2001 年 2 月 计 算 数 学 第 23 卷第 1 期
Feb., 2001 MATHEMATICA NUMERICA SINICA Vol.23, No.1

演化策略的全局收敛性[*1)]

郭崇慧　唐焕文
(大连理工大学应用数学系，大连， 116024)

GLOBAL CONVERGENCE PROPERTIES OF EVOLUTION STRATEGIES

Guo Chonghui　Tang Huanwen
(*Department of Applied Mathematics, Dalian University of Technology, Dalian,* 116024)

Abstract

This paper describes evolution strategy procedures for real-valued function optimization for the purpose of analyzing its asymptotic convergence properties. Two convergence theorems, which show that evolution strategy asymptotically converges to a global minmize point with probability one, are given.

Keywords: evolutionary algorithms, evolution strategies, optimization, global convergence properties

关键词：进化算法，演化策略，最优化，全局收敛性

1. 引　　言

进化算法 (EA, Evolutionary Algorithms) 是近年来兴起的一类基于生物界的自然选择和自然遗传机制的计算方法，如遗传算法 (GA, Genetic Algorithms) 、演化策略 (ES, Evolution Strategies) 和进化规划 (EP, Evolutionary Programming) 等方法. 这类算法的主要优点在于其本质上的并行性、广泛的可应用性和算法的高度稳健性、简明性与全局优化性 [1,2]. 目前，进化算法已被广泛地应用于计算机科学、工程技术、管理科学和社会科学等领域.

进化优化 (EO, Evolutionary Optimization) 是目前进化计算研究和应用的重点. 大多数古典优化方法都是基于目标函数的梯度或更高阶的信息来产生迭代序列. 在通常条件下，迭代序列渐进收敛到局部最优解，收敛速度也比较快. 但是当目标函数受到随机扰动时，则古典方法显得无能为力，特别在实际工程问题中，往往需要求出全局最优解. 而进化算法由于不是基于目标函数的梯度或更高阶的信息来产生迭代序列，并且是一种全局优化技术，因

∗ 1999 年 9 月 10 日收到.
1) 国家自然科学基金资助项目.

此其作为可靠、有效的全局优化方法，已经日益显示出处理现实世界中的复杂优化问题的巨大优势和潜力 [3].

进化算法对于求解复杂优化问题所显示出的潜力越来越引起众多学科研究者的兴趣，但有关进化算法的理论研究成果目前还很少. 特别关于这类算法的数学基础，包括收敛性、收敛速度、计算复杂性、稳健性等，还缺乏普遍意义的工作. 因此，建立坚实的模拟进化类仿生算法的数学基础不仅是这一新兴学科发展的当务之急，而且也更是发展众多以智能信息处理为核心的高、新科学技术的关键所在 [4].

D.B.Fogel 对遗传算法 (GA) 和进化规划 (EP) 的收敛性作了分析, 利用随机过程中的马尔可夫链给出了收敛性证明 [5]. 由于进化算法在最优搜索过程中利用随机性和结构化的信息使最满足目标的决策获得最大的生存可能，是一种概率型的算法 (Probability Algorithms). 因此自然想到：能否利用概率论中的有关知识来证明算法的收敛性？正是在这种思想的指导下，文 [6] 在一定条件下证明了 $(\mu+\lambda)-ES$ 演化策略“依概率收敛”于优化问题的全局极小点. 本文在文 [6] 的基础上继续讨论 $(\mu+\lambda)-ES$ 演化策略的收敛性，在较“依概率收敛”更强的收敛意义下，即“几乎处处收敛”，得到了 $(\mu+\lambda)-ES$ 演化策略的收敛性结果.

2. 预备知识

考虑全局优化问题

$$\text{(P)} \qquad \min\{f(x): x\in\Omega\subset R^n\}, \qquad f:\Omega\subset R^n\longrightarrow R^1,$$

则 (P) 的多个可行解的一个集合可称之为一个种群 (Population)；种群中的每一个元素 (可行解) 可称之为一个个体 (Individual)；种群中个体的数目称之为此种群的规模 (Size). 所谓的 $(\mu+\lambda)-ES$ 演化策略是指在算法中父代种群 (μ 个个体) 将与子代种群 (λ 个个体) 一同参与竞争，以选择产生新一代种群.

