

# 连续型演化算法首次时间分析的平均增益模型

张宇山<sup>1)</sup> 黄翰<sup>2)</sup> 郝志峰<sup>3)</sup> 杨晓伟<sup>2)</sup>

<sup>1)</sup>(广东财经大学统计与数学学院 广州 510320)

<sup>2)</sup>(华南理工大学软件学院 广州 510006)

<sup>3)</sup>(佛山科学技术学院数学与大数据学院 广东 佛山 528000)

**摘要** 连续型演化算法(Evolutionary Algorithms, EAs)的计算时间分析(Runtime analysis)是演化计算理论研究中的难点和热点问题,相较于离散型演化算法,有关前者的理论结果相对较少,数学基础较为薄弱.该文引入鞅论和停时理论,建立了平均增益模型,以估算连续型演化算法的平均首次时间(Expected First Hitting Time, EFHT)上界.平均增益模型建立在一个非负随机过程的基础上,不依赖于算法具体的实现形式.论文介绍了如何应用该模型进行连续型演化算法的计算时间分析.作为案例分析,研究分析了:(1)带自适应步长的非精英(1,  $\lambda$ )ES(Evolution Strategy)求解球函数问题的平均首次时间,得到了3维情形下的时间上界的闭合表达式,并讨论了确保算法收敛条件下步长与子代种群规模 $\lambda$ 之间的关系;(2)(1+ $\lambda$ )ES求解2维倾斜平面问题的平均首次时间,得到了上界的闭合表达式.数值实验的结果表明实际的平均首次时间与理论计算的上界吻合.理论分析和实验结果表明平均增益模型有助于获得连续型演化算法平均首次时间紧致的上界,为连续型演化算法的计算时间分析提供了一种新的有效方法.

**关键词** 连续型演化算法;计算时间分析;首次时间;平均增益模型;鞅;停时理论

**中图法分类号** TP18 **DOI号** 10.11897/SP.J.1016.2019.00624

## An Average Gain Model to Analyze the First Hitting Time of Continuous Evolutionary Algorithms

ZHANG Yu-Shan<sup>1)</sup> HUANG Han<sup>2)</sup> HAO Zhi-Feng<sup>3)</sup> YANG Xiao-Wei<sup>2)</sup>

<sup>1)</sup>(School of Statistics and Mathematics, Guangdong University of Finance and Economics, Guangzhou 510320)

<sup>2)</sup>(School of Software Engineering, South China University of Technology, Guangzhou 510006)

<sup>3)</sup>(School of Mathematics and Big Data, Foshan University, Foshan, Guangdong 528000)

**Abstract** Runtime analysis of continuous evolutionary algorithms (EAs) is a challenging, hard and hot topic in the theoretical foundation research of evolutionary computation, relatively few theoretical results have been derived compared to the discrete EAs. In this paper, we introduce the stopping time and martingale theory to establish an average gain model to estimate the upper bound for the expected first hitting time of continuous EAs. We regard the first hitting time as a stopping time, the proposed model is established on a non-negative stochastic process, and does not depend on the specific realization of the algorithm. Roughly speaking, the average gain can be considered to be the average one step size varying with the generation's status. In the part of theoretical framework, firstly, we use martingale's property and the Lebesgue dominated convergence theorem to prove a theoretical results on the upper bound of the first hitting time

收稿日期:2017-01-16;在线出版日期:2018-01-19.本课题得到国家自然科学基金(61876207)、教育部-中国移动科研基金(MCM20160206)、广东省自然科学基金杰出青年基金项目(2014A030306050)、广东省特支计划人才专项课题(2014TQ01X664)、广州市科学技术局项目(201707010227)资助.张宇山,男,1986年生,博士,副教授,主要研究方向为演化算法的理论基础. E-mail: scuthill@163.com.黄翰(通信作者),男,1980年生,博士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为演化计算方法的理论基础、演化计算方法的优化设计及其应用. E-mail: hhan@scut.edu.cn.郝志峰,男,1969年生,博士,教授,博士生导师,主要研究领域为机器学习、计算智能、代数学及组合优化.杨晓伟,男,1969年生,博士,教授,博士生导师,主要研究领域为机器学习、计算智能、数据挖掘.

when the lower bound of the average gain is a constant, then, we consider the case that the lower bound of the average gain depends on the  $t$ -th generation of population, and derive a tight upper bound for the first hitting time of continuous EAs through integral computation. To demonstrate how the theoretical framework is applied to analyzing the first hitting time of continuous EAs on specific problem, we select two classical problems as case studies. We analyze (1) the expected first hitting time of the non-elitist  $(1, \lambda)$ ES with adaptive step-size solving the Sphere function problem, and derive a closed-form expression of the first hitting time upper bound for 3-dimensional case. We also discuss the relationship between the step size and the offspring size  $\lambda$  to ensure the convergence of EA; (2) the expected first hitting time of the  $(1 + \lambda)$ ES with uniform distribution mutation operator on the 2-dimensional inclined plane, and derive a closed-form expression of the time upper bound. The numerical experimental results show that the practical expected first hitting time is not only lower than the theoretical upper bound but also very close, this means that the proposed model can be used to obtain tight upper bounds for the expected first hitting time of continuous EAs. Both the theoretical analysis and the numerical experimental results show that our model can be effective for the precise runtime analysis of continuous evolutionary algorithms.

**Keywords** continuous evolutionary algorithm; runtime analysis; first hitting time; average gain model; martingale; stopping time theory

## 1 引言

演化算法 (Evolutionary Algorithms, EAs) 是一类受自然进化过程启发而产生的自适应搜索算法。随着演化算法在各类优化问题的应用中取得丰硕的成果, 此类算法的理论研究日益引起学者们的关注<sup>[1-2]</sup>。

计算时间 (runtime) 分析是当前演化算法理论研究的热点问题。直观来讲, 计算时间分析的目标是分析算法在运行中至少找到一个最优解或者好的近似最优解所需的适应值评估次数。计算时间可用首次到达相关的随机过程的某个状态集的时间来衡量<sup>[3]</sup>。由于演化算法的随机性本质, 这类算法的计算时间分析并不容易。

早期的研究主要关注简单演化算法如  $(1+1)$ EA 求解具有良好结构特性的伪布尔函数的计算时间<sup>[4-6]</sup>。这些研究展示了一些有用的数学方法和工具, 并得到了关于某些实例的理论结果。当前,  $(1+1)$ EA 的计算时间分析已经逐渐从简单伪布尔函数拓展到具有实际应用背景的组合优化问题。Oliveto 等人分析了  $(1+1)$ EA 求解顶点覆盖问题的某些实例的计算时间<sup>[7]</sup>。Lehre 等人精选了计算唯一输入输出序列问题的几个实例, 对  $(1+1)$ EA

的计算时间作了分析<sup>[8]</sup>。Zhou 等人针对  $(1+1)$ EA 求解以下几个组合优化问题的某些实例进行了一系列的近似性能分析: 最小标签生成树问题<sup>[9]</sup>、多处理器调度问题<sup>[10]</sup>、最大割问题<sup>[11]</sup>以及最大叶子生成树问题<sup>[12]</sup>, 取得了丰硕的理论成果。

伴随着  $(1+1)$ EA 理论研究的进展, 很多数学方法和工具被提出, 如马尔可夫链<sup>[13]</sup>、与收敛速率相结合的吸收马尔可夫链<sup>[14]</sup>、转换分析 (switch analysis)<sup>[15]</sup>、基于适应值分割的方法<sup>[16-17]</sup>, 等等。特别地, 由 He 等人引入的漂移分析 (drift analysis)<sup>[6]</sup>已被证明是演化算法计算时间分析的一种强有力的技巧。Jägersküpper 将马尔可夫链分析与漂移分析相结合, 重新考察了“标准  $(1+1)$ EA 在线性函数上的计算时间”这一问题, 显著改进了期望计算时间的上界<sup>[18]</sup>。通过把漂移分析和接管时间 (takeover time) 概念结合起来, Chen 等人从理论上讨论了基于种群的演化算法在单模问题上的时间复杂度<sup>[19]</sup>。漂移分析的一些变体也被提出来了, Oliveto 等人<sup>[20]</sup>引进了一个简化的漂移分析定理, 用以证明  $(1+1)$ EA 在 Needle、OneMax 以及 Maximum Matching 问题上的计算时间下界。Rowe 等人<sup>[21]</sup>提出了可变漂移 (variable drift), 即漂移量随着当前的状态单调递减, 以此求得  $(1, \lambda)$ EA 在 OneMax 问题上计算时间从指数式变为多项式的阈值是  $\lambda \approx$

$5 \log_{10} n$ . Witt<sup>[22]</sup> 使用乘法漂移分析 (multiplicative drift analysis) 得到了  $(1+1)$ EA 在线性函数上平均计算时间的紧致的界. 这些研究充分展示了漂移分析是演化算法计算时间分析中的强有力工具, 而且针对它的改进是当前的一个研究重点. 演化算法在带噪声的环境中常常显示出较强的鲁棒性, 但是有关的理论研究很少. Droste<sup>[23]</sup> 分析了  $(1+1)$ EA 在带噪声的 OneMax 函数上的计算时间, 得到了多项式时间复杂度下的最大噪声强度. 在此基础上, Qian 等学者探讨了带噪声的适应值函数是怎样影响演化算法的计算时间<sup>[24]</sup>, 通过严格的计算时间分析指出, 采样可以显著地加快演化算法在带噪声环境中的优化时间<sup>[25]</sup>.

