A Note on the Population Based Incremental Learning with Infinite Population Size

R. Rastegar
Computer Engineering and IT Department
Amirkabir University of Technology, Tehran, Iran
rrastegar@ce.aut.ac.ir

M. R. Meybodi
Computer Engineering and IT Department
Amirkabir University of Technology, Tehran, Iran
meybodi@ce.aut.ac.ir

Abstract. In this paper, we study the dynamical properties of the population based incremental learning (PBIL) algorithm when it uses truncation, proportional, and Boltzmann selection schemas. The results show that if the population size tends to infinity, with any learning rate, the local optima of the function to be optimized are asymptotically stable fixed points of the PBIL.

1 Introduction

Genetic Algorithm (GA) is a class of optimization algorithms motivated from the theory of natural selection and genetic recombination. It tries to find better solutions by selection and recombination of promising solutions. It works well in wide verities of problem domains. The poor behavior of genetic algorithm in some problems, in which the designed operators of crossover and mutation do not guarantee that the building block hypothesis is preserved, has led to the development of other types of algorithms. The search for techniques to preserve building blocks leads to the emergence of the new class of algorithms called Probabilistic Model Building Genetic Algorithms (PMBGAs) also known as Estimation of Distribution Algorithms (EDAs). The principle concept in these new techniques is to prevent disruption of partial solutions contained in a chromosome by giving them high probability of being presented in the child chromosome. Building a probabilistic model to represent correlation between variables in the chromosome and using this model to generate next population can achieve it. The EDAs are classified to three classes based on the interdependencies between variables in chromosomes [8]. Instances of EDAs include Population-based Incremental Learning (PBIL) [1], Univariate Marginal Distribution Algorithm (UMDA) [10], Learning Automata-based Estimation of Distribution Algorithm (LAEDA) [14], Compact Genetic Algorithm (cGA) [6] for no dependency model, Mutual Information Maximization for Input Clustering (MIMIC) [3], Combining Optimizer with Mutual Information Trees (COMIT) [2] for bivariate dependencies model, and Factorized Distribution Algorithm (FDA) [11], Bayesian Optimization Algorithm (BOA) [13] for multiple dependencies model, to name a few.

The PBIL is one of the simplest EDAs that ignore all the variables interactions. The first version of this algorithm was introduced by Baluja [1] in 1994. Recently, there has been many increasing interest in the PBIL and many papers have been published in the literature. These papers are in two domains. In the first domain, applications of the PBIL in solving difficult problems are interested [9][16]. The others are papers, which discuss extensions of the PBIL in to continuous search spaces [15] or theoretical frameworks of the PBIL [5][4][7][19][10].

This paper focuses on the dynamical properties of the PBIL with infinite population size. We have analyzed the stable fix points of the PBIL with respect to three famous selection schemas; truncation, proportional, and Boltzmann selection schemas. The framework, which is used in our analysis, is base on González approach [4]. We prove that the local optima (absolute local optima) are asymptotically stable fixed points of the PBIL.

The rest of paper is organized as follows. Section 2 briefly presents the PBIL algorithm. Related works are described in section 3. Section 4 demonstrates our analyzing results. Finally, Section 5 concludes.

2 The Population-Based Incremental Learning

The combinatorial optimization problem considered in this paper can be described as follows: Given a finite search space $D=\{0,1\}^l$ and a digestive pseudo boolean function $f:D\rightarrow\Re_{>0}$, find $max\{f(\mathbf{x});\mathbf{x}\in D\}$. The algorithm considered here for solving of this optimization problem is the PBIL.

The PBIL is a combination of evolutionary optimization and hill climbing [1]. The goal of this algorithm is to create a real valued probability vector, $\mathbf{p}=(p_1,\ldots,p_m,\ldots,p_l)$, which, when sampled, reveals high quality solutions with high probability. Note p_m is the probability of obtaining 1 in variable m. initially; the values of the probability vector are set to 0.5. Sampling from this probability vector yields random solutions because the probability of generating a 1 or 0 is equal. As search progresses, the values in the probability vector gradually shift to represent high quality solutions. This is done as follows; at instance n, N chromosomes are generated based upon the probabilities specified in the probability vector $\mathbf{p}(n)$. Then based on a selection method schema (often truncation selection), M chromosomes $\mathbf{x}_{1:N},\ldots,\mathbf{x}_{M:N}$ are selected from the generated population. The probability vector is pushed towards the selected chromosomes. The distance, which the probability vector is pushed, depends upon the learning rate parameter

$0<\alpha\leq 1$. After the probability vector is updated, sampling from the updated probability vector produces a new population of chromosomes, and the cycle is continued. As the search progresses, entries in the probability vector move away from their initial settings of 0.5 towards either 0.0 or 1.0. Updating of the probability vector is similar to the weight update rule in supervised competitive learning networks [1]. Fig. 1 shows the pseudocode of the PBIL.

```
Initialize probability vector p(0)
Do
   Using p(n) obtain N chromosomes x_1,...,x_N
   Evaluate them with respect to f.
   Select M chromosomes x_1:N,...,x_M:N .
   Update p(n) as follows,
```

$$\mathbf{p}(n+1)=(1-\alpha)\mathbf{p}(n)+\alpha\frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_{i:N} \qquad (1)$$

```
   where 0<α≤1 is learning rate parameter.
Until p(n) converges
```

Fig. 1. The pseudocode of the PBIL

3 Related works

Up to now, some effort has been devoted to study the working mechanism of the PBIL. Höhfeld and Rudolf [7] have showed that if the expectation of probability vector in time n is denoted by $E\{\mathbf{p}(n)\}$, then for a linear function,

$$\lim_{n\to\infty} E\{\mathbf{p}(n)\}=x^* \qquad (2)$$

where x^* is the optimum point of D.

