On the Convergence of a Factorized Distribution Algorithm with Truncation Selection

Qingfu Zhang *
Department of Computer,
University of Essex,
CO4 3SQ, Colchester, UK.
E-mail: qzhang@essex.ac.uk

Non-Technical Abstract

Optimization is to find the "best" solution to a problem where the quality of a solution can be measured by a given criterion. Estimation of Distribution Algorithms (EDA) generate a sequence of populations (collections) of candidate solutions for solving an optimization problem. They evaluate and analyze each member in the current population, and based on the results estimate the probability of each possible solution being the best. Then a population of new candidate solutions are generated according to these estimated probabilities. EDAs have become an increasing popular optimization tool. However, it is still unclear in theory when EDAs work. In this paper, we prove that a Fractorized Distribution Algorithm (FDA) with truncation selection, an instance of EDAs, can solve a class of optimization problems.

*This work was supported by EPSRC under Grant GR/R64742/01

Abstract

We investigate the global convergence of a Fractorized Distribution Algorithm (FDA) with truncation selection. Like conventional genetic algorithms, FDAs maintain and successively improve a population of solutions. In FDAs, a distribution model is built based on the statistical information extracted from a set of selected solutions in the current population, and then the model thus built is used to generate new solutions for the next generation. The variable-dependence structure of the distribution model in FDAs is determined by the variable-interaction structure of the objective function. We prove that FDA with truncation selection converges globally for optimization of a class of additively decomposable functions (ADF). Our results imply that the utilization of appropriately selected dependence relationships is sufficient to guarantee the global convergence of estimation of distribution algorithms (EDAs) for optimization of ADFs.

Keywords: Genetic Algorithms, Estimation of Distribution Algorithms, Factorized Distribution Algorithms, Convergence, Infinite Population.

1 Introduction

Recently, Estimation of Distribution Algorithms (EDAs) [2] have emerged as an alternative to Genetic Algorithms (GAs) for solving hard search and optimization problems. Like classic genetic algorithms, EDAs maintain and improve a population of candidate solutions. However, there is no crossover or mutation in EDAs. Instead, they explicitly extract global statistical information from the current population and build a posterior probability distribution model of promising solutions, based on the extracted information. New solutions are sampled from the model thus built and fully or in part replace the current population to form the next generation. The probability models used in the existing EDAs can be divided into two classes:

- the models in which all the variables are statistically independent: Population-Based Incremental Learning (PBIL) [3], Univariate Marginal Distribution Algorithm (UMDA) [1] and Compact Genetic Algorithms (CGA) [13] use the models of this type.
- the models in which some selected variable independencies are considered: Mutual Information Maximization for Input Clustering (MIMIC)

[4], Combining Optimizers with Mutual Information Trees (COMIT) [5], Estimation of Bayesian Networks Algorithm (EBNA) [11], Factorized Distribution Algorithm (FDA) [6], Bayesian Optimization Algorithm (BOA) [7], Bayesian Evolutionary Algorithm (BEA) [8], to name a few, use the models of this type.

Muehlenbein [1] Gonzalez, et al [9], and Hoehfeld & Rudolph [10] have studied the behaviors of UMDA and PBIL. Their results implies that EDAs cannot solve problems with nonlinear variable interactions if they use probability models in which all the variables are independent. In [12], Muehlenbein & Mahnig discussed the behaviours of FDA for separable additively decomposable functions (ADFs). Since there are no overlaps in their objective functions, their FDA is equivalent to UMDA. Therefore, more work needs to be performed to study the working mechanics of EDAs using models with variable independencies. Our recent theoretical work on the convergence of FDA for general ADFs under proportional selection presents a first attempt to this goal [14]. However, FDA under other selection schemes and other EDAs using probability models with variable independencies have not yet been studied theoretically.

The purpose of this paper is to study the convergence of FDA under the widely-used truncation selection. Based on the knowledge of the variable-interaction structure of the objective function, FDA chooses a variable-dependence structure for its probability model. Our analysis in this paper shows that this approach is approximate for optimizing ADFs with overlaps. It is extremely difficult to analyze the behaviors of FDA with finite populations. Since the distributions of infinite populations can be studied mathematically and can be regarded as the limit distribution of large populations. this paper focuses on the dynamics of the distributions of infinite populations in FDA. All the results in this paper are on continuous optimization problems. However, they could be extended to discrete optimization problems without difficulty.

