta.$$

Proof. The proof is given in the appendix. $\square$

Corollary 13 indicates that $E\mu_{EP}$ can be studied by the lower bound and upper bound of $P\{\xi_t^{EP} \in \Omega_{EP}^*(\varepsilon) \mid \xi_{t-1}^{EP} \notin \Omega_{EP}^*(\varepsilon)\}$. Therefore, the theorem and corollaries introduce a first-hitting-time framework for the running time analysis of EP. The running times of EP based on Gaussian and Cauchy mutation are studied following the framework.

4. Runtime Upper Bounds of CEP and FEP

In this section, we use the framework proposed in Section 3 to study the running time of EP based on Gaussian and Cauchy mutations, that is, CEP and FEP. Moreover, our running time analysis aims at a class of EP with constant variation, as shown in (2).

4.1. Mean Running Time of CEP. CEP [2] is a classical EP, from which several continuous evolutionary algorithms are derived. This subsection mainly focuses on the running time of CEP proposed by Fogel [3], Schwefel [6], and Schwefel [5, 6]. The mutation of CEP is based on the standard normal distribution indicated by (3).

According to the running time analysis framework, the probability $p_t = P\{\xi_t^C \in \Omega_{EP}^*(\varepsilon) \mid \xi_{t-1}^C \notin \Omega_{EP}^*(\varepsilon)\}$ for CEP is a crucial factor, and its property is introduced by the following.

Theorem 14. Let the stochastic process of CEP be denoted by $\{\xi_t^C\}_{t=0}^{+\infty}$. Let k be the population size of CEP with solution space $S = \prod_{i=1}^n [-b_i, b_i]$ where $b_i = b > 0$. Then $\forall \varepsilon > 0$, one has the following:

- (1) For fixed individual $(\mathbf{x}, \bar{\sigma})$, $P\{\bar{\mathbf{x}} \in S^*(\varepsilon)\} \geq (m(S^*(\varepsilon)) / (\sqrt{2\pi})^n) (\prod_{j=1}^n (1/\sigma_j)) \exp\{-\sum_{j=1}^n (2b^2/\sigma_j^2)\}$ where the renewing function $V(\sigma_{ij}) = \sigma$.
- (2) The right part of the inequality of conclusion (1) is maximum when $\sigma = 2b$.
- (3) $P\{\xi_t^C \in \Omega_{EP}^*(\varepsilon) \mid \xi_{t-1}^C \notin \Omega_{EP}^*(\varepsilon)\} \geq 1 - (1 - m(S^*(\varepsilon)) / (4\sqrt{e\pi}b)^n)^k$ if $\sigma = 2b$.

Proof. The proof is given in the appendix. $\square$

In Theorem 14, Part (1) indicates a lower bound of the probability that an individual $\mathbf{x}$ can be renewed to be $\bar{\mathbf{x}}$ in the optimal neighborhood $S^*(\varepsilon)$. As a result, the lower bound is a function of variation σ . Part (2) presents the lower bound that Part (1) will be tightest if $\sigma = 2b$. By this, the key factor of Corollary 13 $P\{\xi_t^C \in \Omega_{EP}^*(\varepsilon) \mid \xi_{t-1}^C \notin \Omega_{EP}^*(\varepsilon)\}$ may have a tight lower bound shown by Part (3).

Theorem 14 indicates that a larger $m(S^*(\varepsilon))$ leads to faster convergence for CEP since $P\{\xi_t^C \in \Omega_{EP}^*(\varepsilon) \mid \xi_{t-1}^C \notin \Omega_{EP}^*(\varepsilon)\}$ becomes larger when $m(S^*(\varepsilon))$ is larger. Moreover, Theorem 14 produces the lower bound of $P\{\xi_t^C \in \Omega_{EP}^*(\varepsilon) \mid \xi_{t-1}^C \notin \Omega_{EP}^*(\varepsilon)\}$, with which the first-hitting time of CEP can be estimated following Corollary 13. Corollary 15 indicates the convergence capacity and running time upper bound of CEP.

Corollary 15. Supposed conditions of Theorem 14 are satisfied:

- (1) $\lim_{t \rightarrow +\infty} \lambda_t^C = 1$ ($\lambda_t^C = P\{\xi_t^C \in \Omega_{EP}^*(\varepsilon)\}$);
- (2) $\forall \varepsilon > 0$,

$$E\mu_C \leq (1 - \lambda_0^C) \left(1 - \left(1 - \frac{m(S^*(\varepsilon))}{(4\sqrt{e\pi}b)^n} \right)^k \right)^{-1}. \quad (6)$$

Proof. The proof is given in the appendix. $\square$

Corollary 15 shows that CEP converges globally and that the upper bound of CEP's running time decreases as the Lebesgue measure of the optimal ε neighborhood $m(S^*(\varepsilon))$ of $S^*(\varepsilon)$ increases. A larger $m(S^*(\varepsilon))$ translates to a larger optimal ε neighborhood $\forall \varepsilon > 0$, which allows the solution of CEP more easily. Moreover, an increase in problem dimension number can increase the upper bound, and enlarging the individual number will make the upper bound smaller.