When α is 1, the PBIL is equivalent to the UMDA. Mühlenbein and Mahnig [10] and Zhang [19] have proved that the UMDA with infinite population size stops at local optima when proportional and 2-tournoment selection schemas are used respectively.

González et al [5] have showed that the PBIL could be modeled by means of markov chain. By this model, they have proved that for PBIL with $M=1$ and $N=2$ applied to the OneMax problem in two dimensions:

$$P(\lim_{n\to\infty}\mathbf{p}(n)=(a,b))\to 1 \qquad (3)$$

when α tends to 1, $\mathbf{p}(0)\to(a,b)$, and $(a,b)\in\{0,1\}^2$. It shows a strong dependence of the PBIL on the initial parameters.

In [4], González et al have modeled the PBIL as a discrete dynamical system when the learning rate tends to zero and the population size is finite. They have assumed that in each instance one chromosome is selected using truncation selection schema. By this modeling, they have proved that local optima are stable fixed points of the PBIL. In their analysis, the cases in which M is greater than one are ignored.

4 Analyzing the PBIL

In this section, we analyze the stable fix points of the PBIL. Our approach is strongly inspired by [4], in which the PBIL is modeled as a discrete dynamical system in such a way that the trajectory of the PBIL is related to the iterations of a deterministic discrete dynamical system. As we stated before, in [4], González assumed that one chromosome is selected for updating the probability vector, i.e. $M=1$, and her analysis is for the cases which α tends *to 0* and the truncation selection method is used. We use the methodology of [4][18] but we assume $M, N\to\infty$. In our analysis, we consider three famous selection schemas; truncation, proportional, and Boltzmann selection schemas. We also assume that the learning rate can be any value from $(0,1]$.

First, we model the PBIL as a stochastic sequence and then obtain a deterministic model for it.

According to (1), the PBIL can be seen as a sequence of probability vectors each of which given by the stochastic rule $\tau:[0,1]^l\to[0,1]^l$:

$$\tau(\mathbf{p}(n))=\mathbf{p}(n+1)=(1-\alpha)\mathbf{p}(n)+\alpha\frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_{i:N} \qquad (4)$$

We are interested in the trajectory follows by the iteration of τ, and its behavior.

$$\lim_{n\to\infty}\tau^n(\mathbf{p}(0)) \qquad (5)$$

Now, define the operator $\Gamma:[0,1]^l\to[0,1]^l$ as follows:

$$\Gamma(\mathbf{p}(n))=E\{\mathbf{p}(n+1)\mid\mathbf{p}(n)\}=(1-\alpha)\mathbf{p}(n)+\alpha E\{\frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_{i:N}\mid\mathbf{p}(n)\} \qquad (6)$$

The operator Γ is a deterministic operator that gives the expected value of the stochastic operator τ.

When the population size tends to infinity, by the law of large numbers, we conclude that τ converges in probability to Γ. Because τ corresponds to Γ and follows the iterations of deterministic operator Γ. therefore, we study the behavior of Γ instead τ. We rewrite (6) as follows,

$$\Gamma(\mathbf{p})=(1-\alpha)\mathbf{p}+\alpha\Gamma'(\mathbf{p}) \qquad (7)$$

where

$$\Gamma'(\mathbf{p})=E\{\frac{1}{M}\sum_{i=1}^{M}\mathbf{X}_{i:N}\mid\mathbf{p}\} \qquad (8)$$

Note that, with respect to $0<\alpha\leq 1$, the dynamical system $\Gamma'(\mathbf{p})$ has the same behavior as $\Gamma(\mathbf{p})$ in the points of search space [4]. In the following, some necessary lemmas needed are stated.

Lemma 1. $\Gamma'(\mathbf{p})$ can be computed as follows,

$$\Gamma'(\mathbf{p})=\lim_{M\to\infty}E\{\frac{1}{M}\sum_{i=1}^{M}\mathbf{X}_{i:N}\mid\mathbf{p}\}=\sum_{x\in D}\mathbf{x}P_M(\mathbf{X}=\mathbf{x}\mid\mathbf{p}) \qquad (9)$$

Where $P_M(\mathbf{X}=\mathbf{x}|\mathbf{p})$ denotes the probability which the chromosome **x** belongs to the set of selected chromosomes.

Proof. With respect to the infinite population size, the proof is trivial from the law of large numbers. **Q.E.D.**

According to (9), the calculation of $P_M(\mathbf{X}=\mathbf{x}|\mathbf{p})$ is needed to compute $\Gamma'(\mathbf{p})$. Three cases with respect to the three considered selection schemas are stated.

