Cauchy Annealing Schedule:
An Annealing Schedule for Boltzmann Selection Scheme in Evolutionary Algorithms

Ambedkar Dukkipati, M. Narasimha Murty, and Shalabh Bhatnagar

Dept. of Computer Science and Automation
Indian Institute of Science
Bangalore 560012, India
Email: {ambedkar,mnm,shalabh}@csa.iisc.ernet.in

Abstract— **Boltzmann selection is an important selection mechanism in evolutionary algorithms as it has theoretical properties which help in theoretical analysis. However, Boltzmann selection is not used in practice because a good annealing schedule for the 'inverse temperature' parameter is lacking. In this paper we propose a Cauchy annealing schedule for Boltzmann selection scheme based on a hypothesis that selection-strength should increase as evolutionary process goes on and distance between two selection strengths should decrease for the process to converge. To formalize these aspects, we develop formalism for selection mechanisms using fitness distributions and give an appropriate measure for selection-strength. In this paper, we prove an important result, by which we derive an annealing schedule called Cauchy annealing schedule. We demonstrate the novelty of proposed annealing schedule using simulations in the framework of genetic algorithms.**

I. INTRODUCTION

Selection is a central concept in evolutionary algorithms. There are several selection mechanisms in genetic algorithms, like proportionate selection, ranking selection, tournament selection, truncation selection and Boltzmann selection [1]. Among all these selection mechanisms, Boltzmann selection has an important place because it has some nice theoretical properties in some models of evolutionary algorithms [2]. For example, Boltzmann selection is extensively used in statistical mechanics theory of evolutionary algorithms [3]–[6].

Moreover, Boltzmann selection scheme is not used often in solving practical problems because, similar to simulated annealing, it needs an annealing schedule for perturbing the value of inverse temperature parameter used in Boltzmann selection, which is difficult to choose [2]. This problem is well known from simulated annealing [7], an optimization algorithm where noise is introduced by means of a formal temperature. Lowering, or "annealing," the temperature from high to low values in the course of the optimization leads to improved results compared to an optimization at fixed temperature [8]. However, there remains the problem of choosing a suitable annealing schedule for a given optimization problem. The same problem occurs in population-based optimization algorithms, and this paper address this problem for evolutionary algorithms.

Usually, in evolutionary algorithms, probabilistic selection mechanisms are characterized by *selection probabilities* [9]. For a population $P = \{\omega_i\}_{i=1}^{n_P}$, selection probabilities $\{p_i\}_{i=1}^{n_P}$ are defined as,

$$p_i = \text{Prob}(\omega_i \in \text{selection}(P)|\omega_i \in P) \ \forall i = 1 \ldots n_P \ ,$$

and $\{p_i\}_{i=1}^{n_P}$ satisfies the condition: $\sum_{i=1}^{n_P} p_i = 1$.

Let $\{f(\omega_i)\}_{i=1}^{n_P}$ be the corresponding fitness values. The proportionate selection assigns selection probabilities according to the relative fitness of individuals as [10]:

$$p_i = \frac{f(\omega_i)}{\sum_{j=1}^{n_P} f(\omega_j)} \ . \tag{1}$$

Similarly Boltzmann selection is represented as [11]:

$$p_i = \frac{e^{\gamma f(\omega_i)}}{\sum_{j=1}^{n_P} e^{\gamma f(\omega_j)}} \ , \tag{2}$$

where γ is called *inverse temperature*. The strength of selection is controlled by the parameter γ. A higher value of γ (low temperature) gives a stronger selection, and a lower value of γ gives a weaker selection. For details of representation of other selection mechanisms refer [1], [9], [12].

Some properties of selection mechanisms that are desirable in order to control the search process are [9]:

- The impact of the control parameters on selective pressure should be simple and predictable.
- One single control parameter for selective pressure is preferable.
- The range of selective pressure that can be realized by varying the control parameter should be as large as possible.

Boltzmann selection satisfies above properties. Boltzmann selection gives faster convergence, but without good annealing schedule for γ, it might lead to premature convergence.

In this paper we propose Cauchy criteria for choosing the Boltzmann selection schedule. Based on this we derive an annealing schedule for the inverse temperature parameter γ, using a result we proved. Since selection depends only on the fitnesses of candidate solutions of population, in this paper we

characterize the selection using *normalized fitness distribution* (normalized fitness distribution is precisely normalization of fitness distribution of population) instead of selection probabilities which are defined for all the members of population. We also give a new measure for selection-strength which is suitable for the theoretical analysis presented in this paper.

