

Running-time analysis of evolutionary programming based on Lebesgue measure of searching space

Yushan Zhang¹ · Han Huang² · Zhiyong Lin³ · Zhifeng Hao⁴ · Guiwu Hu¹

Received: 31 May 2015 / Accepted: 24 October 2016 / Published online: 19 November 2016
© The Natural Computing Applications Forum 2016

Abstract There have been many studies on the runtime analysis of evolutionary algorithms in discrete optimization, and however, relatively few homologous results have been obtained on continuous optimization, such as evolutionary programming (EP). This paper presents an analysis of the running time (as approximated by the mean first hitting time) of two EP algorithms based on Gaussian and Cauchy mutations, using an absorbing Markov process model. Given a constant variation, we analyze the running-time upper bound of special Gaussian mutation EP and Cauchy mutation EP, respectively. Our analysis shows that the upper bounds are impacted by individual number, problem dimension number, searching range, and the Lebesgue measure of the optimal neighborhood. Furthermore, we provide conditions whereby the mean running time of the considered EPs 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 computation of an exponential and the given polynomial of n . In the end, we present a case study on sphere function,

and the experiment validates the theoretical result in the case study.

Keywords Evolutionary algorithm · Running-time analysis · Evolutionary programming · Lebesgue measure · Optimal ε neighborhood

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 times 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, 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

✉ Han Huang
hhan@scut.edu.cn
Yushan Zhang
scuthill@163.com

¹ School of Mathematics and Statistics, Guangdong University of Finance and Economics, Guangzhou 510320, China

² School of Software Engineering, South China University of Technology, Guangzhou 510006, China

³ Department of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China

⁴ School of Mathematics and Big Data, Foshan University, Foshan 528000, China

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 EPs 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, 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. Rodolph [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, i.e., lacking of running-time analysis. To date, running-time analyses have mainly focused on Boolean-individual EAs like $(1 + 1)$ EA [12], $(N + N)$ EA [13], multi-objective EA [14–16]. Alternative theoretical measures for evaluating the running times 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. Jägerskpper conducted a rigorous

runtime analysis on $(1 + 1)$ ES, $(1 + \lambda)$ ES minimizing the sphere function [32, 33]. Agapie et al. modeled the $(1 + 1)$ ES as a renewal process under some reasonable assumptions and analyzed the first hitting time on inclined plane [34] and sphere function [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 [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 EPs can converge in a timespan less than a polynomial of n .

2 EP algorithms and their stochastic process model

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 reference [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 reference [7]. However, for ease of theoretical analysis, we will consider a kind of simple renewing function, i.e., constant function, in our case studies (see Sect. 4).

$$V(\sigma_{ij}) = \sigma \tag{2}$$

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

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

$$FEP : \quad \bar{x}_{ij} = x_{ij} + \bar{\sigma}_{ij}\delta_j \tag{4}$$

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

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

In step 5, for each individual in the set of all 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 1 (*minimization problem*) Let $S \subseteq \mathbf{R}^n$ be a finite subspace of the n -dimensional real domain \mathbf{R}^n , and let $f : S \rightarrow \mathbf{R}$ be an n -dimensional real function. A minimization problem, denoted by the 2-tuple (S, f) , is to find an n -dimensional vector $\mathbf{x}_{\min} \in S$ such that $\forall \mathbf{x} \in S, f(\mathbf{x}_{\min}) \leq f(\mathbf{x})$.

Without loss of generality, we can assume that $S = \prod_{i=1}^n [-b_i, b_i]$ where $b_i = b > 0$. The function $f : S \rightarrow \mathbf{R}$ 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 S is non-empty.

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

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 Stochastic process model of EP algorithms

Our running-time analyses are based on representing the EP algorithms as Markov processes. In this section, we explain the terminologies and notations used throughout the rest of this study.

Definition 2 (*stochastic process of EP*) The stochastic process of an evolutionary programming algorithm EP is denoted by $\{\zeta_t^{EP}\}_{t=0}^{+\infty}$, $\zeta_t^{EP} = (\mathbf{v}_1^{(t)}, \mathbf{v}_2^{(t)}, \dots, \mathbf{v}_k^{(t)})$, where $\mathbf{v}_i^{(t)} = (\mathbf{x}_i^{(t)}, \boldsymbol{\sigma}_i^{(t)})$ is the i th individual at the t th iteration.