The proportional selection model: The proportional selection model is the most widely selection schema in evolutionary algorithms. When the PBIL uses finite population the probability of a chromosome being selected is proportional to its fitness. Therefore, as the population size tends to infinity this selection schema should be modeled as [20]:

$$P_M(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p}) = f(\mathbf{x})P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})/E\{f(\mathbf{X})\,|\,\mathbf{p}\} \qquad (10)$$

Thus by lemma (1), we have:

$$\Gamma'(\mathbf{p}) = \sum_{\mathbf{x}\in D}\{\mathbf{x}f(\mathbf{x})P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})/E\{f(\mathbf{X})\,|\,\mathbf{p}\}\} \qquad (11)$$

The truncation selection model: In the truncation selection schema, all chromosomes of the population are ranked according to their fitnesses and the best ones are selected as the parents of the next generation. In the truncation selection schema with threshold $\xi>0$ only the $100\xi\%$ best chromosomes are selected to become the parents for the next generation [10]. It can be modeled as [20]:

$$P_M(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p}) = \begin{cases} \dfrac{P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\xi} & f(\mathbf{x})\geq\beta \\ 0 & otherwise \end{cases} \qquad (12)$$

Where β is determined by

$$\xi = \sum_{f(\mathbf{x})\geq\beta} P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p}) \qquad (13)$$

Therefore, by the definition of ξ and lemma (1), we have

$$\Gamma'(\mathbf{p}) = \begin{cases} 0 & \mathbf{p}\in\{\mathbf{x}\,|\,f(\mathbf{x})<\beta\} \\ \dfrac{\sum_{f(\mathbf{x})\geq\beta}\mathbf{x}P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\sum_{f(\mathbf{x})\geq\beta}P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})} & \mathbf{p}\in\{\mathbf{x}\,|\,f(\mathbf{x})\geq\beta\} \end{cases} \qquad (14)$$

The Boltzmann selection model: The Boltzmann selection schema is another selection method used in evolutionary algorithms. When the population size tends to infinity, this selection method should be modeled as:

$$P_M(\mathbf{X}=\mathbf{x}|\mathbf{p}) = \exp\{\theta f(\mathbf{x})\}\times P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})/E\{\exp\{\theta f(\mathbf{X})\}\,|\,\mathbf{p}\} \qquad (15)$$

Where $\theta>0$ is the selection parameter.

Now, we have to calculate the probability of sampling a particular chromosome **x** by given **p**.

Lemma 2. By given a probability vector **p,** the probability of sampling an chromosome $\mathbf{x}=(x_1,\ldots,x_n)$ is

$$P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p}) = \prod_{i=1}^{l} p_i^{\,x_i}(1-p_i)^{(1-x_i)} \qquad (16)$$

Proof: the proof is trivial by the fact that all x_is are independent.

In the reminder of this section, we find the properties of $\Gamma'(\mathbf{p})$, that will give us some information about the behavior of the PBIL. At first, we define the local optima of the function to be optimized and then we state Lemma 3 and Theorem 1 where are very useful for analyzing the system.

Definition 1. Let $f{:}D\rightarrow\Re_{>0}$ be a positive real function. **x** is a local maximum with respect to the hamming distance, d_H, if

$$\forall\mathbf{x}'\in D \;\; where \;\; d_H(\mathbf{x},\mathbf{x}')=1\rightarrow f(\mathbf{x}')\leq f(\mathbf{x}) \qquad (17)$$

x is said to be an absolute local maximum, if the above inequality is strict. It is clear that if f is a digestive function, each local maximum is an absolute local maximum.

Lemma 3. The following equalities are true.

$$P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{x}')=0 \quad for\ all \;\; \mathbf{x}\neq\mathbf{x}' \qquad (18)$$

$$P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{x}')=1 \quad for \;\; \mathbf{x}=\mathbf{x}' \qquad (19)$$

$$\left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}} = \begin{cases} 1 & if \;\; x_m=1 \\ -1 & if \;\; x_m=0 \end{cases} \qquad (20)$$

$$\left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = 0,\ \forall\mathbf{x}'\ d_H(\mathbf{x},\mathbf{x}')\geq 2 \qquad (21)$$

$$\left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = \begin{cases} 1 & if\ d_H(\mathbf{x},\mathbf{x}')=1,\ x_m=1,\ x'_m=0 \\ -1 & if\ d_H(\mathbf{x},\mathbf{x}')=1,\ x_m=0,\ x'_m=1 \end{cases} \qquad (22)$$

$$\left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = 0 \;\; if\ d_H(\mathbf{x},\mathbf{x}')=1, x_m=x'_m \qquad (23)$$

where **x** and **x**′ belong to D.

Proof: the proof is simplicity trivial by looking lemma 2. **Q.E.D.**

Definition 2. For each, **x** which belongs to D**,** D1, D2 and D3 are three subsets of D that are defined as follows,

$$\forall\mathbf{x}': D1=\{\mathbf{x}\,|\,d_H(\mathbf{x},\mathbf{x}')=1\}$$
$$D2=\{\mathbf{x}\,|\,d_H(\mathbf{x},\mathbf{x}')\geq 2\},\ D3=\{\mathbf{x}\} \qquad (24)$$
$$where\ D1\cup D2\cup D3=D$$

Theorem 1. Assume **x**′ be a fixed point of a discrete dynamical system Γ′[17]:

1) If all eigenvalues of the Jacobean matrix of Γ′ have absolute values less than one, then **x**′ is an asymptotically stable fixed point of Γ′.

2) If some eigenvalues of the Jacobean matrix of Γ′ have absolute values greater than one, then **x′** is an unstable fixed point of Γ′.