现将求解问题 (P) 的 $(\mu+\lambda)-ES$ 演化策略实现步骤描述如下：

步 1. 初始化：在 Ω 中按均匀分布随机选取 μ 个点 $x_1(0), x_2(0), \ldots, x_\mu(0)$. 这样我们得到初始种群 $X(0)=\{x_1(0), x_2(0), \ldots, x_\mu(0)\}$, 置 $k:=0$.

步 2. 产生中间种群：执行 λ 步 $(\lambda\ge\mu)$.

2.1. 置 $i:=1$.

2.2. 以等概率从 $X(k)$ 中选取两个个体 $x_{i1}(k), x_{i2}(k)$.

2.3. 以杂交算子作用于 $x_{i1}(k), x_{i2}(k)$ 以产生中间个体 $\hat{x}_{\mu+i}(k)$，其中杂交算子为中间杂交 (Intermediate Crossover), 即：

$$\hat{x}_{\mu+i}(k)=(x_{i1}(k)+x_{i2}(k))/2.$$

2.4. 变异：令 $x_{\mu+i}(k)=\hat{x}_{\mu+i}(k)+\Delta x$, 其中 $\Delta x\sim N(0,\sigma)=(N(0,\sigma_1), N(0,\sigma_2), \cdots, N(0,\sigma_n))^T$, $N(0,\sigma_i)$ 表示均值为 0, 方差为 σ_i^2 的正态分布，且 Δx 的 n 个分量 $\Delta x^1, \Delta x^2, \ldots, \Delta x^n$ 之间相互独立.

2.5. 若 $x_{\mu+i}(k) \in \Omega$, 转 2.6；否则转 2.4.

2.6. 若 $i = \lambda$, 转 Step 3；否则置 $i := i + 1$, 转 2.2.

步 3. 选择：从 $\{x_1(k), x_2(k), \dots, x_{\mu+\lambda}(k)\}$ 中选取 μ 个函数值最小的个体组成新一代种群 $X(k+1) = \{x_1(k+1), x_2(k+1), \dots, x_\mu(k+1)\}$.

步 4. 终止检验：检验当前种群 $X(k+1)$ 是否产生满意解或已达到预设的进化时限，若满足则停止；否则置 $k := k + 1$, 转步 2.

为了对上述算法进行收敛性分析，引入概率论 [7] 中的有关知识.

定义 1. 设 $\xi_n (n = 1, 2, \dots)$ 为概率空间上定义的随机序列，若存在随机变数 ξ, 使对 $\forall \epsilon > 0$, 有

$$\lim_{n\to\infty} P\{|\xi_n - \xi| < \epsilon\} = 1, \tag{1}$$

则称随机序列 $\{\xi_n\}$ 依概率收敛于随机变数 ξ.

定义 2. 对上述随机序列及随机变数，若

$$P\{\lim_{n\to\infty} \xi_n = \xi\} = 1, \tag{2}$$

或者对 $\forall \epsilon > 0$, 有

$$P\{\bigcap_{n=1}^{\infty} \bigcup_{k\ge n} [|\xi_k - \xi| \ge \epsilon]\} = 0, \tag{3}$$

则称随机序列 $\{\xi_n\}$ 以概率为 1 收敛 (几乎处处收敛) 于随机变数 ξ.

引理 1. 波雷尔 – 坎特里 (Borel–Cantelli) 引理

设 $A_1, A_2, \dots$ 是概率空间上的一事件序列，令 $p_k = P\{A_k\}$. 若 $\sum_{k=1}^{\infty} p_k < \infty$, 则

$$P\{\bigcap_{n=1}^{\infty} \bigcup_{k\ge n} A_k\} = 0. \tag{4}$$

若 $\sum_{k=1}^{\infty} p_k = \infty$, 且各 A_k 相互独立，则

$$P\{\bigcap_{n=1}^{\infty} \bigcup_{k\ge n} A_k\} = 1. \tag{5}$$

3. 主要结果

假设 1. (i) 问题 (P) 的可行域 Ω 为 R^n 中的有界闭区域. (ii) 目标函数 $f(x)$ 是区域 Ω 上的连续函数.

由假设 1, 易知

$$S = \{x | arg \min_{x\in\Omega} f(x)\} \neq \emptyset.$$

任给 $\epsilon > 0$, 若记

$$D_0 = \{x \in \Omega || f(x) - f^*| < \epsilon\}; \qquad D_1 = \Omega \setminus D_0, \tag{6}$$

其中 $f^* = \min\{f(x) : x \in \Omega\}$, 则上述算法中产生的 μ 个点可以分成两种状态:

(1) 至少有一个点属于 D_0 的，记为状态 S_0 ;

(2)μ 个点均属于 D_1, 记为状态 S_1.