The stochastic status ζ_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, i.e., $t \in \mathbf{Z}^+$.

Definition 3 (*status space*) The status space of EP is $\Omega_{EP} = \{(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) | \mathbf{v}_i = (\mathbf{x}_i, \boldsymbol{\sigma}_i), \mathbf{x}_i \in S, \sigma_{ij} \leq 2, j = 1, \dots, n\}$.

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

Definition 4 (*optimal status space*) The optimal status space of EP is the subset $\Omega_{EP}^* \subseteq \Omega_{EP}$ such that $\forall \zeta_{EP}^* \in \Omega_{EP}^*, \exists (\mathbf{x}^*, \boldsymbol{\sigma}) \in \zeta_{EP}^*$, where $\boldsymbol{\sigma} = (\sigma, \sigma, \dots, \sigma)$, and $\sigma > 0$.

Hence, all members of Ω_{EP}^* contains at least one optimal solution represented by individual \mathbf{x}^* .

Definition 5 (*Markov process*) A stochastic process $\{\zeta_t\}_{t=0}^{+\infty}$ with status space Ω is a Markov process if $\forall \tilde{\Omega} \subseteq \Omega, P\{\zeta_{t+1} \in \tilde{\Omega} | \zeta_0, \dots, \zeta_t\} = P\{\zeta_{t+1} \in \tilde{\Omega} | \zeta_t\}$.

Lemma 1 *The stochastic process $\{\zeta_t^{EP}\}_{t=0}^{+\infty}$ of EP is a Markov process.*

Proof The proof is given in “Appendix” section. $\{\xi_t^{\text{EP}}\}_{t=0}^{+\infty}$

We now show that the stochastic process of EP is an absorbing Markov process, defined as follows. \square

Definition 6 (absorbing Markov process to optimal status space) A Markov process $\{\xi_t\}_{t=0}^{+\infty}$ is an absorbing Markov process to Ω^* if $\exists \Omega^* \subset \Omega$, such that $P\{\xi_{t+1} \notin \Omega^* | \xi_t \in \Omega^*\} = 0$ for $t = 0, 1, \dots$

Lemma 2 The stochastic process $\{\xi_t^{\text{EP}}\}_{t=0}^{+\infty}$ of EP is an absorbing Markov process to Ω_{EP}^* .

Proof The proof is given in “Appendix” section. \square

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

3 General running-time 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 7 (first hitting time of EP) μ_{EP} is the first hitting time of EP if $\mu_{\text{EP}} = \min\{t \geq 0 : \xi_t^{\text{EP}} \in \Omega_{\text{EP}}^*\}$.

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 $E\mu_{\text{EP}}$ which can be calculated by Theorem 1. Hereinafter, we use the term *running time* and *first hitting time* interchangeably.

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

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

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

Proof The proof is given in “Appendix” section. \square

The conclusions of Theorem 1 and Corollary 1 are a direct approach to calculating the first hitting time of EP. However, Corollary 1 is more practical than Theorem 1 for this purpose since $p_i = P\{\xi_i^{\text{EP}} \in \Omega_{\text{EP}}^* | \xi_{i-1}^{\text{EP}} \notin \Omega_{\text{EP}}^*\}$, 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 1 is similar to the one in [17]. However, the estimating method [17] is based on a probability q_i in discrete optimization $q_i = \sum_{y \notin \Omega_{\text{EP}}^*} P\{\xi_i^{\text{EP}} \in \Omega_{\text{EP}}^* | \xi_{i-1}^{\text{EP}} = y\} P\{\xi_{i-1}^{\text{EP}} = y\}$.

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 2 and 3:

Corollary 2 If $\alpha_t \leq P\{\xi_t^{\text{EP}} \in \Omega_{\text{EP}}^* | \xi_{t-1}^{\text{EP}} \notin \Omega_{\text{EP}}^*\} \leq \beta_t$,

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

Proof The proof is given in “Appendix” section. \square

Corollary 2 indicates that $E\mu_{\text{EP}}$ can be studied by the lower bound and upper bound of $P\{\xi_t^{\text{EP}} \in \Omega_{\text{EP}}^* | \xi_{t-1}^{\text{EP}} \notin \Omega_{\text{EP}}^*\}$. Therefore, the theorem and corollaries introduce a first-hitting-time framework for the running-time analysis of EP. The running times of EPs based on Gaussian and Cauchy mutation are studied following the framework.