2 Factorized Distribution Algorithm

FDA is developed for solving the optimization of additively decomposable functions (ADFs). In FDA, the variable-dependence structure in the probability models is determined by the structure of the ADF to be optimized. In this paper, we consider the following optimization problem.

$$\max f(X) = f_1(X_1) + f_2(X_2) \qquad X \in D \tag{1}$$

where $X = (x_1, x_2, \cdots, x_n) \in R^n$, $X_1 = (x_1, \cdots, x_m)$ and $X_2 = (x_k, x_{k+1}, \cdots, x_n)$, $k \leq m + 1$. The feasible region $D = [a, b]^n$ is a supercube. We denote the overlap between X_1 and X_2 by $X_0 = (x_k, \cdots, x_m)$. $f_1 : [a, b]^m \to R$ and f_2 : $[a, b]^{n-k+1} \to R$ are positive and continuous functions. Therefore, there exists a point X^* such that $f(X) \leq f(X^*)$ for all $X \in D$. x^* is called a globally optimal solution and $G^* = f(x^*)$ is the global maximum[1].

2.1 The algorithm

FDA with truncation selection for the problem (1) can be described as follows:

Step 0 Intialization Set $t = 1$. Generate N points randomly from D to form the current population $Pop(t)$.

Step 1 Trunction Selection Compute the objective function values of all the points in the current population and rank all these points according to their objective function values. Select the $M = [\alpha N]$ $(0 < \alpha < 1)$ best points to form the parent set $Pop^s(t)$.

Step 2 Esimation of the Distribution Estimate the marginal probabilities of X_i $(i = 0, 1, 2)$: $p^s(X_0, t)$, $p^s(X_1, t)$, $p^s(X_2, t)$ of the points in $Pop^s(t)$.

Step 3 Sampling Sample N points from the probability

$$p(X, t + 1) = \frac{p^s(X_1, t) p^s(X_2, t)}{p^s(X_0, t)}$$

to form $Pop(t + 1)$.

Step 4 Set $t = t + 1$, go to Step 1.

In the above algorithm, the distribution of the promising solutions is modeled by $p(X, t+1)$. The structure of the variable dependence in $p(X, t+1)$ is the same as the structure of the variable interaction in the objective function $f(X)$. In a similar way, one can design FDA for general ADFs. Since FDA only needs to estimate the marginal distributions of some subvectors of X, the computational cost of FDA at each generation is mainly determined by the lengths of these subvectors. Thus FDA can be used for optimizing a large scale ADF if the lengths of the subvectors are bounded.

[1] $f(x)$ may have many distinct globally optimal solutions.

2.2 The model of infinite population

Let the underlying probability distribution functions for the points in $Pop(t)$ and $Pop^s(t)$ be $p(X,t)$ and $p^s(X,t)$, respectively. By the famous Glivenko-Canteli theorem, the empirical probability density functions induced by points in $Pop(t)$ and $Pop^s(t)$ will converge to $p(X,t)$ and $p^s(X,t)$, as the sizes of $Pop(t)$ and $Pop^s(t)$ tends to infinity, respectively. Therefore, $p(X,t)$ and $p^s(X,t)$ can be thought as the population and the parent set at time t in EDAs with infinite populations. FDA with infinite populations can be regarded as the following iteration of probability functions:

Step 1 Trunction Selection ($p(X,t) \rightarrow p^s(X,t)$)

$$p^s(X,t) = \begin{cases} \frac{p(X,t)}{\alpha(t)}, & \text{if} \quad f(X) \geq \beta(t), \\ 0, & \text{otherwise} \end{cases}$$

where $\beta(t)$ is a threshold such that $\alpha(t) = \int\limits_{f(X) \geq \beta(t)} p(X,t)dX$. The point is selected if and only if its objective function is greater than $\beta(t)$. As a result, the $100\alpha(t)\%$ best points become the parents for the next generation.