Now, we are ready to state main lemmas and theorems about the dynamical properties of the PBIL. For each selection schema, at first we find the fixed points of the PBIL and then by using theorem 1 we discover the properties of these fixed points. The first selection schema to be considered is proportional selection.

Lemma 4. If proportional selection schema is used, all points of *D* are fixed points of Γ′.
Its proof is given in appendix.

Lemma 5. Let λ_i be i^{th} eigenvalue of $\partial_{\mathbf{p}}\Gamma'(\mathbf{x}')$, by using the proportional selection schema. λ_i is computed as follow,

$$\lambda_i = \frac{f(\mathbf{x}(i,\mathbf{x}'))}{f(\mathbf{x}')} \tag{25}$$

Where $\mathbf{x}' \in D$, and

$$\mathbf{x}(i,\mathbf{x}') = \mathbf{x},\ d_H(\mathbf{x}',\mathbf{x}) = 1,\ x_i \neq x_i' \tag{26}$$

Its proof is given in appendix.

Theorem 2. Let *f* be a positive real function on *D* and proportional selection schema is used by the PBIL. If **x′** is an absolute local maximum of *f* then **x′** is a stable fixed point. The other points of *D*, which aren't absolute maximum points of *f,* are unstable.
Its proof is given in appendix.

Another selection schema that can be used in the PBIL is the truncation selection schema. In the following, we consider the PBIL with truncation selection schema

Lemma 6. In each instance *n*, the points of *D* that have fitnesses equal or higher than *β* (defined before) are the fixed points of Γ′ if truncation selection schema is used.
Its proof is given in appendix.

Lemma 7. If truncation selection schema is used, the r^{th} element of the m^{th} column of $\partial_{\mathbf{p}}\Gamma'(\mathbf{x}')$ is computed as follows,

$$\left.\frac{\partial \Gamma_r'(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = \begin{cases} 1 & if\ r = m \text{ and } f(\mathbf{x}(m,\mathbf{x}')) \geq \beta \\ 0 & otherwise \end{cases} \tag{27}$$

Where $\mathbf{x}' \in D$, and

$$\mathbf{x}(i,\mathbf{x}') = \mathbf{x};\ d_H(\mathbf{x}',\mathbf{x}) = 1, x_i \neq x_i'$$

Its proof is given in appendix.

Theorem 3. Let *f* be a positive real function on *D* and truncation selection schema is used by the PBIL. If **x′** is an absolute local maximum of *f* whose neighborhoods have fitnesses less than *β,* then **x′** is a stable fixed point.
Its proof is given in appendix.

Remark. Because the parameter *β* changes when the PBIL works, based on theorem 3, we can conclude that the stability or unstability of points of D may change.

Finally, we consider the PBIL with Boltzmann selection schema.

Lemma 8. If Boltzmann selection schema is used, all the points of *D* are fixed points of Γ′.
Its proof is given in appendix.

Lemma 9. Let λ_i be i^{th} eigenvalue of $\partial_{\mathbf{p}}\Gamma'(\mathbf{x}')$, by using Boltzmann selection schema, λ_i is computed as follows,

$$\lambda_i = \frac{\exp(\theta\, f(\mathbf{x}(i,\mathbf{x}')))}{\exp(\theta\ f(\mathbf{x}'))} \tag{28}$$

Where $\mathbf{x}' \in D$, and

$$\mathbf{x}(i,\mathbf{x}') = \mathbf{x},\ d_H(\mathbf{x}',\mathbf{x}) = 1,\ x_i \neq x_i' \tag{29}$$

Its proof is given in appendix.

Theorem 4. Let *f* be a positive real function on *D* and Boltzmann selection schema is used by PBIL. If **x′** is an absolute local maximum of *f,* then **x′** is a stable fixed point. The other points of *D*, which aren't absolute maximum points of *f,* are unstable.
Its proof is given in appendix.

Because of theorems 2 and 4, we can conclude that the PBIL will never converge to a point of D, which is not a local maximum. The result of Theorem 3 is weaker than the results of theorems 2 and 4. Theorem 3 indicates that if the probability vector **p** is very close to one of local maximum points (with conditions of theorem 3) the PBIL converges to that local maximum. In other words, we cannot say anything about the other points based on theorem 3.

The theorems 2, 3, and 4 still leave two questions unanswered. 1) Does the PBIL algorithm have some other stable attractors in the probabilistic space $[0,1]^l$? 2) Is it possible that the PBIL does not converge to a point of $[0,1]^l$ which would be the case, for example, if the PBIL exhibits limit cycle behavior or chaotic behavior? At present we have no results concerning these questions.

5 Conclusions

We analyzed the dynamical properties of the PBIL algorithm with proportional, Boltzmann, and truncation selection schemas. Our approach was strongly inspired by González C. et al approach [4], in which the PBIL was modeled as a discrete dynamical system. We proved that, as the population size tends to infinity, the local optima (absolute local optima) are the asymptotically stable fixed points of the PBIL algorithm.