综上所述, 现有的关于演化算法计算时间分析的工作主要集中于离散优化问题, 较少关于演化算法在连续搜索空间上的计算时间分析. 通过研究算法趋于最优解时的单步平均空间增益, Jägersküpper<sup>[26-27]</sup> 运用瓦尔德等式首次对  $(1+1)$ ES 和  $(1+\lambda)$ ES 求解球函数问题的计算时间进行了严格而冗长的分析. Agapie 等人<sup>[28]</sup> 在较强的假设条件下, 建立了  $(1+1)$ ES 的更新过程模型, 并分析了  $(1+1)$ ES 求解 2 维倾斜平面问题的首达时间. 张宇山等人<sup>[29]</sup> 基于停时理论, 将首达时间视为停时, 提出了分析演化算法首达时间的新模型, 并以此分析了  $(1+\lambda)$ ES 在倾斜平面问题上的平均首达时间. 从理论上讲, 漂移分析适用于离散优化和连续优化情形, 然而, 关于漂移分析应用于后者的理论结果却不多见<sup>[30]</sup>. 受漂移分析的思想启发, 黄翰等人<sup>[31]</sup> 提出了平均增益模型以估计  $(1+1)$ EA 在球函数问题上的平均计算时间. 然而, 文献[31]所提方法具有局限性, 只是针对连续型  $(1+1)$ EA 求解球函数的计算时间这一具体案例建立模型, 所有的理论结果(模型、定理、推论)都严重依赖于此具体案例, 故在某种意义上是专用的(ad hoc), 使得该模型与具体的案例纠缠在一起, 没能抽象上升为一般的模型. 此外, 该文原始的证明是基于黎曼积分, 因此数学上的严格性和一般性都需要加强.

本文的目的在于将文献[31]所提出的模型进一步严格化, 并加以推广, 使之成为分析演化算法计算时间的通用模型. 为了将所提的模型从特定的算法和问题分离出来, 本文将模型建立在一个非负随机过程  $\{X_t; t=0, 1, 2, \dots\}$  之上, 从而脱离了具体的案例, 使其理论高度得以提升. 通过引入鞅论, 将首达

时间视为停时. 本文将鞅论和停时理论结合起来, 建立起估计连续型演化算法平均首达时间上界的通用公式. 在实例分析部分也比文献[31]更进一步, 我们用所提的方法分析了: (1) 自适应步长的非精英  $(1, \lambda)$ ES 求解球函数问题的计算时间, 并讨论了步长和子代种群规模  $\lambda$  的关系; (2) 带均匀分布变异算子的  $(1+\lambda)$ ES 求解 2 维倾斜平面问题的平均首达时间. 数值试验结果表明, 本文的理论分析与实际情形相一致. 与  $(1+1)$ ES 和  $(1+\lambda)$ ES 在球函数问题和倾斜平面问题上的计算时间分析<sup>[26-28]</sup> 相比较, 本文的方法更简便且可以得到更紧的时间上界.

## 2 问题描述与数学建模

### 2.1 连续型演化算法

本小节简要介绍连续型演化算法的基本知识.

不失一般性, 假定本文的演化算法被用来处理连续搜索空间中的最小化问题. 给定搜索空间  $S \subset R^n$  和函数  $f: S \rightarrow R$ , 我们的任务是要找到至少一个全局最优化解  $x^* \in S$ , 使得函数的最小值  $f^* \triangleq f(x^*) = \min_{x \in S} f(x)$ .

一个典型的演化算法求解问题的步骤可以简要描述如下<sup>[30]</sup>.

#### 算法 1. 演化算法.

输入: 初始种群

输出: 包含到目前为止找到的最好解的种群

1. 随机地生成一个初始种群  $P = \{\xi_1, \xi_2, \dots, \xi_\mu\}$ ;
2. WHILE(终止条件不满足)
  - (1) 通过交叉和/或变异从  $P$  中产生一个子代  $P' = \{\xi'_1, \xi'_2, \dots, \xi'_\lambda\}$ ;
  - (2) 通过从  $P \cup P'$  或  $P'$  中选择  $\mu$  个个体产生一个新的父代种群  $P$ ;
3. 输出到目前为止找到的含最好解的种群.

将第  $t$  代种群表示为  $P_t = \{\xi_1^t, \xi_2^t, \dots, \xi_\mu^t\}$ ,  $t = 0, 1, 2, \dots$ . 令  $f(P_t) \triangleq \min\{f(x) : x \in P_t\}$  为种群  $P_t$  的适应值, 定义  $X_t = f(P_t) - f^*$ , 则  $X_t$  可被视为基于种群的 EA 到最优解的某种特定的距离值. 由于考虑的是最小化优化问题, 显然,  $\{X_t\}_{t=0}^\infty$  是一个非负随机过程.

### 2.2 上鞅与停时

本论文分析的对象是表达为  $\{X_t\}_{t=0}^\infty$  的离散时间非负随机过程, 其中  $X_t \geq 0, t = 0, 1, \dots$ . 就演化算法而言,  $X_t$  可以代表  $(1+1)$ EA 求解 OneMax 问题

时第  $t$  代的位串中 0-位的个数、基于种群的 EA 到最优解的某种特定的距离值, 等等.

根据随机过程理论<sup>[32]</sup>, 设  $(\Omega, \mathcal{F}, P)$  是一个概率空间,  $\{Y_t\}_{t=0}^{\infty}$  为  $(\Omega, \mathcal{F}, P)$  上的随机过程.  $\mathcal{F}_t = \sigma(Y_0, Y_1, \dots, Y_t) \subset \mathcal{F}$ ,  $t=0, 1, \dots$  为  $\mathcal{F}$  的自然  $\sigma$ -代数流.  $\sigma$ -代数  $\mathcal{F}_t$  包含了由  $Y_0, Y_1, \dots, Y_t$  生成的所有事件, 即直到时刻  $t$  为止的全部信息. 显然,  $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n \subset \dots$ .

在概率论中, 鞅是公平赌博的数学模型, 描述过去事件的知识无助于使未来期望收益最大. 具体来讲, 鞅是一列随机变量即随机过程, 在此过程中的任一特定时刻, 下一时刻的平均值等于当前时刻的观察值, 即使之前所有观测值的信息已知.

鞅描述的是公平赌博, 而上鞅和下鞅则分别描述了不利赌博与有利赌博, 这两者的当前观测值未必等于未来值的条件期望, 前者的当前观测值在条件数学期望的意义下是未来值的上界, 而后者是下界. 下面给出上鞅的正式定义.

**定义 1**<sup>[32]</sup>. 设  $\{F_t, t \geq 0\}$  为  $\mathcal{F}$  的单调递增子  $\sigma$ -代数序列. 称随机过程  $\{S_t\}_{t=0}^{\infty}$  为关于  $\{F_t, t \geq 0\}$  或  $\{Y_t\}_{t=0}^{\infty}$  的上鞅, 如果  $\{S_t\}_{t=0}^{\infty}$  是  $\{F_t, t \geq 0\}$  适应的 (即对任意的  $t=0, 1, 2, \dots$ ,  $S_t$  在  $F_t$  上可测),  $E(|S_t|) < +\infty$ , 且对任意的  $t=0, 1, 2, \dots$ , 有  $E(S_{t+1} | F_t) \leq S_t$ .