According to (6), $E\mu_C$ has a smaller upper bound when the population size k increases. Approximately,

$$E\mu_C \leq (1 - \lambda_0^C) \frac{(4\sqrt{e\pi}b)^n}{km(S^*(\varepsilon))} \quad (7)$$

since $(1 - m(S^*(\varepsilon))/(4\sqrt{e\pi}b)^n)^k \approx 1 - k(m(S^*(\varepsilon))/(4\sqrt{e\pi}b)^n)$ when k is large. Furthermore, the running time of CEP is similar to an exponential order function of size n , if $m(S^*(\varepsilon)) \geq C_0 > 0$ where C_0 is a positive constant. Hence, the running time of CEP is nearly $O((4\sqrt{e\pi}b)^n)$.

Moreover, we can give an approximate condition under which CEP can converge in time polynomial to n when $m(S^*(\varepsilon)) \geq c_C((4\sqrt{e\pi}b)^n/P_n)$, where $c_C = (1 - \lambda_0^C)/P_n > 0$.

Suppose (7) holds. Then $m(\mathbf{S}^*(\varepsilon)) \geq (1 - \lambda_0^C)(4\sqrt{e}\pi b)^n / P_n^2 > 0 \Leftrightarrow E\mu_C \leq (1 - \lambda_0^C)(4\sqrt{e}\pi b)^n / P_n \cdot m(\mathbf{S}^*(\varepsilon)) \leq P_n$. Thus, the running time of CEP can be polynomial to n , under the constraint that, $\forall \varepsilon > 0$, $m(\mathbf{S}^*(\varepsilon)) \geq ((1 - \lambda_0^C)/k) \cdot ((4\sqrt{e}\pi b)^n / P_n)$, where $\mathbf{S}^*(\varepsilon) = \{\mathbf{x} \in \mathbf{S} \mid f(\mathbf{x}) < f^* + \varepsilon\}$.

4.2. Mean Running Time of FEP. FEP was proposed by Yao et al. [7] as an improvement to CEP. The mutation for FEP is based on the Cauchy distribution indicated by (4). The property of the probability $p_t = P\{\xi_t^F \in \Omega_{EP}^*(\varepsilon) \mid \xi_{t-1}^F \notin \Omega_{EP}^*(\varepsilon)\}$ is discussed by Theorem 16 ($t = 0, 1, \dots$).

Theorem 16. Let $\{\xi_t^F\}_{t=0}^{+\infty}$ be the stochastic process of FEP, and then one has the following:

- (1) $P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\} \geq (m(\mathbf{S}^*(\varepsilon)) / (\pi^n)(\sigma + 4b^2/\sigma)^{-n}$.
- (2) The right part of the inequality of conclusion (1) is maximum when $\sigma = 2b$.
- (3) $P\{\xi_t^F \in \Omega_{EP}^*(\varepsilon) \mid \xi_{t-1}^F \notin \Omega_{EP}^*(\varepsilon)\} \geq 1 - (1 - m(\mathbf{S}^*(\varepsilon)) / (4\pi b)^n)^k$ if $\sigma = 2b$.

Proof. The proof is given in the appendix. $\square$

Similar to Theorem 14, FEP with $\sigma = 2b$ may lead to a tight lower bound of the probability in Part (1) and Part (3). Theorem 16 indicates that $m(\mathbf{S}^*(\varepsilon))$ affects the convergence of FEP directly. A larger $m(\mathbf{S}^*(\varepsilon))$ allows FEP to more easily arrive at a status in the optimal state space (Definition 5). Conversely, a bigger b (the bounds of search space) increases the difficulty of the problem for FEP. Given this convergence time framework, Corollary 17 summarizes the convergence capacity and running time upper bound of FEP.

Corollary 17. Supposed conditions of Theorem 16 are satisfied:

- (1) $\lim_{t \rightarrow +\infty} \lambda_t^F = 1$ ($\lambda_t^F = P\{\xi_t^F \in \Omega_{EP}^*(\varepsilon)\}$);
- (2) $\forall \varepsilon > 0$, $E\mu_F \leq (1 - \lambda_0^F)(1 - (1 - m(\mathbf{S}^*(\varepsilon)) / (4b\pi)^n)^k)$.

Proof. The proof is given in the appendix. $\square$

Corollary 17 proves the convergence of FEP and indicates that a larger $m(\mathbf{S}^*(\varepsilon))$ can make FEP converge faster. Using a similar analysis to that for CEP, a larger dimension number of problems n can increase the upper bound of FEP, but larger individual number k can lead to a smaller upper bound. Furthermore, $[-b, b]$ is the maximum interval bound for each dimension of the variables to be optimized (Definition 2). As a result, the second conclusion of Corollary 17 also implies that a larger search space will increase the upper bound of $E\mu_{EP}$.