Step 2 Generating new population ($p^s(X,t) \rightarrow p(X,t+1)$)

$$p(X,t+1) = \frac{p^s(X_1,t)p^s(X_2,t)}{p^s(X_0,t)}$$

where $p^s(X_i,t)$ is the marginal probability of X_i in $p^s(X,t)$ $(i = 1,2,3)^2$.

3 Theoretical Analysis of The Convergence of FDA

We first study the case when $k = m+1$, i.e., there is no overlap between X_1 and X_2, then we consider the case when $k < m+1$.

3.1 The case of k=m+1

In this subsection, we always assume that

(A1.1) $k = m+1$, i.e., $X_0 = \emptyset$.

[2] if $X_0 = \emptyset$, $p^s(X_0,t) = 1$.

(A1.2) For any constant c, let $S_1 = \{X_1 | f_1(X_1) = c\}$ and $S_2 = \{X_2 | f_2(X_2) = c\}$, then $\mu_1(S_1) = 0$ and $\mu_2(S_2) = 0$, where μ_1 and μ_2 are the Borel measures in R^m and R^{n-m}, respectively;

(A1.3) $p(X, 0)$ is positive and continuous in D; and

(A1.1) simply means that there is no overlap between X_1 and X_2. (A1.2) means that f_1 and f_2 are not constant in any sets with positive Borel measure. Intuitively, (A1.3) requires that the initial population should be sampled from a positive and continuous probability model.

Lemma 1 *For FDA with infinite population defined in Section 2. B. If there is a positive ξ such that $\beta(t) < G^* - \xi$ for all $t \geq 0$. Then*

$$\lim_{t \to \infty} \alpha(t) = 1 \tag{2}$$

Proof: Without loss of generality, we assume that $\max_{X_1 \in [a,b]^m} f_1(X_1) = \max_{X_2 \in [a,b]^{n-m}} f_2(X_2) = \frac{G^*}{2}$ (if $f_1(X_1)$ and $f_2(X_2)$ do not satisfy this assumption, let $f_1(X_1) := f_1(X_1) + \frac{G^*}{2} - \max_{X_1 \in [a,b]^m} f_1(X_1)$ and $f_2(X_2) := f_2(X_2) + \frac{G^*}{2} - \max_{X_2 \in [a,b]^{n-m}} f_2(X_2)$, then we will have that $\max_{X_1 \in [a,b]^m} f_1(X_1) = \max_{X_1 \in [a,b]^{n-m}} f_2(X_2) = \frac{G^*}{2}$). By Assumption (A1.1), $X_0 = \emptyset$, then the iteration of the FDA can be expressed as

$$p^s(X,t) = \begin{cases} \frac{p(X,t)}{\alpha(t)}, & \text{if} \quad f(X) \geq \beta(t), \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

and

$$p(X, t+1) = p^s(X_1, t) p^s(X_2, t). \tag{4}$$

Integrating both sides of (3) over X_1 and by (4), we can derive

$$p(X_1, t+1) = \frac{p(X_1, t) g_1(X_1, t)}{\alpha(t)} \tag{5}$$

where

$$g_1(X_1, t) = \int\limits_{\beta(t) - f_1(X_1) \leq f_2(X_2)} p(X_2, t) dX_2. \tag{6}$$

By Assumption (A1.2), $g_1(X_1, t)$ is a nonnegative continuous function of X_1. Clearly. $g_1(X_1, t) = 0$ whenever $f_1(X_1) \leq \beta(t) - \frac{G^*}{2}$, and $g_1(X_1^1, t) \geq g_1(X_1^2, t)$ if $f_1(X_1^1) \geq f_1(X_1^2)$, where X_1^1 and X_1^2 are values of X_1.