定义 1 表明, 上鞅在平均意义上是递减的.

徐宗本等人<sup>[33]</sup>首次运用鞅论研究了遗传算法的几乎必然强收敛性, 这为演化算法的理论研究提供了一种全新的分析工具, 我们在此基础上, 运用鞅论来研究演化算法的计算时间.

在演化算法的计算时间分析中, 人们通常感兴趣于在一次运行的过程中, 算法首次找到最优解所需的迭代次数, 这就是所谓的“首达时间”<sup>[6]</sup>, 可以用停时来描述它. 在概率论中, 特别是随机过程的研究中, 停时 (亦称为马尔可夫时间) 是一种特殊类型的“随机时间”: 一个随机变量, 它的取值代表某个给定的随机过程表现出特定的行为的时间. 停时的正式定义如下.

**定义 2**<sup>[32]</sup>. 设  $\{Y_t\}_{t=0}^{\infty}$  是一个随机过程,  $T$  是一个非负整值随机变量. 若对任意的  $n=0, 1, 2, \dots$ , 有  $\{T \leq n\} \in \mathcal{F}_n = \sigma(Y_0, Y_1, \dots, Y_n)$ , 则称  $T$  为关于  $\{Y_t\}_{t=0}^{\infty}$  的停时.

设  $\mathcal{H}_n = \sigma(X_0, X_1, \dots, X_n)$ ,  $n \geq 0$ ,  $T_0 = \min\{t \geq 0: X_t = 0\}$ . 可得  $\forall n \geq 0, \{T_0 \leq n\} = \bigcup_{k=0}^n \{T_0 = k\} =$

$\bigcup_{k=0}^n \{X_0 > 0, \dots, X_{k-1} > 0, X_k = 0\} \in \mathcal{H}_n$ , 故首达时间  $T_0$  是一个关于  $\{X_t\}_{t=0}^{\infty}$  的停时.

**引理 1**<sup>[32]</sup>. 设  $\{S_t\}_{t=0}^{\infty}$  是关于  $\{Y_t\}_{t=0}^{\infty}$  的上鞅, 且  $T$  是关于  $\{Y_t\}_{t=0}^{\infty}$  的有界停时, 则

$$E(S_T | \mathcal{F}_0) \leq S_0 \quad (1)$$

在式 (1) 中, 上鞅  $\{S_t\}_{t=0}^{\infty}$  的下标从固定的时间  $t$  改为随机变量一停时  $T$ , 而  $S_0$  依然保持为上界.

**引理 2**. 设  $s \wedge T_0 = \min\{s, T_0\}$ ,  $s=0, 1, 2, \dots$ , 则  $s \wedge T_0$  对任意固定的  $s$  是关于  $\{X_t\}_{t=0}^{\infty}$  的有界停时.

证明. 当  $s$  固定时, 显然  $s \wedge T_0$  是有界的. 接下来我们将证明  $s \wedge T_0$  是一个停时.

$$\forall m \geq 0, \{s \wedge T_0 \leq m\} = \{s \leq m\} \cup \{T_0 \leq m\}.$$

(1) 当  $m \geq s$ ,  $\{s \leq m\} \cup \{T_0 \leq m\} = \Omega \cup \{T_0 \leq m\} = \Omega \in \mathcal{H}_m$ ;

(2) 当  $m < s$ ,  $\{s \leq m\} \cup \{T_0 \leq m\} = \Phi \cup \{T_0 \leq m\} = \{T_0 \leq m\} \in \mathcal{H}_m$ .

根据定义 2, 结论得证.

证毕.

### 3 平均增益模型

平均增益模型来源于文献<sup>[31]</sup>, 其中一步平均变化量  $\delta_t = E(X_t - X_{t+1} | \mathcal{H}_t)$ ,  $t \geq 0$  被称为平均增益. 基于平均增益,  $T_0$  的期望的上界可以估计如下.

**定理 1**. 设  $\{X_t\}_{t=0}^{\infty}$  为一个随机过程, 满足对任意的  $t \geq 0$ ,  $X_t \geq 0$ . 假定  $E(T_0) < \infty$ , 若  $\forall t \geq 0$ ,  $E(X_t - X_{t+1} | \mathcal{H}_t) \geq \alpha > 0$ , 则  $E(T_0 | X_0) \leq \frac{X_0}{\alpha}$ .

证明. 定义  $Z_t = X_t + t\alpha$ ,  $t=0, 1, 2, \dots$ , 则

$$\begin{aligned} E(Z_{t+1} | \mathcal{H}_t) &= E(X_{t+1} + (t+1)\alpha | \mathcal{H}_t) \\ &= E(X_t - (X_t - X_{t+1}) + (t+1)\alpha | \mathcal{H}_t) \\ &= X_t - E(X_t - X_{t+1} | \mathcal{H}_t) + (t+1)\alpha \\ &\quad (\text{条件期望的性质}) \\ &\leq X_t - \alpha + (t+1)\alpha \\ &= X_t + t\alpha = Z_t. \end{aligned}$$

由定义 1,  $\{Z_t\}_{t=0}^{\infty}$  是一个关于  $\{X_t\}_{t=0}^{\infty}$  的上鞅. 根据引理 1 和 2, 可得

$$E(Z_{s \wedge T_0} | \mathcal{H}_0) = E(Z_{s \wedge T_0} | X_0) \leq Z_0 = X_0.$$

于是

$$E(X_{s \wedge T_0} + (s \wedge T_0)\alpha | X_0) =$$

$$E(X_{s \wedge T_0} | X_0) + \alpha E(s \wedge T_0 | X_0) \leq X_0.$$

由于  $s \wedge T_0 \leq T_0$ ,  $X_{s \wedge T_0} \geq 0$ , 运用勒贝格控制收

敛定理,得

$$X_0 \geq \lim_{s \rightarrow \infty} E(X_s \wedge T_0 | X_0) + \alpha \lim_{s \rightarrow \infty} E(s \wedge T_0 | X_0) \\ \geq \alpha E(T_0 | X_0).$$

这意味着  $E(T_0 | X_0) \leq \frac{X_0}{\alpha}$ . 证毕.

定理 1 与文献[6]中的定理 1 的结论有相似之处,不同之处在于本文运用鞅论和控制收敛定理进行了严格的数学证明,此为其一;

其二,文献[6]中的定理 1 只适用离散优化情形,本文的定理 1 是为下文的定理 2 做铺垫,定理 2 可以用于连续优化情形,理由如下:

定理 1 中的  $\alpha$  是一个不依赖于  $X_t$  的常数,故必须使用  $\delta_t$  在  $t=0,1,2,\dots$  时的一致下界,很多情况下该一致下界会非常小,甚至接近于零,这将导致首达时间  $T_0$  的上界非常松.此外, $T_0$  通常适用于离散优化中的演化算法,对于连续型演化算法,我们感兴趣的是对于  $\epsilon$ -近似解的首达时间,可表达为  $T_\epsilon = \min\{t \geq 0: X_t \leq \epsilon\}$ . 基于定理 1,可得到关于  $T_\epsilon$  的上界的通用定理.