The outline of the paper is as follows. In § II, we present the formalization of selection methods. We present our main results regarding Cauchy criteria for Boltzmann selection schedule in § III. We present simulation results in § IV.

II. A FORMALIZATION OF SELECTION SCHEMES

A. Definitions

Let $f : \Omega \to \mathbb{R}^+ \cup \{0\}$ be a fitness function, where Ω is the search space. Let $P = \{\omega_k\}_{k=1}^{n_P}$ denote the population. Here we assume that the size of population at any time is finite and need not be a constant.

Fitness distribution is an important macroscopic property of population. Formal definition of fitness distribution of a population is given below [13].

Definition 2.1: Fitness distribution of a population $P = \{\omega_k\}_{k=1}^{n_P}$ is a function $\rho^P : \mathbb{R} \to \mathbb{Z}^+ \cup \{0\}$ defined as

$$\rho^P(x) = \sum_{k=1}^{n_P} \delta(x - f(\omega_k)) , \qquad (3)$$

where $\delta : \mathbb{R} \to \{0, 1\}$ is the Kronecker delta function defined as $\delta(x) = 1$ if $x = 0$, $\delta(x) = 0$ otherwise .
ρ^P assigns each $x \in \mathbb{R}$, the number of individuals in a population P carrying x as the fitness value. The finite set of values associated with the fitness distribution which are mapped to non-zero values is called *support* of fitness distribution of population.

Definition 2.2: Let ρ^P be the fitness distribution of population P, then 'support' of ρ^P is defined as [1]

$$supp(\rho^P) = E_\rho(\text{or} E_P) = \{x : \rho^P(x) \neq 0\} . \qquad (4)$$

For any population P, $supp(\rho^P)$ is finite set, since population size is finite. We can write size of a population P in terms of its fitness distribution ρ^P as,

$$n_P = \sum_{x \in E_P} \rho^P(x) . \qquad (5)$$

We now define *normalized fitness distribution* (NFD).

Definition 2.3: Normalized fitness distribution (NFD) of a population $P = \{\omega_k\}_{k=1}^{n_P}$ with fitness distribution ρ^P is a function $\varphi^P : \mathbb{R} \to [0, 1]$ defined as

$$\varphi^P(x) = \frac{\rho^P(x)}{n_P} , \quad \forall x \in \mathbb{R} . \qquad (6)$$

One can see that φ^P is well defined. From (5), we have

$$\sum_{x \in E_P} \varphi^P(x) = 1 . \qquad (7)$$

Note that $supp(\varphi^P) = supp(\rho^P)$. Support of a NFD φ of population P is represented by E_φ.

---

[1]The actual definition of support of ρ^P is $\overline{\{x : \rho^P(x) \neq 0\}}$. The overline denotes the closure of the set. Since $\{x : \rho^P(x) \neq 0\}$ is finite $\{x : \rho^P(x) \neq 0\} = \overline{\{x : \rho^P(x) \neq 0\}}$

B. Representation of Selection Schemes Via NFD

Instead of giving a mechanistic view of selection, we define selection as an operator on fitness distribution (hence on NFD). For that we need to specify the corresponding space.

Definition 2.3 gives the definition of "NFD of a population". To define space of all NFDs we give a generalized definition of NFD, similar to the generalized definition of fitness distribution given in [13].

Definition 2.4: 'Normalized fitness distribution' (NFD) is a function $\varphi : \mathbb{R} \to [0, 1]$ which satisfies

$$\sharp\{x : \varphi(x) \neq 0\} < \infty \quad (i.e., \sharp supp(\varphi) < \infty) , \qquad (8a)$$

$$\sum_{x \in supp(\varphi)} \varphi(x) = 1 , \qquad (8b)$$

where $\sharp$ denotes the cardinality of a set.
From Definition 2.3, one can easily see that every "NFD of a population" is indeed an "NFD". Space of all NFDs is denoted by $\mathcal{O}$ i.e.,

$$\mathcal{O} = \{\varphi : \mathbb{R} \to [0, 1] : \sharp supp(\varphi) < \infty, \sum_{x \in supp(\varphi)} \varphi(x) = 1\} . \qquad (9)$$