If k becomes large enough, we have $(1 - m(\mathbf{S}^*(\varepsilon)) / (4b\pi)^n)^k \approx 1 - k(m(\mathbf{S}^*(\varepsilon)) / (4b\pi)^n)$ such that

$$E\mu_F \leq \frac{1 - \lambda_0^F}{km(\mathbf{S}^*(\varepsilon))} (4b\pi)^n. \quad (8)$$

Hence, the running time of FEP is nearly $O((4b\pi)^n)$ when $m(\mathbf{S}^*(\varepsilon))$ is a constant greater than zero.

Moreover, we can give an approximate condition where FEP can converge in polynomial time $P(n)$; that is, $m(\mathbf{S}^*(\varepsilon)) \geq ((1 - \lambda_0^F)/k) \cdot ((4b\pi)^n / P(n))$, $\forall \varepsilon > 0$ when (8) is true. The analysis is similar to the one for CEP at the end of Section 4.1.

5. Conclusion

In this paper, we propose a running time framework to calculate the mean running time of EP. Based on this framework, the convergence and mean running time of CEP and FEP with constant variation are studied. We also obtain some results at the worst running time of the considered EP, although the results show that the upper bounds can be tighter if the variation $\sigma = 2b$ where $2b$ is the length of the searching interval per dimension. It is shown that the individual number, problem dimension number, searching range, and the Lebesgue measure of the optimal neighborhood of the optimal solution have a direct impact on the bounds of the expected convergence times of the considered EP. Moreover, the larger the Lebesgue measure of the optimal neighborhood of the optimum, the lower the upper bound of the mean convergence time. In addition, the convergence time of the EP can be polynomial on average on the condition that the Lebesgue measure is greater than a value that is exponential to b .

However, it is possible to make an improvement on the running time framework and analysis given in this study. The deduction process in the proofs for Theorems 14 and 16 uses few properties of the distribution functions of the mutations, and so it is possible to tighten the results. By introducing more information on the specific mutation operations, more sound theoretical conclusions may be derivable. More importantly, unlike the rigorous conditions under which CEP and FEP can converge in polynomial time, the running time analysis of specific EP algorithms for real-world and constructive problems would have a more significant and practical impact. Future research could also focus on the runtime analysis of specific case studies of EP.

The proposed framework and results can be considered as a first step for analyzing the running time of evolutionary programming algorithms. We hope that our results can serve as a basis for further theoretical studies.

Appendix

Proof of the Lemmas, Theorems, and Corollaries

Proof of Lemma 7. Recall that $\xi_t^{\text{EP}} = (\mathbf{v}_1^{(t)}, \mathbf{v}_2^{(t)}, \dots, \mathbf{v}_k^{(t)})$ is a real vector and the status space Ω_{EP} of EP is continuous. Furthermore, ξ_t^{EP} is only dependent on ξ_{t-1}^{EP} for $t = 0, 1, \dots$, as evidenced by Steps (3)–(6) of the EP process. Hence, $\{\xi_t^{\text{EP}}\}_{t=0}^{+\infty}$ is a Markov chain. $\square$

Proof of Lemma 9. $\xi_t^{\text{EP}} \in \Omega_{EP}^*(\varepsilon)$ when $\exists(\mathbf{x}_i^{*(t)}, \sigma_i^{(t)}) \in \xi_t^{\text{EP}}$ where $\mathbf{x}_i^{*(t)}$ is an optimal solution. Given Steps (5)–(6) of EP, $\mathbf{x}_i^{*(t)}$ will have the most “wins” as well as the highest fitness, so $\mathbf{x}_i^{*(t)}$ will be selected for the next iteration with probability one

if there are less than k optimal solutions in the tournament. $\mathbf{x}_i^{*(t)}$ might be lost but other optimal solutions will be selected if there are more than k optimal solutions in the parent and offspring. Thus, $P\{\xi_{t+1}^{\text{EP}} \notin \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_t^{\text{EP}} \in \Omega_{\text{EP}}^*(\varepsilon)\} = 0$, and $\{\xi_t^{\text{EP}}\}_{t=0}^{+\infty}$ is an absorbing Markov chain to $\Omega_{\text{EP}}^*(\varepsilon)$. $\square$