**定理 2.** 设  $\{X_t\}_{t=0}^\infty$  是一个随机过程,对任意的  $t \geq 0$ ,均有  $X_t \geq 0$ . 令  $h: [0, A] \rightarrow \mathbb{R}^+$  为一单调递增可积函数. 若当  $X_t > \epsilon > 0$  时,  $E(X_t - X_{t+1} | \mathcal{H}_t) \geq h(X_t)$ , 则对  $T_\epsilon$  成立

$$E(T_\epsilon | X_0) \leq 1 + \int_\epsilon^{X_0} \frac{1}{h(x)} dx \quad (2)$$

证明. 令  $g(x) = \begin{cases} 0, & x \leq \epsilon \\ \int_\epsilon^x \frac{1}{h(t)} dt + 1, & x > \epsilon \end{cases}$ , 可得

(1) 当  $x > \epsilon, y \leq \epsilon$  时,

$$g(x) - g(y) = \int_\epsilon^x \frac{1}{h(t)} dt + 1 \geq 1.$$

(2) 当  $x > \epsilon, y > \epsilon$  时,

$$g(x) - g(y) = \int_y^x \frac{1}{h(t)} dt \geq \frac{x-y}{h(x)}.$$

于是有

(1') 当  $X_t > \epsilon, X_{t+1} \leq \epsilon$  时,

$$E(g(X_t) - g(X_{t+1}) | \mathcal{H}_t) \geq 1.$$

(2') 当  $X_t > \epsilon, X_{t+1} > \epsilon$  时,

$$E(g(X_t) - g(X_{t+1}) | \mathcal{H}_t) \\ \geq E\left(\frac{X_t - X_{t+1}}{h(X_t)} \middle| \mathcal{H}_t\right) \\ = \frac{1}{h(X_t)} E(X_t - X_{t+1} | \mathcal{H}_t) \geq 1.$$

根据以上推导,可得当  $X_t > \epsilon > 0$  时,  $E(g(X_t) - g(X_{t+1}) | \mathcal{H}_t) \geq 1$ . 注意到  $T_\epsilon = \min\{t \geq 0: X_t \leq \epsilon\} = \min\{t \geq 0: g(X_t) = 0\} \triangleq T_0^g$ . 假定  $X_0 > \epsilon$ , 由定理 1, 得到

$$E(T_\epsilon | X_0) = E(T_0^g | g(X_0)) \leq \frac{g(X_0)}{1} \\ = 1 + \int_\epsilon^{X_0} \frac{1}{h(t)} dt. \quad \text{证毕.}$$

在定理 2 中,平均增益  $\delta_t$  的下界  $h(X_t)$  随  $X_t$  的变化而变化,故无需寻找  $\delta_t$  在  $t=0,1,2,\dots$  上的一致下界,这有助于获得首达时间  $T_\epsilon$  的更紧上界. 从定理 2 可推得以下的特例.

**推论 1.** 设  $\{X_t\}_{t=0}^\infty$  是一个随机过程,对任意的  $t \geq 0$ ,有  $X_t \geq 0$ . 若存在  $0 < q \leq 1$ ,使得  $E(X_t - X_{t+1} | \mathcal{H}_t) \geq qX_t (X_t > \epsilon > 0)$ , 则对  $T_\epsilon$  成立

$$E(T_\epsilon | X_0) \leq \frac{1}{q} \ln\left(\frac{X_0}{\epsilon}\right) + 1 \quad (3)$$

证明. 令  $h(x) = qx$ , 显然  $h(x)$  单调递增,可积. 由定理 2 可得

$$E(T_\epsilon | X_0) \leq \int_\epsilon^{X_0} \frac{1}{h(x)} dx + 1 = \int_\epsilon^{X_0} \frac{1}{qx} dx + 1 \\ = \frac{1}{q} \ln\left(\frac{X_0}{\epsilon}\right) + 1. \quad \text{证毕.}$$

令  $T_\epsilon = \min\{t \geq 0: X_t \leq \epsilon\}$  为连续型演化算法找到  $\epsilon$ -近似解的首达时间,平均增益  $\delta_t$  的计算在估计  $T_\epsilon$  的上界时起到关键作用.

对于大部分演化算法,种群  $P_{t+1}$  的状态仅仅依赖于种群  $P_t$ . 在这种情形下,随机过程  $\{P_t\}_{t=0}^\infty$  可用马尔可夫链来建模<sup>[6,13]</sup>. 相应的,  $\{X_t\}_{t=0}^\infty$  也可视为马尔可夫链. 在这样的场合中,平均增益  $\delta_t = E(X_t - X_{t+1} | \mathcal{H}_t)$  可被简化为  $\delta_t = E(X_t - X_{t+1} | X_t)$ , 这将使得计算相对更简便.

对于具有马尔可夫性的演化算法,如果采用与第 2 节相同的记号,则只需把定理 1、2 和推论 1 中的  $E(X_t - X_{t+1} | \mathcal{H}_t)$  改为  $E(X_t - X_{t+1} | X_t)$ , 所有的结论依然成立. 以下是相应的结果.

**定理 1'.** 设  $\{X_t\}_{t=0}^\infty$  是一个与演化算法相关联的随机过程,对任意的  $t \geq 0$ ,有  $X_t \geq 0$ . 假定  $E(T_0) < +\infty$ . 若对任意的  $t \geq 0$ ,有  $E(X_t - X_{t+1} | X_t) \geq \alpha > 0$ , 则  $E(T_0 | X_0) \leq \frac{X_0}{\alpha}$ .

**定理 2'.** 设  $\{X_t\}_{t=0}^\infty$  是一个与演化算法相关联的随机过程,对任意的  $t \geq 0$ ,有  $X_t \geq 0$ . 设  $h: [0, A] \rightarrow$

$R^+$  是一个单调递增, 可积函数. 如果在  $X_t > \epsilon > 0$  时, 有  $E(X_t - X_{t+1} | X_t) \geq h(X_t)$ , 则对  $T_\epsilon$  有  $E(T_\epsilon | X_0) \leq 1 + \int_\epsilon^{X_0} \frac{1}{h(x)} dx$ .

**推论 1'.** 设  $\{X_t\}_{t=0}^\infty$  是一个与演化算法相关联的随机过程, 对任意的  $t \geq 0$ , 有  $X_t \geq 0$ . 如果存在  $0 < q \leq 1$ , 使得  $E(X_t - X_{t+1} | X_t) \geq qX_t (X_t > \epsilon > 0)$ , 则对  $T_\epsilon$  成立

$$E(T_\epsilon | X_0) \leq \frac{1}{q} \ln\left(\frac{X_0}{\epsilon}\right) + 1.$$

在下一节, 我们将使用所提出的方法分析自适应步长的  $(1, \lambda)$ ES 在球函数问题上的计算时间以及  $(1 + \lambda)$ ES 在倾斜平面问题上的计算时间.

## 4 案例分析

### 4.1 $(1, \lambda)$ ES 在球函数问题上的平均首达时间分析

本小节我们考察非精英  $(1, \lambda)$ ES: 在每一步迭代中, 一个父代个体通过变异产生  $\lambda$  个子代个体, 然后其中适应值最小的个体被选出来作为下一代的父代个体. 值得注意的是新的父代个体不一定比旧的父代个体更优.  $(1, \lambda)$ ES 的流程可描述如下.

#### 算法 2. $(1, \lambda)$ ES.

输入: 初始个体  $x_0$  和初始步长  $l_0$ .

输出: 满足给定精度  $\epsilon$  的解

1. 初始化  $x_0, l_0$ , 置  $t = 0$
2. WHILE(终止条件未满足)
  - (1) FOR  $i = 1$  to  $\lambda$ 

$$y_{t,i} = x_t + l_t \cdot u$$
  - (2) ENDFOR
  - (3)  $x_{t+1} = \arg \min_{x \in \{y_{t,1}, \dots, y_{t,\lambda}\}} f(x)$
  - (4) 调整  $l_{t+1}$
  - (5)  $t \leftarrow t + 1$
3. 输出解

其中的  $u$  是一个均匀分布在  $n$  维单位超球面上的随机向量,  $l_t$  是步长或称变异超球的半径. 我们考虑球函数  $f(x) = \sum_{i=0}^n x_i^2$  的最小化, 显然最小值为  $f^* = f(0, 0, \dots, 0) = 0$ .

假设算法已经到达某个点  $x_t \in S, t = 0, 1, \dots$ ,

定义  $X_t = f(x_t) - f^*$ . 由  $x_t$  产生的新点(解)分布在一个半径为  $l_t$ , 中心为  $x_t$  的超球表面上. 由于问题的等值线和变异超球在旋转后依然是不变的, 故可分析问题在平面上的投影, 如图 1 所示. 令

$f(x_t) = R^2, f(y_{t,i}) = r_i^2, i = 1, 2, \dots, \lambda$ . 半径为  $R$  的大圆代表问题  $f(x) = R^2$  的等值线, 其圆心即为最优解. 半径为  $l$  的小圆代表变异超球. 由于对称性, 我们可将分析局限于  $\omega \in [0, \pi]$  情形.