We define selection as an operator Γ on the space $\mathcal{O}$ i.e., $\Gamma : \mathcal{O} \to \mathcal{O}$. At generation k, for a population P_k, with fitness distribution ρ^k and population size N_k, Boltzmann selection Γ can be represented in terms of fitness distribution as

$$\rho^{k+1}(x) = \Gamma\rho^k(x) = \rho^k(x)\frac{e^{\gamma x}}{\sum_{y \in E} \rho^k(y)e^{\gamma y}}N_{k+1} , \qquad (10)$$

where N_{k+1} is the population size after the selection Γ and $E = supp(\rho^k)$. From Definition 2.3, we have

$$\varphi^{k+1}(x) = \frac{\Gamma\rho^k(x)}{N_{k+1}} = \rho^k(x)\frac{e^{\gamma x}}{\sum_{y \in E} \rho^k(y)e^{\gamma y}}$$

$$= \frac{\varphi^k(x)}{N_k}\frac{e^{\gamma x}}{\sum_{y \in E} \frac{\varphi^k(y)}{N_k}e^{\gamma y}} .$$

$$\varphi^{k+1}(x) = \frac{\varphi^k(x)e^{\gamma x}}{\sum_{y \in E} \varphi^k(y)e^{\gamma y}} .$$

Hence Boltzmann selection operator Γ on $\mathcal{O}$ is defined as

$$\Gamma\varphi(x) = \frac{\varphi(x)e^{\gamma x}}{\sum_{y \in E} \varphi(y)e^{\gamma y}} , \quad \forall x \in \mathbb{R} , \quad \forall\varphi \in \mathcal{O} , \qquad (11)$$

where $\gamma \in \mathbb{R}^+ \cup \{0\}$ corresponds to inverse temperature. Similarly we can define proportionate selection using operator Γ_{prop} as follows:

$$\Gamma_{\text{prop}}\varphi(x) = \frac{x\varphi(x)}{\sum_{y \in E} y\varphi(y)} , \quad \forall x \in \mathbb{R} , \quad \forall\varphi \in \mathcal{O} . \qquad (12)$$

Through out this paper we represent Boltzmann selection by Γ unless mentioned otherwise.

C. Metric on Space of NFDs

One can view NFD as a probability distribution and one can use various distance measures on it. For example, one can use Kullback-Leibler distance measure but it is not a metric [14]. We define a metric $d : \mathcal{O} \times \mathcal{O} \rightarrow \mathbb{R}$ according to

$$d(\varphi_1, \varphi_2) = \sum_{x \in E_{\varphi_1} \cup E_{\varphi_2}} |\varphi_1(x) - \varphi_2(x)| \quad, \quad \forall \varphi_1, \varphi_2 \in \mathcal{O} \ . \tag{13}$$

It is easy to verity that d is indeed a metric on $\mathcal{O}$.

D. Selection Strength

There have been several variants to measure selection strength in evolutionary algorithms. The terminology "selection intensity" or "selection pressure" is often used to describe this property of selection.

The concept of "take over time" quantifies selection pressure by the number of generations required by repeated application of selection, to fill the complete population with copies of the single initially best individual [15]. There have been some adaptations of definitions from population genetics for selection intensity. The change in average fitness of the population due to selection is a reasonable measure of selection intensity [16]. Also note that several of these measures depend on fitness distribution at that instance. Details of selection intensity measures can be found in [9], [15], [16].

We measure selection strength w.r.t an NFD using the metric d as distance between the NFD before the selection and after selection. Let $\Gamma : \mathcal{O} \rightarrow \mathcal{O}$ be the selection operator. The selection strength can be measured as:

$$d(\varphi, \Gamma\varphi) = \sum_{x \in E_\varphi} |\varphi(x) - \Gamma\varphi(x)| \ . \tag{14}$$

We give the formal definition of selection strength as follows.

Definition 2.5: Selection strength of a selection scheme Γ with respect to an NFD $\varphi \in \mathcal{O}$ is denoted by $S_\varphi(\Gamma)$ and is defined as

$$S_\varphi(\Gamma) = d(\varphi, \Gamma\varphi) \ . \tag{15}$$

For example, for proportionate selection the NFD φ selection strength can be measured as:

$$d(\varphi, \Gamma_{prop}\varphi) = \sum_{x \in E_\varphi} \left| \varphi(x) - \frac{x\varphi(x)}{\sum_{y \in E_\varphi} y\varphi(y)} \right|$$

$$= \sum_{x \in E_\varphi} \varphi(x) \left| \frac{\sum_{y \in E_\varphi} y\varphi(y) - x}{\sum_{y \in E_\varphi} y\varphi(y)} \right| \ , \tag{16}$$

$$d(\varphi, \Gamma_{prop}\varphi) = \frac{\sum_{x \in E_\varphi} \varphi(x) |\mu_\varphi - x|}{\mu_\varphi} \ . \tag{17}$$

where $\mu_\varphi = \sum_{x \in E_\varphi} x\varphi(x)$ is expectation of φ. The numerator is nothing but mean absolute error of φ. If one observes (17) carefully, it justifies the definition of selection strength as $d(\varphi, \Gamma\varphi)$.

