An Initial Analysis of Approximation Error for Evolutionary Algorithms

Jun He [1]   Yuren Zhou [2]   Guangming Li [3]

[1]Aberystwyth University, UK [2]Sun Yat-sen University, China [3]Shenzhen Institute of Information Technology, China

July 26, 2017

Motivation

This work aims at rigorously analyzing the approximation error of evolutionary algorithms (EAs).

Background

Consider an EA for solving a maximization problem:

$$\max f(x), \text{ subject to } x \in \mathcal{S}. \tag{1}$$

Let $f_{\max}$ denote the fitness of the optimal solution and F_t the expected fitness of the best solution found in the tth generation.

Definitions

- The **approximation error** of the EA in the tth generation is

$$E_t := f_{\max} - F_t. \tag{2}$$

- If some positive constants α and β exist with

$$\lim_{t \to +\infty} \frac{E_t}{(E_{t-1})^{\alpha}} = \beta, \tag{3}$$

then $\{E_t; t = 0, 1, \cdots\}$ is called to **converge to 0 in the order** α, with **asymptotic error constant** β [1, 2].

Research Questions

1. Order $\alpha =$?
2. Asymptotic error constant $\beta =$?

An experimental study

EA-I: $(1+1)$ EA using onebit mutation and elitist selection

EA-II: $(1+1)$ EA using bitwise mutation and elitist selection

$f(x)$: OneMax function

α: set to 1

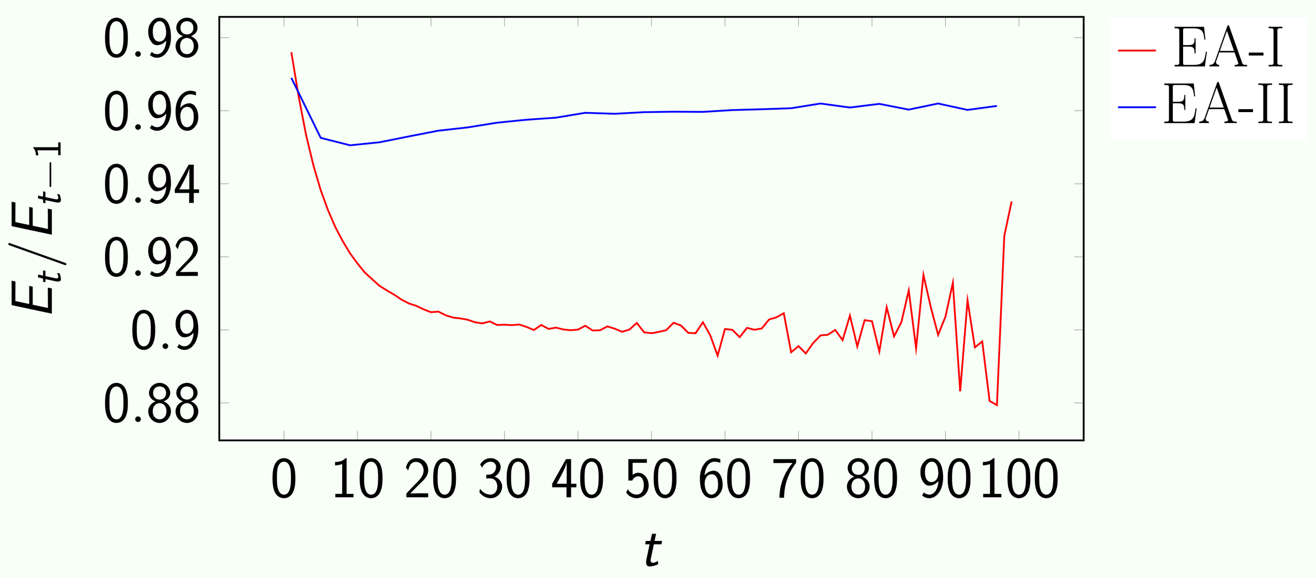

Figure: For EA-I and II, E_t/E_{t-1} converge to some β but stochastic disturbance exists on EA-II.

Analysis Tool

The analysis tool is Markov chain theory [3, 4].

Label all populations by indexes $\{0, 1, \cdots, L\}$ where indexes are sorted according to the fitness value of populations from high to low:

$$f_{\max} = f_0 > f_1 \geq \cdots \geq f_L = f_{\min}, \tag{4}$$

where f_i denotes the fitness of the best individual in the i-th population.