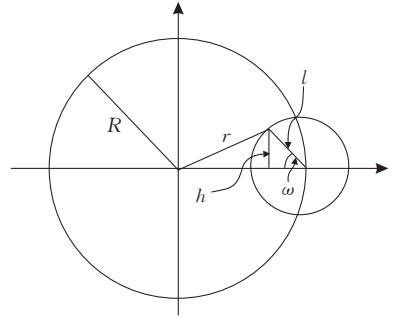


图 1 参数空间的横截面

根据余弦定理,  $r^2 = R^2 - 2lR \cos \omega + l^2$ , 于是  $R^2 - r^2 = 2lR \cos \omega - l^2$ . 为了应用推论 1, 需计算  $E\left(\frac{X_t - X_{t+1}}{X_t} \middle| X_t\right)$ , 为此, 我们定义相对增量  $V$  为

$$V = \frac{R^2 - r^2}{R^2} = 2a \cos \omega - a^2 \quad (4)$$

其中  $a = l/R$ .

此处的角度  $\omega$  是一个随机变量, 因为  $r$  是随机生成的. 关于角度  $\omega$  的概率密度函数, 有以下的引理.

**引理 3**<sup>[34]</sup>. 角度  $\omega$  的概率密度函数为

$$p_\omega(x) = \frac{\sin^{n-2} x}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)} \cdot 1_{[0, \pi]}(x) \quad (5)$$

其中  $B(\cdot, \cdot)$  代表 Beta 函数,  $1_A(x)$  表示集合  $A$  的示性函数.

为了避免复杂而冗长的计算, 不失一般性, 我们在余下的部分考虑  $n = 3$  的情形.

如式(4)所示, 相对增量  $V$  由角度  $\omega$  决定, 故也是一个随机变量, 它的概率密度函数可计算如下.

**引理 4.** 对于  $n = 3$ , 给定  $X_t = R^2$ , 相对增量  $V$  的概率密度函数为

$$p_V(v) = \frac{1}{4a} \cdot 1_{S_a}(v) \quad (6)$$

其中  $S_a = [-2a - a^2, 2a - a^2]$ .

证明. 对于  $n = 3$ , 注意到  $B\left(\frac{1}{2}, 1\right) = 2$ , 故根据引理 3 得到  $p_\omega(x) = \frac{\sin x}{2} \cdot 1_{[0, \pi]}(x)$ .

由  $V = 2a \cos \omega - a^2$  得  $\omega = \arccos\left(\frac{V + a^2}{2a}\right)$ ,

故

$$p_V(v) = p_w \left[ \arccos \left( \frac{v+a^2}{2a} \right) \right] \cdot \left| \left[ \arccos \left( \frac{v+a^2}{2a} \right) \right]'_v \right|,$$

$$2a \cos \pi - a^2 \leq v \leq 2a \cos 0 - a^2.$$

经过化简,得到  $p_V(v) = \frac{1}{4a} \cdot 1_{S_a}(v)$ . 也就是说

$V$  服从区间  $S_a = [-2a - a^2, 2a - a^2]$  上的均匀分布. 证毕.

根据算法 2, 当  $(1, \lambda)$  ES 到达点  $x_t (t=0, 1, \dots)$  时, 产生  $\lambda$  个子代个体, 表示为  $y_{t,i} (i=1, 2, \dots, \lambda)$ , 然后  $x_{t+1} = \operatorname{argmin}_{x \in \{y_{t,1}, \dots, y_{t,\lambda}\}} f(x)$  被选为下一代的父代个体. 令  $f(y_{t,i}) = r_i^2, V_i = \frac{R^2 - r_i^2}{R^2}$ , 则  $V_i (i=1, 2, \dots, \lambda)$  与  $V$  独立同分布.

由于  $x_{t+1} = \operatorname{argmin}_{x \in \{y_{t,1}, \dots, y_{t,\lambda}\}} f(x)$ , 这等价于  $\frac{X_t - X_{t+1}}{X_t} = \max\{V_i; i=1, 2, \dots, \lambda\}$  并记为  $V_{\max}$ . 其概率密度函数计算如下.

**引理 5.** 给定  $X_t = R^2$ , 则  $V_{\max}$  的概率密度函数为

$$p_{V_{\max}}(x) = \lambda \cdot p_V(x) \cdot P^{\lambda-1} \{V \leq x\} \cdot 1_{S_a}(x) \quad (7)$$

证明.  $V_{\max}$  的分布函数为

$$P\{V_{\max} \leq x\} = P\{V_1 \leq x, \dots, V_\lambda \leq x\}$$

$$= P^\lambda \{V \leq x\},$$

因此

$$p_{V_{\max}}(x) = \frac{dP^\lambda \{V \leq x\}}{dx}$$

$$= \lambda \cdot p_V(x) \cdot P^{\lambda-1} \{V \leq x\} \cdot 1_{S_a}(x). \text{ 证毕.}$$

令  $T_\epsilon = \min\{t \geq 0; X_t \leq \epsilon\}$ , 我们基于推论 1 估计首次达时间  $T_\epsilon$  的上界.

**定理 3.** 当  $a = l_t / \sqrt{X_t} = (\lambda - 1) / (\lambda + 1), \lambda \geq 2$  时, 对于  $T_\epsilon$  有下式成立

$$E(T_\epsilon | X_0) \leq \left( \frac{\lambda + 1}{\lambda - 1} \right)^2 \ln \left( \frac{X_0}{\epsilon} \right) + 1 \quad (8)$$

证明. 根据引理 4、5, 可得

$$E \left( \frac{X_t - X_{t+1}}{X_t} \middle| X_t \right) = E(V_{\max} | X_t)$$

$$= \int_{-2a-a^2}^{2a-a^2} x \cdot p_{V_{\max}}(x) dx$$

$$= \int_{-2a-a^2}^{2a-a^2} x \cdot \lambda \cdot \frac{1}{4a} \cdot \left( \frac{x+2a+a^2}{4a} \right)^{\lambda-1} dx$$

$$= 2a \cdot \frac{\lambda-1}{\lambda+1} - a^2 \quad (9)$$

将式(9)关于  $a$  求导, 可得当  $a^* = \frac{\lambda-1}{\lambda+1}, \lambda \geq 2$

时, 式(9)取得其最大值  $\left( \frac{\lambda-1}{\lambda+1} \right)^2 \in (0, 1)$ , 其中  $\lambda \geq 2$ .

根据推论 1, 下式成立

$$E(T_\epsilon | X_0) \leq \left( \frac{\lambda+1}{\lambda-1} \right)^2 \ln \left( \frac{X_0}{\epsilon} \right) + 1. \text{ 证毕.}$$

定理 3 表明如果选择步长  $l_t = \left( \frac{\lambda-1}{\lambda+1} \right) \cdot \sqrt{X_t}$ , 则非精英  $(1, \lambda)$  ES 能够收敛到球函数的最优解的  $\epsilon$ -邻域且平均首次达时间的上界为  $\left( \frac{\lambda+1}{\lambda-1} \right)^2 \ln \left( \frac{X_0}{\epsilon} \right) + 1$ .

## 4.2 $(1+\lambda)$ ES 在倾斜平面问题上的平均首次达时间分析

倾斜平面(inclined plane)问题是一个基本的连续优化问题, Agapie 等人<sup>[28]</sup>引入更新过程理论, 分析了带均匀分布变异算子的  $(1+1)$  ES 在该问题上的平均运行时间. 本小节将他们的分析推广到  $(1+\lambda)$  ES 情形. 本小节讨论的目标函数是二维倾斜平面(inclined plane)<sup>[28]</sup>:

$$f(x_1, x_2) = x_2, (x_1, x_2) \in S = [-a, a] \times [0, a], a > 0.$$

显然,  $\max_{(x_1, x_2) \in S} f(x_1, x_2) = a$ , 当  $x_2 = a$  时取到最大值.

图 2 描述了倾斜平面投影到  $x_1 O x_2$  面的示意图, 平面沿着坐标轴  $x_2$  正向以 45 度角倾斜.

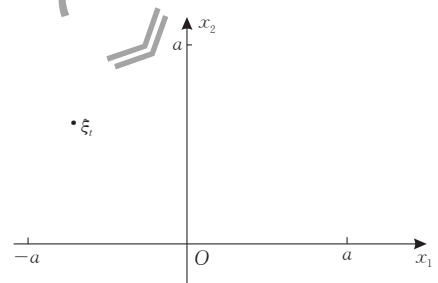


图 2 倾斜平面的投影

$(1+\lambda)$  ES 的算法流程如下所示.

**算法 3.**  $(1+\lambda)$  ES.

输入: 初始个体;

输出: 最优个体;

1. 初始化, 随机生成 1 个个体  $x = (x_1, x_2, \dots, x_n) \in R^n$ ;
2. 通过变异产生  $\lambda$  个中间个体;
3. 从这  $1+\lambda$  个个体中选择 1 个最佳个体;
4. 若终止条件满足, 则停止; 否则回到 2.