Similarly, we have

$$p(X_2, t+1) = \frac{p(X_2, t) g_2(X_2, t)}{\alpha(t)} \tag{7}$$

where

$$g_2(X_2, t) = \int\limits_{\beta(t) - f_2(X_2) \le f_1(X_1)} p(X_1, t) dX_1. \tag{8}$$

Therefore

$$p(X_1, t+1) = p(X_1, 1) G_1(X_1, t) \tag{9}$$

and

$$p(X_2, t+1) = p(X_2, 1) G_2(X_2, t) \tag{10}$$

where $G_1(X_1, t) = \prod_{j=1}^{t} g_1(X_1, j)[\alpha(j)]^{-1}$ and $G_2(X_2, t) = \prod_{j=1}^{t} g_2(X_2, j)[\alpha(j)]^{-1}$. By the continuity and monotonicity of $g_1(X_1, t)$ and $g_2(X_2\ , t)$, $G_1(X_1, t)$ and $G_2(X_2, t)$ are nonnegative continuous functions, and

$$G_1(X_1^1, t) \ge G_1(X_1^2, t) \tag{11}$$

if $f_1(X_1^1) \ge f_1(X_1^2)$, where X_1^1 and X_1^2 are values of X_1. Similarly

$$G_2(X_2^1, t) \ge G_2(X_2^2, t) \tag{12}$$

if $f_2(X_2^1) \ge f_2(X_2^2)$, where X_2^1 and X_2^2 are values of X_2. Let $\beta = sup\{\beta(0), \beta(1), \ldots\}$. By the assumption in the lemma

$$\beta \le G^* - \xi. \tag{13}$$

Now we prove the following claims.

Claim 1: $p(X_1, t) > 0$ for all $t \ge 1$ and for all X_1 with $f_1(X_1) > \beta - \frac{G^*}{2}$, and $p(X_2, t) > 0$ for all $t \ge 1$ and for all X_2 with $f_2(X_2) > \beta - \frac{G^*}{2}$.

By (3) and (4), we have

$$p(X_1, t+1) = \frac{1}{\alpha(t)} \int\limits_{\beta(t) - f_1(X_1) \le f_2(X_2) \le \frac{G^*}{2}} p(X, t) dX_2 \tag{14}$$

and

$$p(X_2, t+1) = \frac{1}{\alpha(t)} \int\limits_{\beta(t) - f_2(X_2) \le f_1(X_1) \le \frac{G^*}{2}} p(X, t) dX_1. \tag{15}$$

By Assumption (A1.3) and the condition in the lemma, $p(X_1, 1) > 0$ for all X_1 with $f_1(X_1) > \beta - \frac{G^*}{2}$. Similarly, we can prove $p(X_2, 1) > 0$ for all X_2 with $f_2(X_2) > \beta - \frac{G^*}{2}$. Therefore

$$p(X, 1) = p(X_1, 1)p(X_2, 1) > 0 \tag{16}$$

whenever $f_1(X_1) > \beta - \frac{G^*}{2}$ and $f_2(X_2) > \beta - \frac{G^*}{2}$. Applying induction on (14), (15) and (16) will easily lead to Claim 1.

Claim 2 There exists $A > 0$ such that $G_1(X_1, t) < A$ and $G_2(X_2, t) < A$ for all $t \geq 0$, for all X_1 with $f_1(X_1) < \frac{\beta}{2}$ and for all X_2 with $f_2(X_2) < \frac{\beta}{2}$.

Note that

$$\begin{aligned} 1 &\geq \int\limits_{\frac{\beta}{2} < f_1(X_1)} p(X_1, t+1) dX_1 = \int\limits_{\frac{\beta}{2} < f_1(X_1)} p(X_1, 1) G_1(X_1, t) dX_1 \\ &\geq \min_{\frac{\beta}{2} < f_1(X_1)} G_1(X_1, t) \int\limits_{\frac{\beta}{2} < f_1(X_1)} p(X_1, 1) dX_1. \end{aligned}$$

Noting that $\mu_1(\{X_1 | \frac{\beta}{2} < f_1(X_1, t)\}) > 0$. By Claim 1, we have $\int\limits_{\frac{\beta}{2} < f_1(X_1)} p(X_1, 1) dX_1 > 0$. Let $M_1 = \frac{1}{\int\limits_{\frac{\beta}{2} < f_1(X_1)} p(X_1, 1) dX_1}$, then

$$\min_{\frac{\beta}{2} < f_1(X_1)} G_1(X_1, t) \leq M_1. \tag{17}$$

By the monotoneity of $G_1(X_1, t)$, we obtain

$$G_1(X_1, t) \leq M_1 \tag{18}$$

for all $t \geq 1$ and all X_1 with $f_1(X_1) < \frac{\beta}{2}$. Similarly, there exists $M_2 > 0$ such that

$$G_2(X_2, t) \leq M_2 \tag{19}$$

for all $t \geq 1$ all X_2 with $f_2(X_2) < \frac{\beta}{2}$. Therefore, Claim 2 holds.