4 Running-time upper bounds of CEP and FEP

In this section, we use the framework proposed in Sect. 3 to study the running times of EPs based on Gaussian and Cauchy mutations, i.e., CEP and FEP. Moreover, our running-time analysis aims at a class of EPs with constant variation, as shown in Eq. (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], Bäck and Schwefel [5, 6]. The mutation of CEP is based on the standard normal distribution indicated by Eq. (3).

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

Theorem 2 Let the stochastic process of CEP be denoted by $\{\xi_t^{\text{C}}\}_{t=0}^{+\infty}$. Let k be the population size of CEP with solution space $\mathbf{S} = \prod_{i=1}^n [-b_i, b_i]$ where $b_i = b > 0$. Then $\forall \varepsilon > 0$, we have

- For fixed individual $(\mathbf{x}, \boldsymbol{\sigma})$, $P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\} \geq \frac{m(\mathbf{S}^*(\varepsilon))}{(\sqrt{2\pi})^n} \left(\prod_{j=1}^n \frac{1}{\sigma_j}\right) \exp\left\{-\sum_{j=1}^n \frac{2b_j^2}{\sigma_j^2}\right\}$ where the renewing function $V(\sigma_{ij}) = \sigma$.
- The right part of the inequality of conclusion (1) is maximum when $\sigma = 2b$.

$$3. P\{\zeta_t^C \in \Omega_{EP}^* | \zeta_{t-1}^C \notin \Omega_{EP}^*\} \geq 1 - (1 - \frac{m(\mathbf{S}^*(\varepsilon))}{(4\sqrt{\varepsilon\pi b})^n})^k \quad \text{if } \sigma = 2b.$$

Proof The proof is given in “Appendix” section. \square

In Theorem 2, Part (1) indicates a lower bound of the probability that an individual \mathbf{x} can be renewed to be the $\bar{\mathbf{x}}$ in the optimal neighborhood $\mathbf{S}^*(\varepsilon)$. 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 2 $P\{\zeta_t^C \in \Omega_{EP}^* | \zeta_{t-1}^C \notin \Omega_{EP}^*\}$ may have a tight lower bound shown by Part (3).

Theorem 2 indicates that a larger $m(\mathbf{S}^*(\varepsilon))$ leads to faster convergence for CEP since $P\{\zeta_t^C \in \Omega_{EP}^* | \zeta_{t-1}^C \notin \Omega_{EP}^*\}$ becomes larger when $m(\mathbf{S}^*(\varepsilon))$ is larger. Moreover, Theorem 2 produces the lower bound of $P\{\zeta_t^C \in \Omega_{EP}^* | \zeta_{t-1}^C \notin \Omega_{EP}^*\}$, with which the first hitting time of CEP can be estimated following Corollary 2. Corollary 3 indicates the convergence capacity and running-time upper bound of CEP.

Corollary 3 *Supposed conditions of Theorem 2 are satisfied,*

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

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

Proof The proof is given in “Appendix” section. \square

Corollary 3 shows that CEP converge globally and that the upper bound of CEP’s running time decreases as the Lebesgue measure of the optimal ε neighborhood $m(\mathbf{S}^*(\varepsilon))$ of $\mathbf{S}^*(\varepsilon)$ increases. A larger $m(\mathbf{S}^*(\varepsilon))$ translates to a larger 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 Eq. (6), $E\mu_C$ has a smaller upper bound when the population size k increases. Approximately,

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

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

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

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

4.2 Mean running time of FEP

FEP was proposed by Yao [7] as an improvement to CEP. The mutation for FEP is based on the Cauchy distribution indicated by Eq. (4). The property of the probability $p_t = P\{\zeta_t^F \in \Omega_{EP}^* | \zeta_{t-1}^F \notin \Omega_{EP}^*\}$ is discussed by Theorem 3 ($t = 0, 1, \dots$).