步骤 2 的变异算子采用  $[-0.5, 0.5]$  上的均匀

分布: 对于父代个体  $\mathbf{x}=(x_1, x_2, \dots, x_n)$  中的每个分量  $x_i (i=1, 2, \dots, n)$ , 令  $y_i = x_i + z_i, z_i \sim U[-0.5, 0.5]$ ,  $z_1, \dots, z_n$  相互独立, 生成一个中间个体  $\mathbf{y}=(y_1, y_2, \dots, y_n)$ . 对父代个体  $\mathbf{x}$  重复  $\lambda$  次变异操作, 共独立产生  $\lambda$  个中间个体  $\mathbf{y}$ .

由于本小节讨论的是二维连续优化问题, 故可设  $\xi_t=(x_1^t, x_2^t) \in S \subset R^2, t=0, 1, \dots$  为算法的第  $t$  代个体. 定义  $X_t = a - f(\xi_t) = a - x_2^t$ , 记  $\eta_t = X_t - X_{t+1} = x_2^{t+1} - x_2^t, t=0, 1, 2, \dots$ . 下面的引理描述了  $\eta_t (t=0, 1, 2, \dots)$  的分布.

**引理 6.**  $\eta_t (t=0, 1, 2, \dots)$  独立同分布, 分布函数  $F(x) = \mathbb{P}\{\eta_t \leq x\}$  为

$$F(x) = \begin{cases} 0, & x < 0 \\ \left(\frac{1}{2}\right)^\lambda, & x = 0 \\ \left(x + \frac{1}{2}\right)^\lambda, & 0 < x \leq \frac{1}{2} \\ 1, & x > \frac{1}{2} \end{cases} \quad (10)$$

证明. 记  $x_2^t (t=0, 1, \dots)$  表示第  $t$  代个体的第 2 个分量,  $\tilde{x}_{2,1}^t, \tilde{x}_{2,2}^t, \dots, \tilde{x}_{2,\lambda}^t$  分别为第  $t$  代个体经变异后产生的  $\lambda$  个中间个体的第 2 个分量, 则  $\tilde{x}_{2,k}^t = x_2^t + z_k, z_k \sim U[-0.5, 0.5] (k=1, 2, \dots, \lambda)$  且相互独立. 于是第  $t+1$  代个体的第 2 个分量为

$$x_2^{t+1} = \begin{cases} x_2^t, & \max_{k=1,2,\dots,\lambda} \tilde{x}_{2,k}^t \leq x_2^t \\ \max_{k=1,2,\dots,\lambda} \tilde{x}_{2,k}^t, & \max_{k=1,2,\dots,\lambda} \tilde{x}_{2,k}^t > x_2^t \end{cases}$$

根据  $\eta_t (t=0, 1, \dots)$  的定义, 可得

$$\eta_t = \begin{cases} 0, & \max_{k=1,2,\dots,\lambda} z_k \leq 0 \\ \max_{k=1,2,\dots,\lambda} z_k, & 0 < \max_{k=1,2,\dots,\lambda} z_k \end{cases}$$

由算法的搜索过程可知, 每一次迭代时均匀随机变量  $z_k (k=1, 2, \dots, \lambda)$  都是独立产生的, 故  $\eta_t (t=0, 1, \dots)$  独立同分布, 其分布函数  $F(x) = \mathbb{P}\{\eta_t \leq x\}$  为

(1) 当  $x < 0$  时:  $F(x) = 0$ ;

(2) 当  $x = 0$  时:

$$\begin{aligned} F(x) &= \mathbb{P}\{\eta_t \leq x\} = \mathbb{P}\{\eta_t = 0\} = \mathbb{P}\left\{\max_{k=1,2,\dots,\lambda} z_k \leq 0\right\} \\ &= \prod_{k=1}^{\lambda} \mathbb{P}\{z_k \leq 0\} = \left(\frac{1}{2}\right)^\lambda; \end{aligned}$$

(3) 当  $0 < x \leq \frac{1}{2}$  时:

$$\begin{aligned} F(x) &= \mathbb{P}\{\eta_t \leq x\} = \mathbb{P}\left\{\max_{k=1,2,\dots,\lambda} z_k \leq x\right\} \\ &= \prod_{k=1}^{\lambda} \mathbb{P}\{z_k \leq x\} = \left(x + \frac{1}{2}\right)^\lambda; \end{aligned}$$

(4) 当  $x > \frac{1}{2}$  时:  $F(x) = 1$

$$\text{即 } F(x) = \begin{cases} 0, & x < 0 \\ \left(\frac{1}{2}\right)^\lambda, & x = 0 \\ \left(x + \frac{1}{2}\right)^\lambda, & 0 < x \leq \frac{1}{2} \\ 1, & x > \frac{1}{2} \end{cases} \quad \text{证毕.}$$

不妨假定算法在初始化时从原点出发, 即  $\xi_0 = (x_1^0, x_2^0) = (0, 0)$ , 此时  $X_0 = a$ , 由定理 2 可得如下结论.

**定理 4.** 在本案例中, 首达时间  $T_\epsilon$  满足

$$E(T_\epsilon | X_0) \leq 1 + \frac{2(\lambda+1)(a-\epsilon)}{\lambda-1 + \left(\frac{1}{2}\right)^\lambda} \quad (11)$$

证明. 对于  $t=0, 1, 2, \dots$

$$\begin{aligned} E(X_t - X_{t+1} | X_t) &= E(\eta_t | X_t) \\ &= \int_{-\infty}^{+\infty} x dF(x) \\ &= \int_0^{\frac{1}{2}} x \cdot \lambda \left(x + \frac{1}{2}\right)^{\lambda-1} dx \\ &= \int_0^{\frac{1}{2}} \left(x + \frac{1}{2}\right) \cdot \lambda \left(x + \frac{1}{2}\right)^{\lambda-1} dx - \\ &\quad \int_0^{\frac{1}{2}} \frac{1}{2} \cdot \lambda \left(x + \frac{1}{2}\right)^{\lambda-1} dx \\ &= \int_0^{\frac{1}{2}} \lambda \left(x + \frac{1}{2}\right)^\lambda d\left(x + \frac{1}{2}\right) - \int_0^{\frac{1}{2}} \frac{1}{2} d\left(x + \frac{1}{2}\right)^\lambda \\ &= \frac{\lambda}{\lambda+1} \left(x + \frac{1}{2}\right)^{\lambda+1} \Big|_0^{\frac{1}{2}} - \frac{1}{2} \left(x + \frac{1}{2}\right)^\lambda \Big|_0^{\frac{1}{2}} \\ &= \frac{\lambda-1 + \left(\frac{1}{2}\right)^\lambda}{2(\lambda+1)}. \end{aligned}$$

于是对于  $\epsilon > 0$ , 成立

$$\begin{aligned} E(T_\epsilon | X_0) &\leq 1 + \int_\epsilon^a \frac{2(\lambda+1)}{\lambda-1 + \left(\frac{1}{2}\right)^\lambda} dx \\ &= 1 + \frac{2(\lambda+1)(a-\epsilon)}{\lambda-1 + \left(\frac{1}{2}\right)^\lambda}. \quad \text{证毕.} \end{aligned}$$

## 5 数值实验

### 5.1 (1, $\lambda$ ) ES 求解球函数问题

定理 3 给出了 3 维情形下平均首达时间上界的



解析表达式. 我们通过实验来验证理论结果, 实验设置为: 固定误差  $\epsilon=0.01$ ; 置  $\mathbf{x}_0 = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$ , 此时  $X_0=1$ ; 对每一个  $\lambda$ , 将算法 2 在 3 维球函数上运行 300 轮; 用  $T_{ei}$  表示第  $i$  轮的  $\epsilon$ -邻域首达时间; 定

义  $\widehat{E(T_\epsilon | X_0)} = \frac{\sum_{i=1}^{300} T_{ei}}{300}$ , 可被认为是平均首达时间的估计量. 表 1 比较了实际的平均首达时间  $\widehat{E(T_\epsilon | X_0)}$  与理论的时间上界  $\left(\frac{\lambda+1}{\lambda-1}\right)^2 \ln\left(\frac{X_0}{\epsilon}\right)+1$ .