Proof of Theorem 11. Observe that $\lambda_t^{\text{EP}} - \lambda_{t-1}^{\text{EP}} = P\{\mu_{\text{EP}} \leq t\} - P\{\mu_{\text{EP}} \leq t-1\} = P\{\mu_{\text{EP}} = t\}$. Therefore, we have

$$\begin{aligned} E\mu_{\text{EP}} &= \sum_{t=1}^{+\infty} t \cdot (\lambda_t^{\text{EP}} - \lambda_{t-1}^{\text{EP}}) \\ &= \lim_{N \rightarrow +\infty} \sum_{t=1}^N t \cdot (\lambda_t^{\text{EP}} - \lambda_{t-1}^{\text{EP}}) \\ &= \lim_{N \rightarrow +\infty} \sum_{i=1}^N (\lambda_N^{\text{EP}} - \lambda_{i-1}^{\text{EP}}) = \sum_{i=1}^{+\infty} (1 - \lambda_{i-1}^{\text{EP}}) \\ &= \sum_{i=0}^{+\infty} (1 - \lambda_i^{\text{EP}}). \end{aligned} \quad (\text{A.1})$$

Proof of Corollary 12. Let $p_i = P\{\xi_i^{\text{EP}} \in \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{i-1}^{\text{EP}} \notin \Omega_{\text{EP}}^*(\varepsilon)\}$. Based on the total probability equation, $\lambda_t^{\text{EP}} = (1 - \lambda_{t-1}^{\text{EP}})p_t + \lambda_{t-1}^{\text{EP}}(1 - p_t)$. Substituting it into Theorem 11, we have $E\mu_{\text{EP}} = \sum_{t=0}^{+\infty} [(1 - \lambda_0^{\text{EP}}) \prod_{i=1}^t (1 - p_i)]$. $\square$

Proof of Corollary 13. According to Corollary 12, $E\mu_{\text{EP}} \geq \sum_{t=0}^{+\infty} [(1 - \lambda_0^{\text{EP}}) \prod_{i=1}^t (1 - \beta_i)]$. Similarly, we have $E\mu_{\text{EP}} \leq \sum_{t=0}^{+\infty} [(1 - \lambda_0^{\text{EP}}) \prod_{i=1}^t (1 - \alpha_i)]$. Thus, $\beta^{-1}(1 - \lambda_0^{\text{EP}}) \leq E\mu_{\text{EP}} \leq \alpha^{-1}(1 - \lambda_0^{\text{EP}})$ when $\alpha_t = \alpha$ and $\beta_t = \beta$. $\square$

Proof of Theorem 14. (1) For fixed individual $(\mathbf{x}, \bar{\sigma})$, $\bar{\mathbf{x}} = \mathbf{x} + \mathbf{u} \otimes \bar{\sigma}$, where $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2, \dots, \bar{\sigma}_n)$ is n -dimension variation variable and the stochastic disturbance $\mathbf{u} = (u_1, u_2, \dots, u_n)$ follows n -dimensional standard Gaussian distribution. Let $\mathbf{z} = (z_1, z_2, \dots, z_n)$; that is, $\mathbf{z} = \mathbf{u} \otimes \bar{\sigma} = (u_1 \bar{\sigma}_1, u_2 \bar{\sigma}_2, \dots, u_n \bar{\sigma}_n)$. According to (1) and (2), $z_j = \mu_j \bar{\sigma}_j = \mu_j \sigma$. Then, we have $z_j \sim N(0, \bar{\sigma}_j^2)$ since $u_j \sim N(0, 1)$, where $j = 1, \dots, n$. Therefore, $\bar{\mathbf{x}} = \mathbf{x} + \mathbf{u} \otimes \bar{\sigma} = \mathbf{x} + \mathbf{z}$.

Given $\mathbf{S} = \prod_{j=1}^n [-b_j, b_j]$, $\mathbf{S}^*(\varepsilon) \subseteq \mathbf{S}$. Let $\tilde{\mathbf{S}} = \{\mathbf{z} \mid \mathbf{z} = \bar{\mathbf{x}} - \mathbf{x}, \bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon), \mathbf{x} \in \mathbf{S}\}$. According to the property of measure, $m(\mathbf{S}^*(\varepsilon)) = m(\tilde{\mathbf{S}})$. One has $P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\} = P\{\mathbf{z} \in \tilde{\mathbf{S}}\} = \int \cdots \int_{\tilde{\mathbf{S}}} \prod_{j=1}^n (1/\sqrt{2\pi\bar{\sigma}_j^2}) \exp\{-z_j^2/2\bar{\sigma}_j^2\} dz_1 \cdots dz_n$.

Note that $\bar{\mathbf{x}}, \mathbf{x} \in \mathbf{S}$ and $\mathbf{z} = \bar{\mathbf{x}} - \mathbf{x}$, so we have $|z_j| = |\bar{x}_j - x_j| \leq 2b_j = 2b$ ($j = 1, \dots, n$), which leads to the following inequality:

$$\begin{aligned} &\int \cdots \int_{\tilde{\mathbf{S}}} \prod_{j=1}^n \frac{1}{\sqrt{2\pi\bar{\sigma}_j^2}} \exp\left\{-\frac{z_j^2}{2\bar{\sigma}_j^2}\right\} dz_1 \cdots dz_n \\ &\geq \int \cdots \int_{\tilde{\mathbf{S}}} \prod_{j=1}^n \frac{1}{\sqrt{2\pi\bar{\sigma}_j^2}} \exp\left\{-\frac{(2b_j)^2}{2\bar{\sigma}_j^2}\right\} dz_1 \cdots dz_n \end{aligned}$$