Theorem 3 *Let $\{\zeta_t^F\}_{t=0}^{+\infty}$ be the stochastic process of FEP,*

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

Proof The proof is given in “Appendix” section. \square

Similarly to Theorem 2, FEP with $\sigma = 2b$ may lead to a tight lower bound of the probability in Part (1) and (3). Theorem 3 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 4). Conversely, a bigger b (the bounds of search space) increases the difficulty of the problem for FEP. Given this convergence time framework, Corollary 4 summarizes the convergence capacity and running-time upper bound of FEP.

Corollary 4 *Supposed conditions of Theorem 3 are satisfied,*

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

Proof The proof is given in “Appendix” section. \square

Corollary 4 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 problem 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 1). As a result, the second conclusion of Corollary 4 also implies that a larger search space will increase the upper bound of $E\mu_{EP}$.

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

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

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 $P(n)$, i.e., $m(\mathbf{S}^*(\varepsilon)) \geq \frac{1 - \lambda_0^F}{k} \cdot \frac{(4b\pi)^n}{P(n)}$, $\forall \varepsilon > 0$ when Eq. (8) is true. The analysis is similar to the one for CEP at the end of Sect. 4.1.

4.3 Case study: simple EPs for sphere function problem

In this section, we analyzed the running time of simple EPs for a concrete problem presented by Definition 8.

Definition 8 A sphere function problem is to minimize the value of function $f(\mathbf{y}) = \sum_{i=1}^n y_i^2$ where $y_i \in [-1, 1]$.

Obviously, the optimal solution of sphere function problem is the vector $(0, 0, \dots, 0)$.

Definition 9 A simple EP is an EP algorithm such that $\sigma = (1, 1, \dots, 1)$, there is only one individual in the population and the variation renewing function $V(\sigma_j) = 1$ where $j = 1, \dots, n$.

Then, we have
Simple CEP:

$$\bar{x}_j = x_j + N_j(0, 1) \tag{9}$$

Simple FEP:

$$\bar{x}_j = x_j + \delta_j \tag{10}$$

We will use the proposed theories of running time to analyze simple CEP and FEP to investigate which algorithm may have a smaller upper bound of running time.

Given the solution space $\mathbf{S} = \prod_{j=1}^n [-1, 1]$, optimal neighborhood $\mathbf{S}^*(\varepsilon) = \{\mathbf{y} \in \mathbf{S} | f(\mathbf{y}) < \varepsilon\}$, $\varepsilon \leq 1$. Let $\tilde{\mathbf{S}} = \{\mathbf{z} | \mathbf{z} = \bar{\mathbf{x}} - \mathbf{x}, \bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon), \mathbf{x} \in \mathbf{S}\}$. We have the following results.

Theorem 4 Let μ_1 and μ_2 be the first hitting time of simple CEP and FEP, then

1. $E\mu_1 \leq \left(\frac{\sqrt{2\pi\varepsilon}}{\varepsilon}\right)^n$
2. $E\mu_2 \leq \left(\frac{2\pi}{\varepsilon}\right)^n$

Proof The proof is given in ‘‘Appendix’’ section. \square

As a result, we find simple CEP has a little smaller upper bound than simple FEP. CEP was proved to be a little better than FEP in solving sphere function problem by experimental data [7], respectively. Therefore, Theorem 4 might also reveal that Gauss distribution is better for sphere function problem than Cauchy distribution.

5 Experiment

Theorem 4 gives the upper bounds of the expected first hitting time for simple CEP and FEP on n -dimensional sphere function problem. We conduct experiment to validate the theoretical result, the experiment settings: fix the error $\varepsilon = 0.1$; fix the initial solution $\mathbf{x}_0 = (1, 1, \dots, 1)$; for simple CEP and FEP, respectively, we conduct 500 runs and denote T_i as the first hitting time at the i th run, then $\frac{\sum_{i=1}^{500} T_i}{500}$ is considered to be the estimation of the expected first hitting time.

From Fig. 1, we can see that the expected first hitting time of simple CEP and FEP grows exponentially when the dimension n of the sphere function problem increases, and the time complexity of the FEP is higher than that of CEP, which validates the theoretical analysis in Sect. 4.3. We need to point out that the time complexity of CEP and FEP grows so fast that when the dimension $n > 7$, the calculation of the first hitting time is too long to obtain a result on our computer, so we only present the case of $n \leq 7$.