表 1  $(1, \lambda)$ ES 平均首达时间和理论上限的比较

$\lambda$	$\widehat{E(T_\epsilon   X_0)}$	$\left(\frac{\lambda+1}{\lambda-1}\right)^2 \ln\left(\frac{X_0}{\epsilon}\right)+1$
3	8.57	19.42
5	6.89	11.36
15	4.13	7.01
25	3.27	6.40
50	2.23	5.99

从表 1 可以看出,  $\widehat{E(T_\epsilon | X_0)}$  随  $\lambda$  增大而减少, 意味着当种群规模增大时, 平均首达时间将减少. 造成这一现象的原因是当种群规模  $\lambda$  变大时, 对搜索空间的采样将更频繁重复, 于是能够更容易地找到最优解. 从表 1 还可看出理论的时间上界与实际实验结果相吻合.

图 3 展示了  $\lambda$  分别为 3、5、15、25、50 的  $(1, \lambda)$ ES 的收敛过程, 可以看出  $(1, \lambda)$ ES 收敛速度很快, 这与表 1 的结果是吻合的.

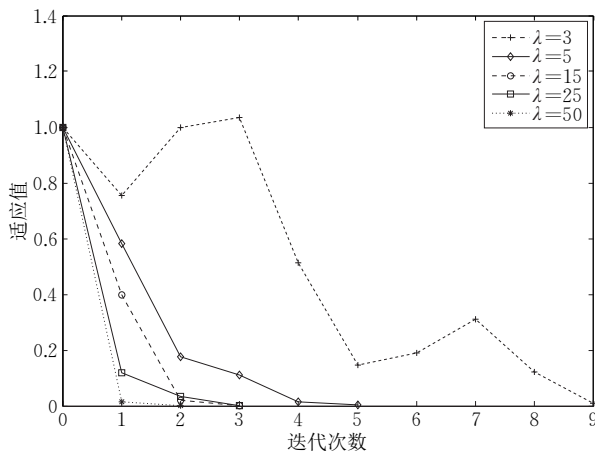


图 3  $(1, \lambda)$ ES 的收敛过程

文献[26]运用瓦尔德等式分析了  $(1+\lambda)$ ES 求解球函数问题的计算时间, 过程冗长而复杂, 而且只给出了渐进的结果. 相比之下, 我们的方法更加简便

直观, 而且可得到较为紧致的界.

## 5.2 $(1+\lambda)$ ES 求解倾斜平面问题

实验设置: 固定误差  $\epsilon=0.01$ ; 置  $\xi_0 = (x_1^0, x_2^0) = (0, 0)$ ,  $a=10$ , 此时  $X_0=10$ ; 对每一个  $\lambda$ , 将算法 3 在 2 维倾斜平面上运行 300 轮. 表 2 对实际的平均首达时间  $\widehat{E(T_\epsilon | X_0)}$  与理论的时间上界  $1 + \frac{2(\lambda+1)(a-\epsilon)}{\lambda-1 + \left(\frac{1}{2}\right)^\lambda}$  作了比较.

表 2  $(1+\lambda)$ ES 平均首达时间和理论上限的比较

$\lambda$	$\widehat{E(T_\epsilon   X_0)}$	$1 + \frac{2(\lambda+1)(a-\epsilon)}{\lambda-1 + \left(\frac{1}{2}\right)^\lambda}$
1	80.37	81.00
3	36.54	38.65
5	28.86	30.77
10	23.78	25.44
20	22.23	23.10
30	22.13	22.38
40	21.89	22.02

从表 2 可以看出, 实际运行结果与理论上界非常接近, 这说明我们得到的理论上界很紧致. 此外, 随着  $\lambda$  的增大, 实际平均首达时间和理论上界都趋于定值, 即

$$\lim_{\lambda \rightarrow \infty} 1 + \frac{2(\lambda+1)(a-\epsilon)}{\lambda-1 + \left(\frac{1}{2}\right)^\lambda} = 1 + 2(a-\epsilon) \approx 21.$$

文献[28]引入更新过程为  $(1+1)$ ES 建模, 分析了  $(1+1)$ ES 在 2 维倾斜平面问题上的平均首达时间, 得到了较为紧致的闭合表达式. 但是更新过程需要对算法的每一步位移作独立同分布的假设, 这个要求比较高, 限制了该模型的适用范围, 而我们的方法无需该假设, 而且我们还将  $(1+1)$ ES 推广到  $(1+\lambda)$ ES, 得到了更一般的结果.

## 6 结 论

本文在文献[31]的基础上进一步深化, 提出的平均增益模型是建立在非负随机过程  $\{X_t; t=0, 1, 2, \dots\}$  之上, 从而使得模型与具体算法和问题脱离开, 从理论高度上进一步发展了文献[31]的模型. 通过将鞅论与停时理论相结合, 我们建立了一个通用的公式(式(2))以估计连续型演化算法的平均首达时间上界. 所提模型的核心是定理 2, 在定理中, 平均增益  $\delta_t$  的下界  $h(X_t)$  依赖于  $X_t$ , 故无需找到  $\delta_t$  在所有  $t=0, 1, 2, \dots$  上的一致下界, 这有助于减少

计算量并通过积分获得首达时间  $T_e$  的较为紧致的上界。

当将该模型应用于连续型演化算法的计算时间分析时, 由于大多数演化算法可以用马尔可夫链建模, 此时平均增益  $\delta_i = E(X_i - X_{i+1} | \mathcal{H}_i)$  可以简化为  $\delta_i = E(X_i - X_{i+1} | X_i)$ , 这会使得计算相对更容易些。

在实例分析中, 我们分析了: (1) 带自适应步长的非精英  $(1, \lambda)$  ES 在 3 维球函数问题上的计算时间, 并得到了平均首达时间上界的解析表达式. 该分析还表明假如步长  $l_i$  和子代种群规模  $\lambda$  满足特定的条件, 则非精英  $(1, \lambda)$  ES 将收敛到全局最优; (2) 带均匀分布变异算子的  $(1 + \lambda)$  ES 在 2 维倾斜平面问题上的计算时间, 并得到了平均首达时间上界的解析表达式. 两个案例分析都表明  $\lambda$  越大, 平均首达时间就越小. 数值实验显示理论分析与实际情况相吻合. 遵循类似的方法, 我们的理论结果也可推广到  $n$  维的场合, 当然这需要更为复杂精细的计算。

值得注意的是, 本文所提模型实际上不仅仅适用于连续型演化算法, 也适用于离散型演化算法, 这可从推导过程看出. 目前, 我们正在利用统计学的方法设计实验, 结合平均增益模型, 从理论和实验数据两方面来分析具有实际背景的演化算法的计算时间, 使得目前的工作可以对算法的实际应用起到指导作用。