Claim 3 For any $\varepsilon > 0$, there exist $T > 0$ and $\eta \in (\beta - \frac{G^*}{2}, \frac{\beta}{2})$ such that

$$\int\limits_{f_2(X_2) \leq \eta} p(X_2, t) dX_2 < \varepsilon \tag{20}$$

for all $t \geq T$.

Let $K = \max_{X_2 \in [a,b]^{n-m}} p(X_2, 1)$. Since $p(X_2, 1)$ is continuous in $[a, b]^{n-m}$ and $K < \infty$. By Claim 2 and (10), we have

$$p(X_2, t) = p(X_2, 1)G_2(X_2, t-1) \le KA \tag{21}$$

for all $t \ge 1$ and for all X_2 with $f_2(X_2) < \frac{\beta}{2}$. Let $\overline{\beta}(t) = \sup\{\beta(0), \beta(1), \cdots, \beta(t)\}$. By (8) and (10), $p(X_2, t) = 0$ whenever $f_2(X_2) < \overline{\beta}(t-1) - \frac{G^*}{2}$. Since $f_2(X_2)$ is continuous and $lim_{t\to\infty}\overline{\beta}(t) = \beta$, for any $\varepsilon_1 > 0$ there exist $T_1 > 0$ and $\eta_1 \in (\beta - \frac{G^*}{2}, \frac{\beta}{2})$ such that $\mu_2(\{X_2 | \overline{\beta}(t-1) - \frac{G^*}{2} \le f_2(X_2) \le \eta_1\}) < \varepsilon_1$ if $t \ge T_1$. Consequently,

$$\int\limits_{f_2(X_2) \le \eta_1} p(X_2, t)dX_2 \le KA\varepsilon_1 \tag{22}$$

if $t \ge T_1$. Therefore, Claim 3 is true.

Claim 4 For any $\varepsilon > 0$, there exist $T \ge 1$ and $\delta > 0$ such that $g_1(X_1, t) > 1 - \varepsilon$ if $t > T$ and $f_1(X_1) > \frac{G^*}{2} - \delta$.

By (6)

$$g_1(X_1, t) = 1 - \int\limits_{f_2(X_2) \le \beta(t) - f_1(X_1)} p(X_2, t)dX_2 \tag{23}$$

It immediately follows from Claim 3 that Claim 4 is true.

Claim 5 $\int\limits_{c \le f_2(X_2)} p(X_2, t)dX_2 \le \int\limits_{c \le f_2(X_2)} p(X_2, t+1)dX_2$ for all $t \ge 0$ and for $c \le \frac{G^*}{2}$.

Let $r = \min_{X_2 \in [a,b]^{n-m}} f_2(X_2)$. Since $f_2(X_2)$ is continuous and defined in a closed set of $[a, b]^{n-m}$, $-\infty < r < +\infty$. Clearly, if $c \le r$, $\int\limits_{c \le f_2(X_2)} p(X_2, t)dX_2 = 1$ for all $t \ge 1$, thus Claim 5 hold. Now we assume that $\frac{G^*}{2} \ge c > r$. By (6), we have

$$\begin{aligned} \int\limits_{r \le f_2(X_2) \le c} p(X_2, t+1)dX_2 &= \int\limits_{r \le f_2(X_2) \le c} p(X_2, t)\frac{g_2(X_2, t)}{\alpha(t)}dX_2 \\ &= \lambda_1 \int\limits_{r \le f_2(X_2) \le c} p(X_2, t)dX_2, \end{aligned} \tag{24}$$