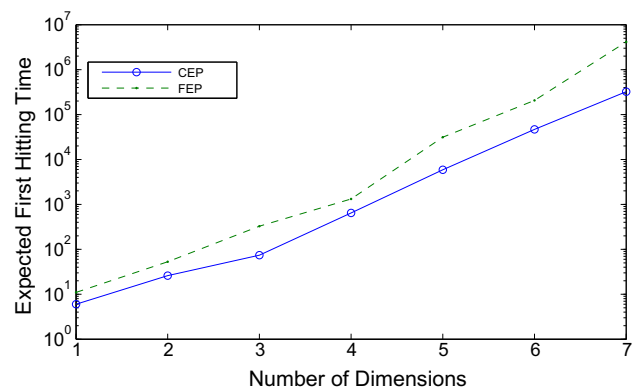


Fig. 1 Expected first hitting time for simple CEP and FEP on n -dimensional sphere function problem

6 Conclusion

In this paper, we propose a running-time framework to calculate the mean running time of EP and present a case study and experiment. 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 EPs, 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 EPs. 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 EPs 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 2 and 3 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 EPs.

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.

Acknowledgements This work is supported by National Natural Science Foundation of China (61370102, 61370177), Humanity and Social Science Youth Foundation of Ministry of Education of China (14YJCZH216), Guangdong Natural Science Funds for Distinguished Young Scholar (2014A030306050), the Fundamental Research Funds for the Central Universities, SCUT (2015PT022) and Guangdong High-level personnel of special support program (2014TQ01X664).

Appendix: Proof of the Lemmas, Theorems and Corollaries

Proof of Lemma 1 Recall that $\zeta_t^{EP} = (\mathbf{v}_1^{(t)}, \mathbf{v}_2^{(t)}, \dots, \mathbf{v}_k^{(t)})$ is a real vector and the status space Ω_{EP} of EP is continuous.

Furthermore, ζ_t^{EP} is only dependent on ζ_{t-1}^{EP} for $t = 0, 1, \dots$, as evidenced by steps 3–6 of the EP process. Hence, $\{\zeta_t^{EP}\}_{t=0}^{+\infty}$ is a Markov process.

Proof of Lemma 2 $\zeta_t^{EP} \in \Omega_{EP}^*$ when $\exists(\mathbf{x}_i^{*(t)}, \sigma_i^{(t)}) \in \zeta_t^{EP}$ where $\mathbf{x}_i^{*(t)}$ is an optimal solution. Given steps 5 and 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, $P\{\zeta_{t+1}^{EP} \notin \Omega_{EP}^* | \zeta_t^{EP} \in \Omega_{EP}^*\} = 0$, and $\{\zeta_t^{EP}\}_{t=0}^{+\infty}$ is an absorbing Markov process to Ω_{EP}^* .

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

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

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

Proof of Corollary 2 According to Corollary 1, $E\mu_{EP} \geq \sum_{t=0}^{+\infty} [(1 - \lambda_0^{EP}) \prod_{i=1}^t (1 - \beta_i)]$. Similarly, we have $E\mu_{EP} \leq \sum_{t=0}^{+\infty} [(1 - \lambda_0^{EP}) \prod_{i=1}^t (1 - \alpha_i)]$. Thus, $\beta^{-1}(1 - \lambda_0^{EP}) \leq E\mu_{EP} \leq \alpha^{-1}(1 - \lambda_0^{EP})$ when $\alpha_t = \alpha$ and $\beta_t = \beta$.

Proof of Theorem 2 For fixed individual $(\mathbf{x}, \bar{\sigma})$, $\bar{\mathbf{x}} = \mathbf{x} + \mathbf{u} \otimes \bar{\sigma}$, where $\sigma = (\bar{\sigma}_1, \bar{\sigma}_2, \dots, \bar{\sigma}_n)$ is a 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 Expressions (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 $\tilde{\mathbf{S}} = \{\mathbf{z} | \mathbf{z} = \bar{\mathbf{x}} - \mathbf{x}, \bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon), \mathbf{x} \in \mathbf{S}\}$. According to the property of measure, $m(\mathbf{S}^*(\varepsilon)) = m(\tilde{\mathbf{S}})$. $P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\} = P\{\mathbf{z} \in \tilde{\mathbf{S}}\} = \int \dots \int_{\tilde{\mathbf{S}}} \prod_{j=1}^n \frac{1}{\sqrt{2\pi\bar{\sigma}_j}} \exp\{-\frac{z_j^2}{2\bar{\sigma}_j^2}\} dz_1 \dots dz_n$.