首先给出如下引理:

引理 2. 设问题 (P) 中的目标函数和可行域满足假设 1, 若记 $q_{ij}(i,j=0,1)$ 为当 $X(k)$ 为状态 S_i 时， $X(k+1)$ 为状态 S_j 的概率，则有:

(a) 对于任意状态为 S_0 的点集 $X(k)$, $q_{00}=1$;

(b) 对于任意状态为 S_1 的点集 $X(k)$, 存在常数 $c \in (0,1)$, 使得 $q_{11} \le c$.

证明. 根据算法的步骤可知，状态为 S_0 的种群不会进化到状态为 S_1 的种群，故 (a) 显然成立.

下面我们来证明 (b) 成立.

因为目标函数 $f(x)$ 和可行域 Ω 满足假设 1, 所以算法初始种群对应的水平集 $L_\alpha = \{x \in \Omega | f(x) \le \alpha\}$ 有界，其中

$$\alpha = \max_{1 \le i \le \mu} f(x_i(0)).$$

设 $M = \overline{Co(L_\alpha)}$ 为水平集 L_α 的闭凸包. 由算法初始点的选取可知，对于状态为 S_1 的 $X(k)$, 其中任意两个个体 $x_{i1}, x_{i2} \in X(k)$. 由 M 的定义知，杂交点 $(x_{i1}+x_{i2})/2 \in M$.

由于 $f(x)$ 在 Ω 上连续，设 x_0 为其一个最小值点，故存在 $r>0$, 使得当

$$||x - x_0||_\infty \le r, \tag{7}$$

时，有

$$|f(x) - f(x_0)| < \epsilon/2. \tag{8}$$

若记

$$Q_{x_0,r} = \{x \in \Omega |\, ||x - x_0||_\infty \le r\}, \tag{9}$$

则显然有

$$Q_{x_0,r} \subset D_0. \tag{10}$$

由于

$$\begin{aligned} P\{(x+\Delta x) \in Q_{x_0,r}\} &= \prod_{i=1}^{n} P\{|x^i + \Delta x^i - x_0^i| \le r\} \\ &= \prod_{i=1}^{n} P\{|x_0^i - x^i - r \le \Delta x^i \le x_0^i - x^i + r\}, \end{aligned} \tag{11}$$

其中 x^i, x_0^i 分别表示点 x 和 x_0 的第 i 个分量.

又由于

$$\Delta x^i \sim N(0, \sigma_i),$$

故

$$P\{(x+\Delta x)\in Q_{x_0,r}\}=\prod_{i=1}^{n}\int_{x_0^i-x^i-r}^{x_0^i-x^i+r}\frac{1}{\sqrt{2\pi}\sigma_i}e^{-\frac{y^2}{2\sigma_i^2}}dy. \tag{12}$$

若记

$$P_1(x)=P\{(x+\Delta x)\in Q_{x_0,r}\},\qquad x\in M, \tag{13}$$

则由式 (12), 有

$$0<P_1(x)<1,\qquad x\in M. \tag{14}$$

由于 Δx^i 为服从正态分布的连续型随机变量以及式 (12)-(13), 知 $P_1(x)$ 连续. 又由于 M 为有界闭集，故存在 $y_0\in M$, 使得

$$P_1(y_0)=\min_{x\in M}P_1(x),\qquad 0<P_1(y_0)<1. \tag{15}$$

由于 q_{10} 表示当 $X(k)$ 为状态 S_1 时， $X(k+1)$ 为状态 S_0 的概率，利用式 (10) 和式 (15) 可得

$$q_{10}\geq P_1(x)\geq P_1(y_0). \tag{16}$$

令

$$c=1-P_1(y_0). \tag{17}$$

由式 (16) 、 (17) 以及等式 $q_{11}+q_{10}=1$, 可得

$$q_{11}=1-q_{10}\leq 1-P_1(y_0)=c,\qquad (0<c<1).$$ ■

定理 1. 设 $\{X(m)\}$ 是由算法产生的种群序列，其中 $x^*(m)\in X(m)$ 为第 m 代种群中的最优点，即

$$x^*(m)=arg\min_{1\leq i\leq\mu}f(x_i(m)),$$

如果问题 (P) 中的目标函数和可行域满足假设 1, 则有

$$P\{\lim_{m\to\infty}f(x^*(m))=f^*\}=1 \tag{18}$$

成立，即种群序列以概率为 1 收敛于全局最优解.