$$\begin{aligned} &= m(\tilde{\mathbf{S}}) \prod_{j=1}^n \frac{1}{\sqrt{2\pi\bar{\sigma}_j^2}} \exp\left\{-\frac{2b_j^2}{\bar{\sigma}_j^2}\right\} \\ &= m(\tilde{\mathbf{S}}) \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\prod_{j=1}^n \frac{1}{\bar{\sigma}_j}\right) \exp\left\{-\sum_{j=1}^n \frac{2b_j^2}{\bar{\sigma}_j^2}\right\}. \end{aligned} \quad (\text{A.2})$$

That is, we obtain a lower bound of $P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\}$ as follows:

$$\begin{aligned} &P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\} \\ &\geq m(\tilde{\mathbf{S}}) \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\prod_{j=1}^n \frac{1}{\bar{\sigma}_j}\right) \exp\left\{-\sum_{j=1}^n \frac{2b_j^2}{\bar{\sigma}_j^2}\right\}. \end{aligned} \quad (\text{A.3})$$

Noting that $m(\mathbf{S}^*(\varepsilon)) = m(\tilde{\mathbf{S}})$, we also have $P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\} \geq m(\mathbf{S}^*(\varepsilon))(1/\sqrt{2\pi})^n (\prod_{j=1}^n (1/\bar{\sigma}_j)) \exp\{-\sum_{j=1}^n (2b_j^2/\bar{\sigma}_j^2)\}$. (2) Let $f(\bar{\sigma}) = \log(\prod_{j=1}^n (1/\bar{\sigma}_j) \exp\{-\sum_{j=1}^n (2b_j^2/\bar{\sigma}_j^2)\}) = -\sum_{j=1}^n (\log \bar{\sigma}_j + a_j/(1/\bar{\sigma}_j^2))$ where $a_j = 2b_j^2$. One has $\max f(\bar{\sigma}) \Leftrightarrow \min \sum_{j=1}^n (\log \bar{\sigma}_j + a_j/(1/\bar{\sigma}_j^2))$, and $d(\log \bar{\sigma}_j + a_j/(1/\bar{\sigma}_j^2))/d\bar{\sigma}_j = 0 \Rightarrow 1/\bar{\sigma}_j - 2a_j/(1/\bar{\sigma}_j^3) = 0 \Rightarrow \bar{\sigma}_j = \sqrt{2a_j}$.

Noticing that $a_j = 2b_j^2 = 2b^2$, we have $\bar{\sigma}_j = 2b$, which implies the lower bound of $P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\}$ can be improved into

$$\begin{aligned} &m(\mathbf{S}^*(\varepsilon)) \left(\frac{1}{\sqrt{2\pi}}\right)^n \prod_{j=1}^n \frac{1}{2b} \exp\left\{-\sum_{j=1}^n \frac{2b_j^2}{(2b)^2}\right\} \\ &= m(\mathbf{S}^*(\varepsilon)) \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-n/2} \prod_{j=1}^n \frac{1}{2b} \\ &= m(\mathbf{S}^*(\varepsilon)) \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-n/2} (2b)^{-n} \\ &= m(\mathbf{S}^*(\varepsilon)) (4\sqrt{e\pi}b)^{-n}. \end{aligned} \quad (\text{A.4})$$

Therefore, $P\{\bar{\mathbf{x}}_t^C \in \mathbf{S}^*(\varepsilon)\} \geq m(\mathbf{S}^*(\varepsilon))(4\sqrt{e\pi}b)^{-n}$.

(3) By the property of $\Omega_{\text{EP}}^*(\varepsilon)$ and $m(\mathbf{S}^*(\varepsilon))$ in Definition 5, $\forall \varepsilon > 0$ and $t = 1, 2, \dots$, $P\{\xi_t^C \in \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^C \notin \Omega_{\text{EP}}^*(\varepsilon)\} = 1 - \prod_{i=1}^k (1 - P\{\bar{\mathbf{x}}_i \in \mathbf{S}^*(\varepsilon)\})$. Thus, $P\{\xi_t^C \in \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^C \notin \Omega_{\text{EP}}^*(\varepsilon)\} \geq 1 - (1 - m(\mathbf{S}^*(\varepsilon))(4\sqrt{e\pi}b)^{-n})^k \forall \varepsilon > 0$. $\square$

Proof of Corollary 15. (1) Based on the total probability equation, we have

$$\begin{aligned} \lambda_t^C &= (1 - \lambda_{t-1}^C) P\{\xi_t^C \in \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^C \notin \Omega_{\text{EP}}^*(\varepsilon)\} \\ &\quad + \lambda_{t-1}^C P\{\xi_t^C \in \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^C \in \Omega_{\text{EP}}^*(\varepsilon)\}. \end{aligned} \quad (\text{A.5})$$