where $\lambda_1 = \frac{g_2(X_2^1, t)}{\alpha(t)}$ and $r \le f_2(X_2^1) \le c$. Moreover

$$\int\limits_{c \le f_2(X_2)} p(X_2, t+1)dX_2 = \int\limits_{c \le f_2(X_2)} p(X_2, t)\frac{g_2(X_2, t)}{\alpha(t)}dX_2$$

$$= \lambda_2 \int\limits_{r \le f_2(X_2) \le c} p(X_2, t) dX_2, \qquad (25)$$

where $\lambda_2 = \frac{g_2(X_2^2, t)}{\alpha(t)}$ and $c \le f_2(X_2^2)$. By the monotoneity of $g_2(X_2^2, t)$.

$$0 \le \lambda_1 \le \lambda_2. \qquad (26)$$

Noting that

$$\begin{aligned} 1 &= \int\limits_{r \le f_2(X_2) \le c} p(X_2, t) dX_2 + \int\limits_{c \le f_2(X_2)} p(X_2, t) dX_2 = \\ & \int\limits_{r \le f_2(X_2) \le c} p(X_2, t+1) dX_2 + \int\limits_{c \le f_2(X_2)} p(X_2, t+1) dX_2, \end{aligned} \qquad (27)$$

we have $\lambda_2 \ge 1$. Therefore

$$\int\limits_{c \le f_2(X_2)} p(X_2, t+1) dX_2 \ge \int\limits_{r \le f_2(X_2) \le c} p(X_2, t) dX_2.$$

Thus, Claim 5 holds.

Now we prove the theorem by contradiction. Assume that $\lim_{t \to \infty} \alpha(t) = 1$ is not true. Since $\alpha(t) \le 1$ for all $t \ge 0$, there exists $\theta < 1$ and $0 < t_1 < t_2 < \cdots$ such that $\alpha(t_k) < \theta$ for $k = 0, 1, \cdots$. By claim 4, there exist $\overline{T} \ge 1$ and $\overline{\delta} > 0$ such that $g_1(X_1, t) > \sqrt{\theta}$ if $t > \overline{T}$ and $f(X_1) > \frac{G^*}{2} - \overline{\delta}$. Therefore

$$\begin{aligned} \int\limits_{f_1(X_1) > \frac{G^*}{2} - \overline{\delta}} p(X_1, t_k + 1) dX_1 &= \int\limits_{f_1(X_1) > \frac{G^*}{2} - \overline{\delta}} p(X_1, t_k) \frac{g_2(X_1, t_k)}{\alpha(t)} dX_1 \\ &\ge \frac{1}{\sqrt{\theta}} \int\limits_{f_1(X_1) > \frac{G^*}{2} - \overline{\delta}} p(X_1, t_k) dX_1 \end{aligned} \qquad (28)$$

for all t_k. By (28) and Claim 5, we obtain

$$\int\limits_{f_1(X_1) > \frac{G^*}{2} - \overline{\delta}} p(X_1, t_k + 1) dX_1 \ge (\frac{1}{\sqrt{\theta}})^k \int\limits_{f_1(X_1) > \frac{G^*}{2} - \overline{\delta}} p(X_1, 1) dX_1. \qquad (29)$$

Thus

$$\int\limits_{f_1(X_1) > \frac{G^*}{2} - \overline{\delta}} p(X_1, t_k + 1) dX_1 \to +\infty, \quad \text{as} \quad k \to +\infty$$

which contradicts that $p(X_1, t)$ is a probability density function. This completes the proof. This Lemma directly leads to

Theorem 2 *For FDA with infinite population defined in Section 2. B. If* $\alpha(t) \leq \theta < 1$ *and* $\beta(t+1) \geq \beta(t)$ *for all* $t \geq 0$, *then*

$$\lim_{t\to\infty} \beta(t) = G^* \tag{30}$$

This theorem implies that if the threshold $\beta(t)$ is not deceasing and the selection percentage is always smaller than θ, then as t tends to infinity, all the points in the parent set will tend to the global optimal points.