Appendix

Proof (Lemma 4). Assume **y** belong to *D* and **p** be equal to **y**. It is clear that the probability of sampling a chromosome different from **y** is zero, therefore by (11), (18) and (19) we have,

$$\Gamma'(\mathbf{y}) = \sum_{\mathbf{x}\in D}\{\mathbf{x}f(\mathbf{x})P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{y})/E\{f(\mathbf{X})\,|\,\mathbf{y}\}\} = \mathbf{y} \quad (30)$$

So **y** is a fixed point of Γ′. **Q.E.D.**

Proof (Lemma 5). We compute the r^{th} element of the m^{th} column of $\partial_{\mathbf{p}}\Gamma'(\mathbf{x}')$,

$$\left.\frac{\partial\Gamma'_r(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = \frac{(\sum_{\mathbf{x}\in D} x_r f(\mathbf{x}) \left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'})(\sum_{\mathbf{x}\in D} f(\mathbf{x})P(\mathbf{X}=\mathbf{x}|\mathbf{x}'))}{(\sum_{\mathbf{x}\in D} f(\mathbf{x})P(\mathbf{X}=\mathbf{x}|\mathbf{x}'))^2} - \frac{(\sum_{\mathbf{x}\in D} x_r f(\mathbf{x})P(\mathbf{X}=\mathbf{x}|\mathbf{x}'))(\sum_{\mathbf{x}\in D} f(\mathbf{x}) \left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'})}{(\sum_{\mathbf{x}\in D} f(\mathbf{x})P(\mathbf{X}=\mathbf{x}|\mathbf{x}'))^2} \quad (31)$$

By (18) and (19), we conclude that

$$\sum_{\mathbf{x}\in D} f(\mathbf{x})P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{x}') = f(\mathbf{x}') \quad (32)$$

$$\sum_{\mathbf{x}\in D} x_r f(\mathbf{x})P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{x}') = x'_r f(\mathbf{x}') \quad (33)$$

Using (32) and (33), we can rewrite (31) as follows,

$$\left.\frac{\partial\Gamma'_r(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = \frac{(\sum_{\mathbf{x}\in D} x_r f(\mathbf{x}) \left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}) f(\mathbf{x}')}{f(\mathbf{x}')^2} - \frac{x'_r f(\mathbf{x}')(\sum_{\mathbf{x}\in D} f(\mathbf{x}) \left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'})}{f(\mathbf{x}')^2} = \sum_{\mathbf{x}\in D}(x_r - x'_r)\frac{f(\mathbf{x})}{f(\mathbf{x}')}\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} \quad (34)$$

By definition 2, we have

$$\left.\frac{\partial\Gamma'_r(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = \underbrace{\sum_{\substack{\mathbf{x}\in D1\\ x_m = x'_m}}(x_r - x'_r)\frac{f(\mathbf{x})}{f(\mathbf{x}')}\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}}_{=0} + \sum_{\substack{\mathbf{x}\in D1\\ x_m \neq x'_m}}(x_r - x'_r)\frac{f(\mathbf{x})}{f(\mathbf{x}')}\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} + \underbrace{\sum_{\mathbf{x}\in D2}(x_r - x'_r)\frac{f(\mathbf{x})}{f(\mathbf{x}')}\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}}_{=0} + \underbrace{\sum_{\mathbf{x}\in D3}(x_r - x'_r)\frac{f(\mathbf{x})}{f(\mathbf{x}')}\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}}_{=0} \quad (35)$$

By looking (35), we conclude

$$\left.\frac{\partial\Gamma'_r(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = \begin{cases} \frac{f(\mathbf{x})}{f(\mathbf{x}')} & if\ m = r\ and\ \mathbf{x}(m,\mathbf{x}') = \mathbf{x} \\ 0 & otherwise \end{cases} \quad (36)$$

or

$$\partial\Gamma'_p(\mathbf{x}') = diag\{\frac{f(\mathbf{x}(1,\mathbf{x}'))}{f(\mathbf{x}')},\ldots,\frac{f(\mathbf{x}(l,\mathbf{x}'))}{f(\mathbf{x}')}\} \quad (37)$$

So the i^{th} eigenvalue of $\partial_{\mathbf{p}}\Gamma'(\mathbf{x}')$ is computed as follows,

$$\lambda_i = \frac{f(\mathbf{x}(i,\mathbf{x}'))}{f(\mathbf{x}')} \quad (38)$$

and hence the proof**. Q.E.D.**

Proof (Theorem 2). By theorem 1 and lemma 5, the stability condition of **x**′ is as follows:

$$\forall 1\le i\le l:\ \lambda_i = \frac{f(\mathbf{x}(i,\mathbf{x}'))}{f(\mathbf{x}')} < 1 \Rightarrow f(\mathbf{x}(i,\mathbf{x}')) < f(\mathbf{x}') \quad (39)$$

In other words, the fitnesses of all **x**s whose hamming distances to **x**′ are 1 are lower than the fitness of **x**′. Therefore by definition 1, **x**′ is an absolute local maximum of *f*.

On the other hand, if **x**′ is not an absolute local maximum of *f,* there is a *j* where **x**(*j*,**x**′)=**x** and *f*(**x**)>*f*(**x**′). Therefore λ_j >1 and by theorem 1, **x**′ is an unstable point. **Q.E.D.**

Proof (Lemma 6). Assume **y** belong to *D* and **p** be equal to **y**. It is clear that the probability of sampling a chromosome different from **y** is zero, therefore by (14), (18) and (19) we have,

$$\Gamma'(\mathbf{y}) = \begin{cases} 0 & \mathbf{y}\in\{\mathbf{x}\,|\,f(\mathbf{x})<\beta\} \\ \frac{\sum_{f(\mathbf{x})\ge\beta}\mathbf{x}P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{y})}{\sum_{f(\mathbf{x})\ge\beta}P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{y})} = \mathbf{y} & \mathbf{y}\notin\{\mathbf{x}\,|\,f(\mathbf{x})<\beta\} \end{cases} \quad (40)$$

By looking (12), it is clear that if *f*(**y**)≥*β,* **y** is a fixed point of Γ′. **Q.E.D.**

Proof (Lemma 7). We consider two cases: **x**′∈{**x**|*f*(x)<*β*} and **x**′∈{**x**|*f*(x)≥*β*}.