Because $\{\xi_t^C\}_{t=0}^{+\infty}$ can be considered as an absorbing Markov chain following Lemma 9, $P\{\xi_t^C \in \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^C \in \Omega_{\text{EP}}^*(\varepsilon)\} = 1$. Also, $\lambda_t^C = (1 - \lambda_{t-1}^C)P\{\xi_t^C \in \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^C \notin \Omega_{\text{EP}}^*(\varepsilon)\} + \lambda_{t-1}^C$,

and we have $1 - \lambda_t^{\text{CEP}} = (1 - \lambda_{t-1}^{\text{C}})(1 - P\{\xi_t^{\text{C}} \in \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^{\text{C}} \notin \Omega_{\text{EP}}^*(\varepsilon)\})$. According to Theorem 14,

$$P\{\xi_t^{\text{C}} \in \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^{\text{C}} \notin \Omega_{\text{EP}}^*(\varepsilon)\} \geq 1 - \left(1 - m(\mathbf{S}^*(\varepsilon)) \left(\frac{1}{4\sqrt{e}\pi b}\right)^n\right)^k. \quad (\text{A.6})$$

For $m(\mathbf{S}^*(\varepsilon))(1/4\sqrt{e}\pi b)^n > 0$, $d = 1 - (1 - m(\mathbf{S}^*(\varepsilon))(1/4\sqrt{e}\pi b)^n)^k > 0$ and $\lim_{t \rightarrow +\infty} (1 - d)^t = 0$. Thus, $1 - \lambda_t^{\text{C}} \leq (1 - d)(1 - \lambda_{t-1}^{\text{C}}) = (1 - \lambda_0^{\text{C}})(1 - d)^t$, so $\lim_{t \rightarrow +\infty} \lambda_t^{\text{C}} \geq 1 - (1 - \lambda_0^{\text{C}})\lim_{t \rightarrow +\infty} (1 - d)^t = 1$. Since $\lambda_t^{\text{C}} \leq 1$, $\lim_{t \rightarrow +\infty} \lambda_t^{\text{C}} = 1$.

(2) Given $\alpha = 1 - (1 - m(\mathbf{S}^*(\varepsilon))(1/4\sqrt{e}\pi b)^n)^k > 0$, we know that $P\{\xi_t^{\text{C}} \in \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^{\text{C}} \notin \Omega_{\text{EP}}^*(\varepsilon)\} \geq \alpha$ based on Theorem 14. According to Corollary 15, $\forall \varepsilon > 0$, $E\mu_{\text{C}} \leq (1 - \lambda_0^{\text{C}})(1 - (1 - m(\mathbf{S}^*(\varepsilon))(1/4\sqrt{e}\pi b)^n)^k)^{-1}$. $\square$

Proof of Theorem 16. (1) For fixed individual $(\mathbf{x}, \bar{\sigma})$, $\bar{\mathbf{x}} = \mathbf{x} + \delta \otimes \bar{\sigma}$, where the stochastic disturbance δ follows n -dimensional standard Cauchy distribution. Let $\mathbf{z} = \delta \otimes \bar{\sigma}$; that is, $z_j = \sigma \delta_j$ ($j = 1, \dots, n$). Then, we have $\bar{\mathbf{x}} = \mathbf{x} + \mathbf{z}$ and $z_j \sim (1/\sigma)C_{\phi=1}(y/\sigma)$ since $\delta_j \sim C_{\phi=1}(y)$ ($j = 1, \dots, n$).

In a manner similar to the proof of Theorem 14, we derive

$$\begin{aligned} P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\} &= P\{\mathbf{z} \in \tilde{\mathbf{S}}\} \\ &= \int \cdots \int_{\tilde{\mathbf{S}}} \left(\frac{1}{\sigma\pi}\right)^n \frac{1}{\prod_{j=1}^n (1 + z_j^2/\sigma^2)} dz_1 \cdots dz_n \\ &= \int \cdots \int_{\tilde{\mathbf{S}}} \left(\frac{1}{\sigma\pi}\right)^n \frac{1}{\prod_{j=1}^n (1 + z_j^2/\sigma^2)} dz_1 \cdots dz_n \\ &\geq \int \cdots \int_{\tilde{\mathbf{S}}} \left(\frac{1}{\sigma\pi}\right)^n \frac{1}{\prod_{j=1}^n (1 + 4b^2/\sigma^2)} dz_1 \cdots dz_n \quad (\text{A.7}) \\ &= \left(\frac{1}{\sigma\pi}\right)^n \frac{m(\mathbf{S}^*(\varepsilon))}{(1 + 4b^2/\sigma^2)^n} \\ &= \frac{m(\mathbf{S}^*(\varepsilon))}{(\pi)^n} \left(\sigma + \frac{4b^2}{\sigma}\right)^{-n}. \end{aligned}$$

(Note that $|z_j| = |\bar{x}_j - x_j| \leq 2b$ ($j = 1, \dots, n$)).