3.2 The case of $k < m + 1$

In this subsection, we assume that

(A2.1) $k < m + 1$, i.e., $X_0 \neq \emptyset$;

(A2.2) Denote $Y_1 = (x_1, \cdots, x_{k-1})$ and $Y_2 = (x_{m+1}, \cdots, x_n)$. For any given X_0 and any given constant c, let $S_1 = \{ Y_1 \in [a, b]^{k-1} \,| f_1(X_1) = c\}$ and $S_2 = \{Y_2 \in [a, b]^{n-m}| f_2(X_2) = c\}$, then $\mu_1(S_1) = 0$ and $\mu_2(S_2) = 0$, where μ_1 and μ_2 are the Borel measures in R^{k-1} and R^{n-m}, respectively; and

(A2.3) $p(X, 0)$ is positive and continuous in D.

Theorem 3 *For FDA with infinite population defined in Section 2. B. If* $\alpha(t) \leq \theta < 1$ *and* $\beta(t+1) \geq \beta(t)$ *for all* $t \geq 0$, *then*

$$\lim_{t\to\infty} \beta(t) = G^* \tag{31}$$

Proof: From the assumptions in the theorem, $lim_{t\to\infty}\beta(t)$ exists. Denote $G = lim_{t\to\infty}\beta(t)$. Now we prove this theorem by contradiction. Assume that $G < G^*$. By (A2.1), FDA can be expressed as

$$p^s(X, t) = \begin{cases} \frac{p(X,t)}{\alpha(t)}, & \text{if} \quad f(X) \geq \beta(t), \\ 0, & \text{otherwise} \end{cases} \tag{32}$$

where

$$\alpha(t) = \int_{f(X)\geq\beta(t)} p(X, t)dX$$

and

$$p(X, t + 1) = \frac{p^s(X_1, t)p^s(X_2, t)}{p^s(X_0, t)}. \tag{33}$$

Therefore

$$p(X_0, t+1) = \frac{1}{\alpha(t)} \int\limits_{f(X) \geq \beta(t)} p(X,t) dY_1 dY_2. \tag{34}$$

Denote $B(X_0) = \max_{Y_1 \in [a,b]^{k-1}, Y_2 \in [a,b]^{n-m}} f(X)$ By Assumption (A2.3), we have

$$p(X_0, t) > 0$$

for all $t \geq 0$ and for any X_0 with $B(X_0) > G$. By (32) and (33), we obtain

$$p^s(Y_1, Y_2|X_0, t) = \begin{cases} \frac{p(Y_1, Y_2|X_0,t)}{\int\limits_{f(X)\geq\beta(t)} p(Y_1,Y_2|X_0,t) dY_1 dY_2}, & \text{if} \quad f(X) \geq \beta(t), \\ 0, & \text{otherwise}, \end{cases} \tag{35}$$

and

$$p(Y_1, Y_2|X_0, t+1) = p(Y_1|X_0, t) p(Y_1|X_0, t) \tag{36}$$

for all $t > 0$ and X_0 with $B(X_0) > G$. For a fixed X_0 with $B(X_0) > G$. By assumptions (A2.2), (A2.3)) and assumption that $G < G^*$, (35) and (36) can be regarded as FDA in (Y_1, Y_2)-space satisfying all the conditions in Lemma 1. Therefore

$$\lim_{t\to\infty} \int\limits_{f(X)\geq\beta(t)} p(Y_1, Y_2|X_0, t) dX = 1. \tag{37}$$

Moreover

$$\int\limits_{f(X)\geq\beta(t)} p(Y_1, Y_2|X_0, t) dX = \frac{p(X_0, t+1)}{p(X_0, t)\alpha(t)} \tag{38}$$

Thus, we have

$$\lim_{t\to\infty} \frac{p(X_0, t+1)}{p(X_0, t)} > 1 \tag{39}$$

for all X_0 with $B(X_0) > G$. Therefore,

$$\lim_{t\to\infty} p(X_0, t) = \infty \tag{40}$$

if $B(X_0) > G$.

Note that the Borel measure of the set $\{X_0 | B(X_0) > G\}$ is positive, then

$$\lim_{t\to\infty} \int\limits_{B(X_0)>G} p(X_0, t) dX_0 = \infty$$

which contradicts the fact that $p(X_0, t)$ is a probability density function. This completes the proof of this theorem.