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:

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

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

$$P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\} \geq m(\tilde{\mathbf{S}}) \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\prod_{j=1}^n \frac{1}{\sigma}\right) \exp\left\{-\sum_{j=1}^n \frac{2b^2}{\sigma^2}\right\}.$$

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

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

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

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

(3) By the property of Ω_{EP}^* and $m(\mathbf{S}^*(\varepsilon))$ in Definition 4, $\forall \varepsilon > 0$ and $t = 1, 2, \dots$, $P\{\xi_t^C \in \Omega_{EP}^* \mid \xi_{t-1}^C \notin \Omega_{EP}^*\} = 1 - \prod_{i=1}^k (1 - P\{\bar{\mathbf{x}}_i \in \mathbf{S}^*(\varepsilon)\})$

Thus, $P\{\xi_t^C \in \Omega_{EP}^* \mid \xi_{t-1}^C \notin \Omega_{EP}^*\} \geq 1 - (1 - m(\mathbf{S}^*)) (4\sqrt{e\pi}b)^{-n}$ for $\forall \varepsilon > 0$.

Proof of Corollary 3 1. Based on the total probability equation, we have

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

Because $\{\xi_t^C\}_{t=0}^{+\infty}$ can be considered as an absorbing Markov process following Lemma 2, $P\{\xi_t^C \in \Omega_{EP}^* \mid \xi_{t-1}^C \in \Omega_{EP}^*\} = 1 \cdot \lambda_t^C = (1 - \lambda_{t-1}^C) P\{\xi_t^C \in \Omega_{EP}^* \mid \xi_{t-1}^C \notin \Omega_{EP}^*\} + \lambda_{t-1}^C$, we have $1 - \lambda_t^{CEP} = (1 - \lambda_{t-1}^C)(1 - P\{\xi_t^C \in \Omega_{EP}^* \mid \xi_{t-1}^C \notin \Omega_{EP}^*\})$. According to Theorem 2,

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

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

2. Given $\alpha = 1 - \left(1 - m(\mathbf{S}^*(\varepsilon)) \left(\frac{1}{4\sqrt{e\pi}b}\right)^n\right)^k > 0$, we know that $P\{\xi_t^C \in \Omega_{EP}^* \mid \xi_{t-1}^C \notin \Omega_{EP}^*\} \geq \alpha$ based on Theorem 2. According to Corollary 3, $\forall \varepsilon > 0$, $E\mu_C \leq (1 - \lambda_0^C) \left(1 - \left(1 - m(\mathbf{S}^*(\varepsilon)) \left(\frac{1}{4\sqrt{e\pi}b}\right)^n\right)^k\right)^{-1}$.

Proof of Theorem 3 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 $z_j \sim \frac{1}{\sigma} C_{\phi=1}(\frac{\cdot}{\sigma})$ since $\delta_j \sim C_{\phi=1}(y)$ ($j = 1, \dots, n$).

In a manner similar to the Proof of Theorem 2, we derive

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

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

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

$$\max f(\bar{\sigma}) \Leftrightarrow \min(\sigma + \frac{4b^2}{\sigma}),$$

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

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

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

Proof of Corollary 4 1. According to Theorem 3, $P\{\zeta_t^F \notin \Omega_{EP}^* | \zeta_{t-1}^F \in \Omega_{EP}^*\} \geq 1 - (1 - \frac{m(\mathbf{S}^*(\varepsilon))}{(4b\pi)^n})^k$. $0 < \frac{m(\mathbf{S}^*(\varepsilon))}{(4b\pi)^n} < 1$