Case 1: If **x**′∈{**x**|*f*(x)<*β*}, by assumption *β>0*, we have

$$\xi = \sum_{f(\mathbf{x})\geq\beta} P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{x}')=0 \quad (41)$$

That is a contradiction, therefore in the following, we assume $\mathbf{x}'\in\{\mathbf{x}|f(\mathbf{x})\geq\beta\}$.

Case 2: Similar to the proof of Lemma 5, to compute $\partial_{\mathbf{p}}\Gamma'(\mathbf{x}')$, its elements are computed. Let $(\partial\Gamma'_r(\mathbf{p})/\partial p_m)$ be the r^{th} element of the m^{th} column of $\partial_{\mathbf{p}}\Gamma'(\mathbf{x}')$,

$$\left.\frac{\partial\Gamma'_r(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = \frac{\left(\sum_{f(\mathbf{x})\geq\beta} x_r \left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}\right)\times\left(\sum_{f(\mathbf{x})\geq\beta} P(\mathbf{X}=\mathbf{x}|\mathbf{x}')\right)}{\left(\sum_{f(\mathbf{x})\geq\beta} P(\mathbf{X}=\mathbf{x}|\mathbf{x}')\right)^2} - \frac{\left(\sum_{f(\mathbf{x})\geq\beta} x_r P(\mathbf{X}=\mathbf{x}|\mathbf{x}')\right)\times\left(\sum_{f(\mathbf{x})\geq\beta} \left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}\right)}{\left(\sum_{f(\mathbf{x})\geq\beta} P(\mathbf{X}=\mathbf{x}|\mathbf{x}')\right)^2} \quad (42)$$

By (18) and (19), we have

$$\sum_{f(\mathbf{x})\geq\beta} P(\mathbf{X}=\mathbf{x}|\mathbf{x}') = \begin{cases} 1 & f(\mathbf{x}')\geq\beta \\ 0 & f(\mathbf{x}')<\beta \end{cases} \quad (43)$$

$$\sum_{f(\mathbf{x})\geq\beta} x_r P(\mathbf{X}=\mathbf{x}|\mathbf{x}') = \begin{cases} x'_r & f(\mathbf{x}')\geq\beta \\ 0 & f(\mathbf{x}')<\beta \end{cases} \quad (44)$$

With respect to $f(\mathbf{x}')\geq\beta$, (43), and (44), we rewrite (42) as

$$\left.\frac{\partial\Gamma'_r(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = \sum_{f(\mathbf{x})\geq\beta}(x_r-x'_r)\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}$$

$$= \sum_{\substack{f(\mathbf{x})\geq\beta\\ \mathbf{x}\in D1\\ x_m\neq x'_m}}(x_r-x'_r)\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} + \underbrace{\sum_{\substack{f(\mathbf{x})\geq\beta\\ \mathbf{x}\in D1\\ x_m=x'_m}}(x_r-x'_r)\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}}_{=0} + \underbrace{\sum_{\substack{f(\mathbf{x})\geq\beta\\ \mathbf{x}\in D2}}(x_r-x'_r)\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}}_{=0} + \underbrace{\sum_{\substack{f(\mathbf{x})\geq\beta\\ \mathbf{x}\in D3}}(x_r-x'_r)\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}}_{=0} =$$

$$\begin{cases} 1 & if\ \ r=m\ ,\ \ f(\mathbf{x}(m,\mathbf{x}'))\geq\beta \\ 0 & otherwise \end{cases} \quad (45)$$

and hence the proof. **Q.E.D.**

Proof (Theorem 3). We consider two cases,

Case 1- for all $\mathbf{x}(i,\mathbf{x}')$, where $i=1,\ldots,l$, we have

$$f(\mathbf{x}(i,\mathbf{x}'))<\beta \quad (46)$$

In this case by Lemma 7,

$$\partial_{\mathbf{p}}\Gamma'(\mathbf{x}')=\mathbf{0} \quad (47)$$

On the other hand, with respect to $f(\mathbf{x})\geq\beta$, $\mathbf{x}'$ is an absolute local maximum. Thus by theorem 1, $\mathbf{x}'$ is a stable fixed point.

Case 2- there exists a $\mathbf{x}(k,\mathbf{x}')$, where $1\leq k\leq l$ and

$$f(\mathbf{x}(k,\mathbf{x}'))>\beta \quad (48)$$

In this case, by Lemma 7, k^{th} eigenvalue of $\partial_{\mathbf{p}}\Gamma'(x')$ is equal to one and we cannot say anything about its stability based on Theorem 1. **Q.E.D.**

Proof (Lemma 8). Assume $\mathbf{y}$ belong to D and $\mathbf{p}$ be equal to $\mathbf{y}$, it is clear that the probability of sampling a chromosome different from $\mathbf{y}$ is zero, therefore by (15), (18) and (19) we have,

$$\Gamma'(\mathbf{y}) = \sum_{\mathbf{x}\in D}\{\mathbf{x}\exp(\theta f(\mathbf{x}))P(\mathbf{X}=\mathbf{x}|\mathbf{y})/E\{\exp(\theta f(\mathbf{X}))|\mathbf{y}\}\} \quad (49)$$
$$=(\mathbf{y}\exp(\theta f(\mathbf{y})))/\exp(\theta f(\mathbf{y}))=\mathbf{y}$$

So $\mathbf{y}$ is a fixed point of Γ'.