(2) Letting $f(\bar{\sigma}) = -\log(\sigma + 4b^2/\sigma)$, we have

$$\max f(\bar{\sigma}) \iff \min \left(\sigma + \frac{4b^2}{\sigma}\right), \quad (\text{A.8})$$

$$\frac{d(\sigma + 4b^2/\sigma)}{d\sigma} = 0 \implies 1 - \frac{4b^2}{\sigma^2} = 0 \implies \sigma = 2b.$$

It can be improved into $P\{\bar{\mathbf{x}} \in \mathbf{S}^*\} \geq m(\mathbf{S}^*(\varepsilon))/(4b\pi)^n$ when $\sigma = 2b$.

(3) Hence, $\forall \varepsilon > 0$ and $t = 1, 2, \dots$, if $\sigma = 2b$, $P\{\xi_t^{\text{F}} \notin \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^{\text{F}} \in \Omega_{\text{EP}}^*(\varepsilon)\} \geq 1 - (1 - m(\mathbf{S}^*(\varepsilon))/(4b\pi)^n)^k$. $\square$

Proof of Corollary 17. (1) According to Theorem 16, $P\{\xi_t^{\text{F}} \notin \Omega_{\text{EP}}^*(\varepsilon) \mid \xi_{t-1}^{\text{F}} \in \Omega_{\text{EP}}^*(\varepsilon)\} \geq 1 - (1 - m(\mathbf{S}^*(\varepsilon))/(4b\pi)^n)^k$.

$0 < m(\mathbf{S}^*(\varepsilon))/(4b\pi)^n < 1$ due to the second assumption in Section 2.2. Let $\theta = 1 - (1 - m(\mathbf{S}^*(\varepsilon))/(4b\pi)^n)^k$ for $t = 1, 2, \dots$. Then $0 < \theta < 1$ and $\lim_{t \rightarrow +\infty} (1 - \theta)^t = 0$.

Thus, $\lim_{t \rightarrow +\infty} \lambda_t^{\text{F}} \geq 1 - (1 - \lambda_0^{\text{F}})\lim_{t \rightarrow +\infty} (1 - \theta)^t = 1$. Also, $\lim_{t \rightarrow +\infty} \lambda_t^{\text{F}} = 1$ since $\lambda_t^{\text{F}} \leq 1$.

(2) Since $\lim_{t \rightarrow +\infty} \lambda_t^{\text{F}} = 1$, following Corollary 13 and Theorem 16, we find that

$$E\mu_{\text{F}} \leq \frac{(1 - \lambda_0^{\text{F}})}{(1 - (1 - m(\mathbf{S}^*(\varepsilon))/(4b\pi)^n)^k)}, \quad \forall \varepsilon > 0. \quad (\text{A.9})$$

$\square$

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by Humanity and Social Science Youth Foundation of Ministry of Education of China (14YJCZH216) and National Natural Science Foundation of China (61370177).

References

- [1] L. J. Fogel, A. J. Owens, and M. J. Walsh, *Artificial Intelligence through Simulated Evolution*, Wiley, New York, NY, USA, 1966.
- [2] D. B. Fogel, *System Identification Through Simulated Evolution: A Machine Learning Approach to Modeling*, Ginn Press, Washington, DC, USA, 1991.
- [3] D. B. Fogel, *Evolving artificial intelligence [Ph.D. thesis]*, University of California at San Diego, La Jolla, Calif, USA, 1992.
- [4] D. B. Fogel, "Applying evolutionary programming to selected traveling salesman problems," *Cybernetics and Systems*, vol. 24, no. 1, pp. 27–36, 1993.
- [5] T. Bäck and H. Schwefel, "An overview of evolutionary algorithms for parameter optimization," *Evolutionary Computation*, vol. 1, no. 1, pp. 1–23, 1993.
- [6] H.-P. Schwefel, *Evolution and Optimum Seeking*, Sixth-Generation Computer Technology Series, John Wiley & Sons, New York, NY, USA, 1995.
- [7] X. Yao, Y. Liu, and G. Lin, "Evolutionary programming made faster," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 2, pp. 82–102, 1999.
- [8] C.-Y. Lee and X. Yao, "Evolutionary programming using mutations based on the Lévy probability distribution," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 1, pp. 1–13, 2004.
- [9] G. Rudolph, "Self-adaptive mutations may lead to premature convergence," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 4, pp. 410–414, 2001.
- [10] G. Rudolph, "Convergence of non-elitist strategies," in *Proceedings of the 1st IEEE Conference on Evolutionary Computation*, pp. 63–66, June 1994.
- [11] G. Rudolph, "Convergence of evolutionary algorithms in general search spaces," in *Proceedings of the IEEE International Conference on Evolutionary Computation (ICEC '96)*, pp. 50–54, May 1996.