证明. 对于 $\forall\epsilon>0$, 令 $p_k=P\{|f(x^*(k))-f^*|\geq\epsilon\}$, 则

$$p_k=\begin{cases}0 & 当\ \exists T\in\{0,1,2,\cdots,k\},\quad x^*(T)\in D_0\\ \bar{p}_k & 当\ x^*(m)\notin D_0, m=0,1,2,\cdots,k\end{cases}. \tag{19}$$

由引理 2 知

$$\bar{p}_k=P\{x^*(m)\notin D_0,m=0,1,2,\cdots,k\}=q_{11}^k\leq c^k, \tag{20}$$

故有

$$\sum_{k=1}^{\infty}p_k\leq\sum_{k=1}^{\infty}c^k=\frac{c}{1-c}<\infty. \tag{21}$$

由引理 1 知

$$P\{\bigcap_{n=1}^{\infty}\bigcup_{k\geq n}[|f(x^*(k))-f^*|\geq\epsilon]\}=0. \tag{22}$$

因此，由“以概率为 1 收敛”的定义可知 (见 (2) 式)，结论成立.

由定理 1 的证明过程可知，如下定理成立.

定理 2. 在算法中，如果问题 (P) 中的目标函数和可行域满足假设 1，遗传产生的点采用随机杂交算子，即在步 2.3 中令 $\hat{x}_{\mu+i}(k)$ 为 $x_{i1}(k)$ 和 $x_{i2}(k)$ 的凸组合中的点，而且随机产生的每一个 Δx^i 服从的概率密度分布函数 $\psi(x)$ 满足下列条件:

1)$\psi(x)$ 连续;

2)$\forall a,b\in R, a<b$, 有 $\int_a^b\psi(x)dx>0$.

则此算法产生的序列 $\{x^*(m)\}$ 满足

$$P\{\lim_{m\to\infty}f(x^*(m))=f^*\}=1. \tag{23}$$

由于以概率为 1 收敛蕴涵着依概率收敛，故本文是在较依概率收敛更强的收敛意义下 (即 $P\{\lim_{m\to\infty}f(x^*(m))=f^*\}=1$) 得到了 $(\mu+\lambda)-ES$ 演化策略的收敛性结果.

参 考 文 献

[1] 徐宗本，李国，解全局优化问题的仿生类算法（Ⅰ)- 模拟进化算法，运筹学杂志， 14:2 (1995), 1–13.
(Xu Zongben, Li Guo. The simulated-biology-like algorithms for global optimization, Part I: simulated evolutionary algorithms, *Chinese Journal of Operations Research,* 14:2(1995), 1–13.)

[2] 谢金星，进化计算简要综述，控制与决策， 12:1(1997),1–7.
(Xie Jinxing, A brief review on evolutionary computation, *Control and Decision,* 12:1(1997), 1–7.)

[3] David B. Fogel, An introduction to simulated evolutionary optimization, *IEEE Transactions On Neural Networks,* 5:1(1994), 3–14.

[4] 李国杰，计算智能：一个重要的研究方向，《智能计算机基础研究'94》，清华大学出版社，北京，1994.
(Li Guojie, Computational intelligence: an important research direction, Basic research on intelligent computer '94, Tsinghua University Press, Beijing, 1994.)

[5] David B.Fogel, Asymptotic convergence properties of genetic algorithms and evolutionary programming: analysis and experiments, Cybernetics and Systems: *An International Journal,* 25:3(1994), 389–407.

[6] 李宏，唐焕文，郭崇慧，一类进化策略的收敛性分析，运筹学学报， 3:4(1999),79–83.
(Li Hong, Tang Huanwen and Guo Chonghui, The convergence of a class of evolution strategies, *OR Transactions,* 3:4(1999), 79–83.)

[7] 梁之舜，邓集贤等，概率论与数理统计，高等教育出版社，北京， 1988.
(Liang Zhishun, Deng Jixian et al, The theory of probability and mathematical statistics, Publisher of Higher Education, Beijing, 1988.)