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

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

2. Since $\lim_{t \rightarrow +\infty} \lambda_t^F = 1$, following Corollary 2 and Theorem 4, we find that

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

Proof of Theorem 4 1. According to Eq. (9), for the only individual \mathbf{x} per iteration, $\bar{\mathbf{x}} = \mathbf{x} + \mathbf{u}$, where the stochastic disturbance \mathbf{u} follows n -dimensional standard Gauss distribution. Then, we have $\mathbf{u} = \bar{\mathbf{x}} - \mathbf{x}$ and $u_j \sim N(0, 1)$ ($j = 1, 2, \dots, n$). Following the Proof of Theorem 2,

$$\begin{aligned} P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\} &= P\{\mathbf{u} \in \tilde{\mathbf{S}}\} \\ &= \int \dots \int_{\tilde{\mathbf{S}}} \prod_{j=1}^n \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u_j^2}{2}\right\} du_1 \dots du_n \\ &\geq \int \dots \int_{\tilde{\mathbf{S}}} \prod_{j=1}^n \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\right\} du_1 \dots du_n \\ &= \int \dots \int_{\tilde{\mathbf{S}}} \left(\frac{1}{\sqrt{2\pi e}}\right)^n du_1 \dots du_n \\ &= m(\mathbf{S}^*(\varepsilon)) \left(\frac{1}{\sqrt{2\pi e}}\right)^n \\ &= \left(\frac{\varepsilon}{\sqrt{2\pi e}}\right)^n \end{aligned}$$

Based on Corollary 3, $E\mu_1 \leq (\frac{\sqrt{2\pi e}}{\varepsilon})^n$

2. According to Eq. (10), $\bar{\mathbf{x}} = \mathbf{x} + \boldsymbol{\delta}$ where the stochastic disturbance $\boldsymbol{\delta}$ follows n -dimensional standard Cauchy distribution. Then, we have $\boldsymbol{\delta} = \bar{\mathbf{x}} - \mathbf{x}$ and $\delta_j \sim C_{\phi=1}(y)$ ($j = 1, 2, \dots, n$).

Following the Proof of Theorem 3,

$$\begin{aligned} P\{\bar{\mathbf{x}} \in \mathbf{S}^*(\varepsilon)\} &= P\{\boldsymbol{\delta} \in \tilde{\mathbf{S}}\} \\ &= \int \dots \int_{\tilde{\mathbf{S}}} \prod_{j=1}^n C_{\phi=1}(\delta_j) d\delta_1 \dots d\delta_n \\ &= \int \dots \int_{\tilde{\mathbf{S}}} \left(\frac{1}{\pi}\right)^n \frac{1}{\prod_{j=1}^n (1 + \delta_j^2)} d\delta_1 \dots d\delta_n \\ &\geq \int \dots \int_{\tilde{\mathbf{S}}} \left(\frac{1}{\pi}\right)^n \frac{1}{\prod_{j=1}^n (1 + 1)} d\delta_1 \dots d\delta_n \\ &= \int \dots \int_{\tilde{\mathbf{S}}} \left(\frac{1}{2\pi}\right)^n d\delta_1 \dots d\delta_n \\ &= \left(\frac{\varepsilon}{2\pi}\right)^n \end{aligned}$$

Hence, $E\mu_2 \leq (\frac{2\pi}{\varepsilon})^n$ owing to Corollary 4.

References

1. Fogel LJ, Owens AJ, Walsh MJ (1996) Artificial intelligence through simulated evolution. Wiley, New York
2. Fogel DB (1991) System identification through simulated evolution: a machine learning approach to modeling. Ginn Press, Washington
3. Fogel DB (1992) Evolving artificial intelligence. Dissertation, University of California at San Diego
4. Fogel DB (1993) Applying evolutionary programming to selected traveling salesman problems. *Cybern Syst* 24(1):27–36
5. Bäck T, Schwefel HP (1993) An overview of evolutionary algorithms for parameter optimization. *Evol Comput* 1(1):1–23
6. Schwefel HP (1995) Evolution and optimum seeking. Wiley, New York
7. Yao X, Liu Y, Lin G (1999) Evolutionary programming made faster. *IEEE Trans Evol Comput* 3(2):82–102
8. Lee CY, Yao X (2004) Evolutionary programming using mutations based on the Levy probability distribution. *IEEE Trans Evol Comput* 8(1):1–13
9. Rudolph G (2001) Self-adaptive mutations may lead to premature convergence. *IEEE Trans Evol Comput* 5(4):410–414
10. Rudolph G (1994) Convergence of non-elitist strategies. *Proceedings of the first IEEE conference on evolutionary computation* 600:63–66
11. Rudolph G (1996) Convergence of evolutionary algorithms in general search spaces. *Proceedings of IEEE international conference on evolutionary computation* 180:50–54
12. Droste S, Jansen T, Wegener I (2002) On the analysis of the (1 + 1) evolutionary algorithm. *Theor Comput Sci* 276(1):51–81
13. He J, Yao X (2001) Drift analysis and average time complexity of evolutionary algorithms. *Artif Intell* 127(1):57–85
14. Laumanns M, Thiele L, Zitzler E (2004) Running time analysis of multiobjective evolutionary algorithms on pseudo-boolean functions. *IEEE Trans Evol Comput* 8(2):170–182
15. Kumar R, Banerjee N (2006) Analysis of a multiobjective evolutionary algorithm on the 0C1 knapsack problem. *Theor Comput Sci* 358(1):104–120