- [12] S. Droste, T. Jansen, and I. Wegener, "On the analysis of the (1+1) evolutionary algorithm," *Theoretical Computer Science*, vol. 276, no. 1-2, pp. 51–81, 2002.
- [13] J. He and X. Yao, "Drift analysis and average time complexity of evolutionary algorithms," *Artificial Intelligence*, vol. 127, no. 1, pp. 57–85, 2001.
- [14] M. Laumanns, L. Thiele, and E. Zitzler, "Running time analysis of multiobjective evolutionary algorithms on pseudo-Boolean functions," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 2, pp. 170–182, 2004.
- [15] R. Kumar and N. Banerjee, "Analysis of a multiobjective evolutionary algorithm on the 0-1 knapsack problem," *Theoretical Computer Science*, vol. 358, no. 1, pp. 104–120, 2006.
- [16] F. Neumann, "Expected runtimes of a simple evolutionary algorithm for the multi-objective minimum spanning tree problem," *European Journal of Operational Research*, vol. 181, no. 3, pp. 1620–1629, 2007.
- [17] Y. Yu and Z.-H. Zhou, "A new approach to estimating the expected first hitting time of evolutionary algorithms," *Artificial Intelligence*, vol. 172, no. 15, pp. 1809–1832, 2008.
- [18] T. Chen, K. Tang, G. Chen, and X. Yao, "A large population size can be unhelpful in evolutionary algorithms," *Theoretical Computer Science*, vol. 436, pp. 54–70, 2012.
- [19] P. K. Lehre and X. Yao, "On the impact of mutation-selection balance on the runtime of evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 16, no. 2, pp. 225–241, 2012.
- [20] C. Witt, "Population size versus runtime of a simple evolutionary algorithm," *Theoretical Computer Science*, vol. 403, no. 1, pp. 104–120, 2008.
- [21] T. Chen, J. He, G. Sun, G. Chen, and X. Yao, "A new approach for analyzing average time complexity of population-based evolutionary algorithms on unimodal problems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 5, pp. 1092–1106, 2009.
- [22] Y. Yu, X. Yao, and Z.-H. Zhou, "On the approximation ability of evolutionary optimization with application to minimum set cover," *Artificial Intelligence*, vol. 180, pp. 20–33, 2012.
- [23] Y. Zhou, J. He, and Q. Nie, "A comparative runtime analysis of heuristic algorithms for satisfiability problems," *Artificial Intelligence*, vol. 173, no. 2, pp. 240–257, 2009.
- [24] 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.
- [25] P. K. Lehre and X. Yao, "Runtime analysis of the (1 + 1) EA on computing unique input output sequences," *Information Sciences*, vol. 259, pp. 510–531, 2014.
- [26] 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.
- [27] Y. Zhou, J. Zhang, and Y. Wang, "Performance analysis of the (1+1) evolutionary algorithm for the multiprocessor scheduling problem," *Algorithmica*, pp. 1–21, 2014.
- [28] Y. Zhou, X. Lai, and K. Li, "Approximation and parameterized runtime analysis of evolutionary algorithms for the maximum cut problem," *IEEE Transactions on Cybernetics*, vol. 45, no. 8, pp. 1491–1498, 2014.
- [29] C. Witt, "Tight bounds on the optimization time of a randomized search heuristic on linear functions," *Combinatorics, Probability and Computing*, vol. 22, no. 2, pp. 294–318, 2013.
- [30] C. Witt, "Fitness levels with tail bounds for the analysis of randomized search heuristics," *Information Processing Letters*, vol. 114, no. 1-2, pp. 38–41, 2014.
- [31] D. Sudholt, "A new method for lower bounds on the running time of evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 3, pp. 418–435, 2013.
- [32] J. Jägersküpper, "Algorithmic analysis of a basic evolutionary algorithm for continuous optimization," *Theoretical Computer Science*, vol. 379, no. 3, pp. 329–347, 2007.
- [33] J. Jägersküpper, "Lower bounds for randomized direct search with isotropic sampling," *Operations Research Letters*, vol. 36, no. 3, pp. 327–332, 2008.
- [34] A. Agapie, M. Agapie, and G. Zbaganu, "Evolutionary algorithms for continuous-space optimisation," *International Journal of Systems Science*, vol. 44, no. 3, pp. 502–512, 2013.
- [35] A. Agapie, M. Agapie, G. Rudolph, and G. Zbaganu, "Convergence of evolutionary algorithms on the n-dimensional continuous space," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1462–1472, 2013.
- [36] Y. Chen, X. Zou, and J. He, "Drift conditions for estimating the first hitting times of evolutionary algorithms," *International Journal of Computer Mathematics*, vol. 88, no. 1, pp. 37–50, 2011.