Proof (Lemma 9). We compute the r^{th} element of the m^{th} column of $\partial_{\mathbf{p}}\Gamma'(\mathbf{x}')$,

$$\left.\frac{\partial\Gamma'_r(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = \left\{\frac{\left(\sum_{\mathbf{x}\in D} x_r\exp(\theta\ f(\mathbf{x}))\left.\frac{\partial P(\mathbf{X}=\mathbf{x}|\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}\right)\times\left(\sum_{\mathbf{x}\in D}\exp(\theta\ f(\mathbf{x}))P(\mathbf{X}=\mathbf{x}|\mathbf{x}')\right)}{\left(\sum_{\mathbf{x}\in D}\exp(\theta\ f(\mathbf{x}))P(\mathbf{X}=x|\mathbf{x}')\right)^2}\right\} -$$

$$\left\{\frac{(\sum_{\mathbf{x}\in D} x_r \exp(\theta\ f(\mathbf{x}))P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{x}')) \times (\sum_{\mathbf{x}\in D} \exp(\theta\ f(\mathbf{x})) \left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'})}{(\sum_{\mathbf{x}\in D} \exp(\theta\ f(\mathbf{x}))P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{x}'))^2}\right\} \tag{50}$$

By (18) and (19), we conclude that

$$\sum_{\mathbf{x}\in D} \exp(\theta\ f(\mathbf{x}))P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{x}') = \exp(\theta\ f(\mathbf{x}')) \tag{51}$$

$$\sum_{\mathbf{x}\in D} x_r \exp(\theta\ f(\mathbf{x}))P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{x}') = x'_r \exp(\theta\ f(\mathbf{x}')) \tag{52}$$

Therefore (50) can be rewritten as follows,

$$\begin{aligned}\left.\frac{\partial \Gamma'_r(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} &= \frac{(\sum_{\mathbf{x}\in D} x_r \exp(\theta f(\mathbf{x})) \left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'})\exp(\theta f(\mathbf{x}'))}{\exp(\theta\ f(\mathbf{x}'))^2} - \\ &\frac{x'_r \exp(\theta\ f(\mathbf{x}'))(\sum_{\mathbf{x}\in D} \exp(\theta\ f(\mathbf{x})) \left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'})}{\exp(\theta\ f(\mathbf{x}'))^2} \\ &= \sum_{\mathbf{x}\in D} (x_r - x'_r) \frac{\exp(\theta\ f(\mathbf{x}))}{\exp(\theta\ f(\mathbf{x}'))} \left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}\end{aligned} \tag{53}$$

By definition 2, we conclude that,

$$\begin{aligned}\left.\frac{\partial \Gamma'_r(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} &= \underbrace{\sum_{\substack{\mathbf{x}\in D1\\ x_m = x'_m}} (x_r - x'_r) \frac{\exp(\theta\ f(\mathbf{x}))}{\exp(\theta\ f(\mathbf{x}'))} \left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}}_{=0} \\ &+ \sum_{\substack{\mathbf{x}\in D1\\ x_m \neq x'_m}} (x_r - x'_r) \frac{\exp(\theta f(\mathbf{x}))}{\exp(\theta f(\mathbf{x}'))} \left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} \\ &+ \underbrace{\sum_{\mathbf{x}\in D2} (x_r - x'_r) \frac{\exp(\theta f(\mathbf{x}))}{\exp(\theta f(\mathbf{x}'))} \left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}}_{=0} \\ &+ \underbrace{\sum_{\mathbf{x}\in D3} (x_r - x'_r) \frac{\exp(\theta f(\mathbf{x}))}{\exp(\theta f(\mathbf{x}'))} \left.\frac{\partial P(\mathbf{X}=\mathbf{x}\,|\,\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'}}_{=0}\end{aligned} \tag{54}$$

By (54), we have

$$\left.\frac{\partial \Gamma'_r(\mathbf{p})}{\partial p_m}\right|_{\mathbf{x}'} = \begin{cases} \dfrac{\exp(\theta f(\mathbf{x}))}{\exp(\theta f(\mathbf{x}'))} & \textit{if } m = r \textit{ and } \mathbf{x}(m,\mathbf{x}') = \mathbf{x} \\ 0 & \textit{otherwise} \end{cases} \tag{55}$$

or

$$\partial\Gamma'_p(\mathbf{x}') = diag\{\frac{\exp(\theta\ f(\mathbf{x}(1,\mathbf{x}')))}{\exp(\theta\ f(\mathbf{x}'))},\ldots,\frac{\exp(\theta\ f(\mathbf{x}(l,\mathbf{x}')))}{\exp(\theta\ f(\mathbf{x}'))}\} \tag{56}$$