## 参 考 文 献

- [1] Oliveto P S, He J, Yao X. Time complexity of evolutionary algorithms for combinatorial optimization: A decade of results. *International Journal of Automation and Computing*, 2007, 4(3): 281-293
- [2] Yao X. Unpacking and understanding evolutionary algorithms // *Proceedings of the WCCI 2012*. Berlin, Germany, 2012: 60-76
- [3] Doerr B, Johannsen D, Winzen C. Multiplicative drift analysis. *Algorithmica*, 2012, 64(4): 673-697
- [4] Drosste S, Jansen T, Wegener I. On the analysis of the  $(1+1)$  evolutionary algorithm. *Theoretical Computer Science*, 2002, 276(1-2): 51-81
- [5] Wegener I. Methods for the analysis of evolutionary algorithms on pseudo-Boolean functions // Sarker R, Mohammadian M, Yao X eds. *Evolutionary Optimization*. Norwell, USA: Kluwer Academic Publishers, 2002: 349-369
- [6] He J, Yao X. Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence*, 2001, 127(1): 57-85
- [7] Oliveto P S, He J, Yao X. Analysis of the  $(1+1)$  EA for finding approximate solutions to vertex cover problems. *IEEE Transactions on Evolutionary Computation*, 2009, 13(5): 1006-1029
- [8] Lehre P K, Yao X. Runtime analysis of the  $(1+1)$  EA on computing unique input output sequences. *Information Sciences*, 2014, 259(1): 510-531
- [9] Lai X S, Zhou Y R, He J, et al. Performance analysis of evolutionary algorithms for the minimum label spanning tree problem. *IEEE Transactions on Evolutionary Computation*, 2014, 18(6): 860-872
- [10] Zhou Y R, Zhang J, Wang Y. Performance analysis of the  $(1+1)$  evolutionary algorithm for the multiprocessor scheduling problem. *Algorithmica*, 2015, 73(1): 21-41
- [11] Zhou Y R, Lai X S, Li K. Approximation and parameterized runtime analysis of evolutionary algorithms for the maximum cut problem. *IEEE Transactions on Cybernetics*, 2015, 45(8): 1491-1498
- [12] Xia X, Zhou Y R, Lai X S. On the analysis of the  $(1+1)$  evolutionary algorithm for the maximum leaf spanning tree problem. *International Journal of Computer Mathematics*, 2015, 92(10): 2023-2035
- [13] He J, Yao X. Towards an analytic framework for analysing the computation time of evolutionary algorithms. *Artificial Intelligence*, 2003, 145(1-2): 59-97
- [14] Yu Y, Zhou Z H. A new approach to estimating the expected first hitting time of evolutionary algorithms. *Artificial Intelligence*, 2008, 172(15): 1809-1832
- [15] Yu Y, Qian C, Zhou Z H. Switch analysis for running time analysis of evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 2015, 19(6): 777-792
- [16] Sudholt D. A new method for lower bounds on the running time of evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 2013, 17(3): 418-435
- [17] Witt C. Fitness levels with tail bounds for the analysis of randomized search heuristics. *Information Processing Letters*, 2014, 114(1-2): 38-41
- [18] Jägersküpfer J. Combining Markov-chain analysis and drift analysis: The  $(1+1)$  evolutionary algorithm on linear functions reloaded. *Algorithmica*, 2011, 59(3): 409-424
- [19] Chen T S, He J, Sun G, et al. 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)*, 2009, 39(5): 1092-1106
- [20] Oliveto P S, Witt C. Simplified drift analysis for proving lower bounds in evolutionary computation. *Algorithmica*, 2011, 59(3): 369-386
- [21] Rowe J, Sudholt D. The choice of the offspring population size in the  $(1, \lambda)$  EA // *Proceedings of the GECCO'12*. New York, USA, 2012: 1349-1356

- [22] Witt C. Tight bounds on the optimization time of a randomized search heuristic on linear functions. *Combinatorics Probability and Computing*, 2013, 22(2): 294-318
- [23] Droste S. Analysis of the (1+1)EA for a noisy OneMax// *Proceedings of the 6th ACM Conference on Genetic and Evolutionary Computation*. Seattle, USA, 2004: 1088-1099
- [24] Qian C, Yu Y, Zhou Z H. Analyzing evolutionary optimization in noisy environments. *Evolutionary Computation*, 2015, doi: 10.1162/EVCO\_a\_00170
- [25] Qian C, Yu Y, Jin Y, et al. On the effectiveness of sampling for evolutionary optimization in noisy environments. *Evolutionary Computation*, 2016, 3(1): 35-55
- [26] Jägersküpper J. Algorithmic analysis of a basic evolutionary algorithm for continuous optimization. *Theoretical Computing Science*, 2007, 379(3): 329-347
- [27] Jägersküpper J. Lower bounds for randomized direct search with isotropic sampling. *Operations Research Letters*, 2008, 36(3): 327-332
- [28] Agapie A, Agapie M, Baganu G. Evolutionary algorithms for continuous-space optimisation. *International Journal of Systems Science*, 2013, 44(3): 502-512
- [29] Zhang Yu-Shan, Hao Zhi-Feng, Huang Han, et al. A stopping time theory model of first hitting time analysis of evolutionary algorithms. *Chinese Journal of Computers*, 2015, 38(8): 1582-1591(in Chinese)
- (张宇山, 郝志峰, 黄翰等. 进化算法首达时间分析的停时理论模型. *计算机学报*, 2015, 38(8): 1582-1591)
- [30] Chen Y, Zou X F, He J. Drift conditions for estimating the first hitting times of evolutionary algorithms. *International Journal of Computer Mathematics*, 2011, 88(1): 37-50
- [31] Huang Han, Xu Wei-Di, Zhang Yu-Shan, et al. Runtime analysis for continuous (1+1) evolutionary algorithm based on average gain model. *Scientia Sinica Informationis*, 2014, 44(6): 811-824(in Chinese)
- (黄翰, 徐威迪, 张宇山等. 基于平均增益模型的连续型(1+1)进化算法计算时间复杂性分析. *中国科学 F 辑: 信息科学*, 2014, 44(6): 811-824)
- [32] Zhang Bo, Zhang Jing-Xiao. *Applied Stochastic Processes*. Beijing: Tsinghua University Press, 2004(in Chinese)
- (张波, 张景肖. *应用随机过程*. 北京: 清华大学出版社, 2004)
- [33] Xu Zong-Ben, Nie Zan-Kan, Zhang Wen-Xiu. Almost sure convergence of genetic algorithms: A martingale approach. *Chinese Journal of Computers*, 2002, 25(8): 785-793(in Chinese)
- (徐宗本, 聂赞坎, 张文修. 遗传算法的几乎必然强收敛性——鞅方法. *计算机学报*, 2002, 25(8): 785-793)
- [34] Schumer M A, Steiglitz K. Adaptive step size random search. *IEEE Transactions on Automatic Control*, 1968, 13: 270-276



**ZHANG Yu-Shan**, born in 1986, Ph. D., associate professor. His main research interests include theoretical foundation of evolutionary algorithms, convergence and time complexity analysis of bio-inspired algorithms.

**HUANG Han**, born in 1980, Ph. D., professor, Ph. D. supervisor. His main research interests include theoretical foundation of evolutionary computation, optimization design

## Background

The runtime analysis of evolutionary algorithms(EAs) is an important topic in the research of theoretical foundation of evolutionary algorithms, which currently attracts many scholars' attention. Theoretical study is helpful to understand the mechanism of EAs more precisely, and guides us to design more efficient EAs in practice. In contrast to fruitful research results of EAs in practical applications, theoretical results on the runtime analysis are still few. The inherent randomness of EAs makes it hard to analyze the runtime. The

and application of evolutionary computation.

**HAO Zhi-Feng**, born in 1969, Ph. D., professor, Ph. D. supervisor. His main research interests include machine learning, design and analysis of algorithms, mathematical foundation of bio-inspired algorithms, combinatorial optimization and algebra.

**YANG Xiao-Wei**, born in 1969, Ph. D., professor, Ph. D. supervisor. His main research interests include machine learning, computational intelligence, data mining.

existing work on runtime analysis of EAs mainly concentrates on discrete optimization problems, time complexity analysis of EAs on continuous search space remains relatively few, which is unsatisfactory from a theoretical point of view.

In this paper, we introduce the stopping time and martingale theory to establish a general average gain model to estimate the upper bound for the expected first hitting time of continuous EAs. The proposed model is established on a non-negative stochastic process and does not need the assumption

of Markov property. Afterwards, we demonstrate how the proposed model can be applied to the first hitting time analysis of continuous EAs. In the end, as cases study, we analyze (1) the first hitting time of the non-elitist  $(1, \lambda)$  ES with adaptive step-size on Sphere function using the proposed approach, and derive a closed-form expression of the time upper bound for 3-dimensional case. We also discuss the relationship between the step size and the offspring size  $\lambda$  to ensure convergence; (2) the expected first hitting time of the  $(1+\lambda)$ ES on the 2-dimensional inclined plane, and derive a closed-form expression of the time upper bound. The experimental results show that the proposed model helps to obtain tight upper bound for the expected first hitting time of continuous EAs.

This work is supported by the National Natural Science Foundation of China (61876207), the Guangdong

Natural Science Funds for Distinguished Young Scholar (2014A030306050), the Ministry of Education-China Mobile Research Funds (MCM20160206), and the Guangdong High-Level Personnel of Special Support Program (2014TQ01X664). Supported by these projects, our research group conducts research around the mathematical foundations of bio-inspired algorithms in recent years, and has achieved some theoretical achievements on the convergence and time complexity analysis of some bio-inspired algorithms such as EAs, ACO, PSO. We have published some related research papers on authoritative journals at home and abroad such as *Science China*, *Chinese Journal of Computers*, *Journal of Computer Research and Development*, *IEEE Transactions on System, Man, and Cybernetics, Part B (Cybernetics)*, *Information Sciences*, *Cluster Computing*, *Neural Computing and Applications*, etc.

计算机学报