16. Neumann F (2007) Expected runtimes of a simple evolutionary algorithm for the multi-objective minimum spanning tree problem. *Eur J Oper Res* 181(3):1620–1629
17. Yu Y, Zhou ZH (2008) A new approach to estimating the expected first hitting time of evolutionary algorithms. *Artif Intell* 172(15):1809–1832
18. Chen T, Tang K, Chen G et al (2012) A large population size can be unhelpful in evolutionary algorithms. *Theor Comput Sci* 436:54–70
19. Lehre PK, Yao X (2012) On the impact of mutation-selection balance on the runtime of evolutionary algorithms. *IEEE Trans Evolut Comput* 16(2):225–241
20. Witt C (2008) Population size versus runtime of a simple evolutionary algorithm. *Theor Comput Sci* 403(1):104–120
21. Chen T, He J, Sun G et al (2009) A new approach for analyzing average time complexity of population-based evolutionary algorithms on unimodal problems. *IEEE Trans Syst Man Cybern Part B Cybern* 39(5):1092–1106
22. Yu Y, Yao X, Zhou ZH (2012) On the approximation ability of evolutionary optimization with application to minimum set cover. *Artif Intell* 180:20–33
23. Zhou Y, He J, Nie Q (2009) A comparative runtime analysis of heuristic algorithms for satisfiability problems. *Artif Intell* 173(2):240–257
24. Oliveto PS, He J, Yao X (2009) Analysis of the-EA for finding approximate solutions to vertex cover problems. *IEEE Trans Evol Comput* 13(5):1006–1029
25. Lehre PK, Yao X (2014) Runtime analysis of the $(1 + 1)$ EA on computing unique input output sequences. *Inf Sci Int J* 259:510–531
26. Lai X, Zhou Y, He J et al (2014) Performance analysis of evolutionary algorithms for the minimum label spanning tree problem. *IEEE Trans Evolut Comput* 18(6):860–872
27. Zhou Y, Zhang J, Wang Y (2014) Performance analysis of the $(1 + 1)$ Evolutionary Algorithm for the multiprocessor scheduling problem. *Algorithmica* 5:1–21
28. Zhou Y, Lai X, Li K (2014) Approximation and parameterized runtime analysis of Evolutionary Algorithms for the maximum cut problem. *IEEE Trans Cybern*. doi:10.1109/tcyb.2014.2354343
29. Witt C (2013) Tight bounds on the optimization time of a randomized search heuristic on linear functions. *Combin Probab Comput* 22(02):294–318
30. Witt C (2014) Fitness levels with tail bounds for the analysis of randomized search heuristics. *Inf Process Lett* 114(1):38–41
31. Sudholt D (2013) A new method for lower bounds on the running time of evolutionary algorithms. *IEEE Trans Evolut Comput* 17(3):418–435
32. Jägerskpper J (2007) Algorithmic analysis of a basic evolutionary algorithm for continuous optimization. *Theor Comput Sci* 379(3):329–347
33. Jägerskpper J (2008) Lower bounds for randomized direct search with isotropic sampling. *Oper Res Lett* 36(3):327–332
34. Agapie A, Agapie M, Baganu G (2013) Evolutionary algorithms for continuous-space optimisation. *Int J Syst Sci* 44(3):502–512
35. Agapie A, Agapie M, Rudolph G, Zbaganu G (2013) Convergence of evolutionary algorithms on the n -dimensional continuous space. *IEEE Trans Cybern* 43(5):1462–1472
36. Chen Y, Zou X, He J (2011) Drift conditions for estimating the first hitting times of evolutionary algorithms. *Int J Comput Math* 88(1):37–50