So i^{th} eigenvalue of $\partial_{\mathbf{p}}\Gamma'(\mathbf{x}')$ is computed as follows,

$$\lambda_i = \frac{\exp(\theta\ f(\mathbf{x}(i,\mathbf{x}')))}{\exp(\theta\ f(\mathbf{x}'))} \tag{57}$$

and hence the proof**. Q.E.D.**

Proof (Theorem 4). By theorem 1 and lemma 9, the stability condition of $\mathbf{x}'$ is

$$\forall 1 \le i \le l\ \ \lambda_i = \frac{\exp(\theta\ f(\mathbf{x}(i,\mathbf{x}')))}{\exp(\theta\ f(\mathbf{x}'))} < 1 \quad \underset{\theta>0}{\Rightarrow} f(\mathbf{x}(i,\mathbf{x}')) < f(\mathbf{x}') \tag{58}$$

In other words, the fitnesses of all **x**s whose hamming distance to $\mathbf{x}'$ are 1 are lower than the fitness of $\mathbf{x}'$. Therefore by definition 1, $\mathbf{x}'$ is an absolute local maximum of *f.*

On the other hand, if $\mathbf{x}'$ is not an absolute local maximum of *f,* then there is a *j* where $\mathbf{x}(j,\mathbf{x}')=\mathbf{x}$ and $f(\mathbf{x}) > f(\mathbf{x}')$. Therefore $\lambda_j > 1$ and by theorem 1, $\mathbf{x}'$ is an unstable point. **Q.E.D.**

References

[1] Baluja, S., "Population Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning", Technical Report, Carnegie Mellon University, 1994.

[2] Baluja, S., and Davis, S., "Fast Probabilistic Modeling for Combinatorial Optimization", Proceedings of the 15 National Conference on Artificial Intelligence, Madison, Wisconsin, AAAI Press, pp. 469-476, 1998.

[3] Bonet, D., Isbell, J. S., and Viola, P., "MIMIC: Finding Optima by Estimation Probability Densities", Advances in Neural Information Processing Systems, Vol. 9, pp. 424-431, Cambridge, MIT Press, 1997.

[4] González, C., Lozano, J. A., and Larrañaga, P., "Analyzing the PBIL Algorithm by Means of

Discrete Dynamical Systems, Complex Systems, Vol. 12, pp. 465-479, 2000.
[5] González, C., Lozano, J. A., and Larrañaga, P., "The Convergence Behavior of the PBIL Algorithm: A Preliminary Approach", in 5th International Conference on Artificial Neural Networks and Genetic Algorithms, ICANNGA'2001.
[6] Harik, G. R., Lobo, F., G., and Goldberg, D. E., "The Compact Genetic Algorithm", IEEE Transaction on Evolutionary Computation, Vol. 3, No. 4, 1999.
[7] Höhfeld, M., and Rudolph, G., "Toward A Theory of Population Based Incremental Learning", Proceedings of the 4th IEEE Conferences on Evolutionary Computation, Indianapolis, IEEE Press, pp. 1-5, 1997.
[8] Larrañaga, P., and Lozano, J. A., *Estimation of Distribution Algorithms. A New tools for Evolutionary Computation*, Kluwer Academic Publishers, 2001.
[9] Maxwell, B., and Anderson, S., "Training Hidden Markov Models using Population-Based Incremental Learning", In Genetic and Evolutionary Computation Conference, GECCO-99, 1999.
[10] Mühlenbein, H., "The Equation for Response to Selection and Its Use for Prediction", Evolutionary Computation, Vol. 5, pp. 303-346, 1998.
[11] Mühlenbein, H., and Mahnig, T., "Convergence Theory and Application of the Factorized Distribution Algorithm", Journal of Computing and Information Technology, Vol 7, PP. 19-32, 1999.
[12] Mühlenbein, H., Mahnig, T., and Rodriguez, A. O., "Schemata, Distributions and Graphical Models in Evolutionary Optimization", Journal of Heuristics, Vol. 5, pp. 215-247, 1999.
[13] Pelikan, M., Goldberg D. E., and Cantz Paz, E., "BOA: the Bayesian Optimization Algorithm", Proceedings of the Genetic and Evolutionary Computation Conference, Orlando, Morgan Kaufmann Publishers, pp. 525-532, 1999.
[14] Rastegar, R., and M. R. Meybodi, "A New Estimation of Distribution Algorithm based on Learning Automata", to appear in the proceedings of IEEE Conference on Evolutionary Computation 2005, UK, 2005.
[15] Sebag, M., and Ducoulombier, A., "Extending Population-Based Incremental Learning to Continues Search Spaces", in Parallel Problem Solving from Nature, SN-V, pp. 418-427, Springer, 1998.
[16] Servais, M. P., De Jaer, G., and Greene, J. R., "Function Optimization using Multiple-Base Population Based Incremental Learning", In Proceedings of the Eighth South African Workshop on Pattern Recognition, 1997.
[17] Sheinerman, E. R., *Invitation to Dynamical Systems*, Printice-Hall, 1996.
[18] Vose, M. D., "Random Heuristic Search", Theoretical Computer Science, No. 229, Vol. 1-2, pp. 103-142, 1999.
[19] Zhang, Q., "On Stability of Fixed Points of Limit Models of Univariate Marginal Distribution Algorithm and Factorized Distribution Algorithm", IEEE Transactions on Evolutionary Computation, Vol. 8, No.1, 2004.
[20] Zhang, Q., and Mühlenbein, H., "On the Convergence of a Class of Estimation of Distribution Algorithms", IEEE Transaction on Evolutionary Computation, Vol. 8, No. 2, 2004.