One of the most commonly asked questions about EDAs for optimization problems is what dependence relationships in the distribution of the points in the parent set should be utilized to generate new points for locating the global optimum. Obviously, there is no definite answer for general optimization problems. Our analysis on FDA with truncation selection in this section implies that the utilization of appropriately selected dependence relationships is sufficient to guarantee the convergence of a population-based algorithm for optimization of an ADF function, and other dependence relationships can be neglected.

4 Conclusion

EDAs is population-based optimization algorithms using estimation of distributions instead of the manipulation of strings as done in genetic and other evolutionary algorithms. Though many applications of EDAs have been reported, very few efforts have been made in the theoretical study of the behaviors of the EDAs. FDA is a special EDA in which the structure of the probability model is determined by the variable-interaction structure in the objective function. This paper proved the convergence of FDA with truncation selection. However, the convergence proof is only valid for an infinite populations. The study of the convergence of FDA with finite populations will be our future work.

ACKNOWLEDGMENT: The author thanks Dr. H. Muehlenbein, Mr. H. Li, the editor and reviewers for their valuable suggestion.

References

[1] Muehlenbein, H., *The Equation for Response to Selection and Its Use for Prediction,* Evolutionary Computation, vol. 5, pp. 303-346, 1998.

[2] Larranaga, P. and Lozano, J. A., *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Kluwer Academic Publishers, 2001.

[3] Baluja, S., *Population-Based Incremental Learning: A Method for Integrating Genetic Search Based function Optimization and Competitive Learning*, Technical Report. Carnegie Mellon University, 1994.

[4] De Bonet, Isbell, J.S. and Viola, P. , *MIMIC: Finding Optima By Estimating Probability Densities*, Advances in Neural Information Processing Systems, Vol. 9, 1997. pp. 424-431, The MIT Press, Cambridge.

[5] Baluja, S. AND Davies, *Fast Probabilistic Modelling for Combinatorial Optimization*, AAAI-98, 1998.

[6] Muehlenbein, H., Mahnig, T. and Rodriguez, A. O., *Schemata, Distributions, and Graphical Models in Evolutionary Optimization*, Journal of Heuristics, 5, pp. 215-247, 1999.

[7] Pelikan, M., Goldberg, D. E. and Cantu-Paz, E., *BOA: the Bayesian Optimization Algorithm*, Proceedings of the Genetic and Evolutionary Computation Conference, pp. 525-532, 1999

[8] Zhang B. T., *A Bayesian Framework for Evolutionary Computation,* Proceedings of the 1999 Congress on Evolutionary Computation, Vol.1, pp. 722-228., 1999

[9] González C. , Lozano J. A., Larrañaga, p., *Analyzing the PBIL algorithm by means of discrete dynamical systems Complex Systems.* Vol. 12, pp. 465-479, 2000.

[10] Hoehfeld, M and Rudolph. G., *Towards a Theory of Population-Based Incremental Learning*, Proceedings of the 4th IEEE Conference on Evolutionary Computation, pp. 1-5, 1997.

[11] Etxeberria, R. and Larrañaga P. , *Global Optimization with Bayesian networks,* Proceedings of the II Symposium on Artificial Intelligence, CIMAF'99, La Habana, Cuba, pp. 332-339, 1999.

[12] Muehlenbein, H., & Mahnig, Th,, *Convergence Theory and Application of the Factorized Distribution Algorithm*, Journal of Computing and Information Technology, 7, pp. 19-32, 1999

[13] Harik, G. R., Lobo, F. G. and Goldberg, D. E., *The compact genetic algorithm.* IEEE Trans. on Evol. Comp., vol. 3, pp. 287–297, 1999.

[14] Zhang, Q. and Muehlenbein, H., *On the Convergence of a Class of EDAs*, IEEE Trans on Evol. Comp., Accepted for publication, 2003.

[15] Muehlenbein, H., Mahnig, Th., *FDA – A scalable evolutionary algorithm for the optimization of additively decomposed functions*, Evolutionary Computation, vol.4, pp. 353-376, 1999.