- $r_{i,j}$ denotes the transition probability of an EA from j to i.
- Matrix $\mathbf{R}$ denotes transition probabilities within the set $\{1, \cdots, L\}$.

$$\mathbf{R} := \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,L-1} & r_{1,L} \\ r_{2,1} & r_{2,2} & r_{2,3} & \cdots & r_{2,L-1} & r_{2,L} \\ r_{3,1} & r_{3,2} & r_{3,3} & \cdots & r_{3,L-1} & r_{3,L} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ r_{L,1} & r_{L,2} & r_{L,3} & \cdots & r_{L,L-1} & r_{L,L} \end{pmatrix}. \tag{5}$$

- Vector $\mathbf{q}_0 := (\Pr(1), \Pr(2), \cdots, \Pr(L))^T$ represent the probability distribution of the initial population over the set $\{1, \cdots, L\}$.

Suppose that EAs can be modelled by homogeneous Markov chains and are convergent (approximation error $E_t \to 0$ when $t \to +\infty$).

Main Theoretical Result

1. In many cases, the order of convergence $\alpha = 1$
2. Asymptotic error constant equals to the spectral radius: $\beta = \rho(\mathbf{R})$.

General EAs

Under the particular initialization, that is, set $\mathbf{q}_0 = \mathbf{v}/|\mathbf{v}|$ where $\mathbf{v}$ is an eigenvector corresponding to the eigenvalue $\rho(\mathbf{R})$ [4].

Theorem 1

Let $\mathbf{R}$ be the transition submatrix with $\rho(\mathbf{R}) < 1$. Under particular initialization , it holds for all $t \geq 1$,

$$\frac{E_t}{E_{t-1}} = \rho(\mathbf{R}). \tag{6}$$

That is $\alpha = 1$ and $\beta = \rho(\mathbf{R})$.

EAs with Primitive Transition Matrices

Case 1: transition matrix $\mathbf{R}$ is primitive.

Primitive matrix

A matrix $\mathbf{R}$ is called primitive if there exists a positive integer m such that $\mathbf{R}^m > \mathbf{O}$.

This condition means that starting for any state i, an EA can reach any other state j in m generations.

Under random initialization, that is, the initial population can be chosen to be any non-optimal state with a positive probability. Equivalently, $\mathbf{q}_0 > \mathbf{0}$.

Theorem 2

If $\mathbf{R}$ is primitive, then under random initialization, it holds

$$\lim_{t \to +\infty} \frac{E_t}{E_{t-1}} = \rho(\mathbf{R}). \tag{7}$$

That is $\alpha = 1$ and $\beta = \rho(\mathbf{R})$.

EAs with Reducible Transition Matrices

Case 2: transition matrix $\mathbf{R}$ is reducible.

Definition

$\mathbf{R}$ is reducible if it can be split as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{O} & \mathbf{R}_{22} \end{pmatrix} \tag{8}$$

where $\mathbf{O}$ is a zero-value submatrix.

Consider a special reducible transition matrix $\mathbf{R}$ that is an upper triangular matrix:

$$\mathbf{R} = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,L-1} & r_{1,L} \\ 0 & r_{2,2} & r_{2,3} & \cdots & r_{2,L-1} & r_{2,L} \\ 0 & 0 & r_{3,3} & \cdots & r_{3,L-1} & r_{3,L} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & 0 & r_{L,L} \end{pmatrix}. \tag{9}$$

Theorem 3

If $\mathbf{R}$ is upper triangular with unique diagonal entry $r_{i,i}$, then under random initialization, it holds

$$\lim_{t \to +\infty} \frac{E_t}{E_{t-1}} = \rho(\mathbf{R}). \tag{10}$$

That is $\alpha = 1$ and $\beta = \rho(\mathbf{R})$.

References

[1] R. Burden, J. Faires, and A. Burden, *Numerical Analysis*. Cengage Learning, 2015.

[2] W. Gautschi, *Numerical Analysis*. Birkhäuser Boston, 2011.

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

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

Contact: Jun He, Department of Computer Science, Aberystwyth University, Aberystwyth SY23 3DB, UK. Email: jun.he@aber.ac.uk