III. CAUCHY CRITERIA FOR BOLTZMANN SELECTION SCHEME

A. Boltzmann Selection Scheme

Let $\{P_n\}$ be the evolutionary process, where P_n is population at generation n. We represent corresponding Boltzmann selection scheme as $\{\Gamma_{(n)}\}$ where $\Gamma_{(n)}$ is an operator $\Gamma_{(n)} : \mathcal{O} \rightarrow \mathcal{O}$ and is defined as:

$$\varphi_n(x) = \Gamma_{(n)}\varphi_{n-1}(x) = \frac{\varphi_{n-1}(x)e^{\gamma_n x}}{\sum_{y \in E_{\varphi_{n-1}}} \varphi_{n-1}(y)e^{\gamma_n y}} \ ,$$
$$\forall x \in \mathbb{R} \ , \ \forall n = 1, 2, \ldots \ , \tag{18}$$

where $\varphi_n \in \mathcal{O}$. $\{\gamma_n\}$ is annealing schedule for the Boltzmann selection scheme $\{\Gamma_{(n)}\}$ and $\gamma_n \geq 0 \ \forall n = 1, 2\ldots$. Also $\{\gamma_n\}$ is a non-decreasing sequence since γ_n represents the inverse temperature [2].

B. Cauchy Criteria

Our Hypothesis for Boltzmann selection schedule is:

> The difference between successive selection pressures should decrease as the evolutionary process proceeds.

We formalize above hypothesis as Cauchy criteria for Boltzmann selection schedule as follows:

Definition 3.1: A Boltzmann selection schedule $\{\Gamma_{(n)}\}$ is said to satisfy Cauchy criteria if $\{\Gamma_{(n)}\varphi\} \subset \mathcal{O}$ is Cauchy with respect to metric d, $\forall \varphi \in \mathcal{O}$.

We justify the fact that Cauchy criteria for Boltzmann selection schedule captures the hypothesis by the following lemma.

Lemma 3.2: Let Γ_1 and Γ_2 be two Boltzmann selection operators. Then for any $\varphi \in \mathcal{O}$, difference between these selection strengths satisfies

$$|S_\varphi(\Gamma_1) - S_\varphi(\Gamma_2)| \leq d(\Gamma_1\varphi, \Gamma_2\varphi) \ . \tag{19}$$

Proof: From Definition 2.5 we have

$$|S_\varphi(\Gamma_1) - S_\varphi(\Gamma_2)| = |d(\varphi, \Gamma_1\varphi) - d(\varphi, \Gamma_2\varphi)| \ .$$

From triangular inequality we have

$$d(\varphi, \Gamma_1\varphi) \leq d(\varphi, \Gamma_2\varphi) + d(\Gamma_1\varphi, \Gamma_2\varphi) \ ,$$

which gives

$$d(\Gamma_1\varphi, \Gamma_2\varphi) \geq d(\varphi, \Gamma_1\varphi) - d(\varphi, \Gamma_2\varphi) \ . \tag{20a}$$

Similarly we have

$$d(\Gamma_1\varphi, \Gamma_2\varphi) \geq d(\varphi, \Gamma_2\varphi) - d(\varphi, \Gamma_1\varphi) \ . \tag{20b}$$

From (20a) and (20b) we get

$$d(\Gamma_1\varphi, \Gamma_2\varphi) \geq |d(\varphi, \Gamma_1\varphi) - d(\varphi, \Gamma_2\varphi)| \ . \qquad \blacksquare$$

Hence decrement in $d(\Gamma_1\varphi, \Gamma_2\varphi)$ results in decrement in the difference between selection strengths. From the definition of Cauchy sequence justification is clear.

Note that above criteria is stated in terms of the selection operator. Based on this we derive an annealing schedule for inverse temperature parameter γ_n in the next section.

C. Derivation of Cauchy Annealing Schedule

We summarize Cauchy criteria for Boltzmann selection schedule $\{\Gamma_{(n)}\}$ as:

(CB1) $\{\gamma_n\}$ is non-decreasing sequence

(CB2) $\{\Gamma_{(n)}\varphi\} \subset \mathcal{O}$ is Cauchy $\forall \varphi \in \mathcal{O}$

For $\{\Gamma_{(n)}\}$ to satisfy (CB1) we define

$$\gamma_n = \sum_{k=1}^{n} g_k \ , \ \text{where} \ \{g_k\} \subset \mathbb{R}^+ \cup 0, \ \forall n = 1, 2, \ldots \ . \ (21)$$

Clearly $\{\gamma_n\}$ is non decreasing sequence. Then Boltzmann selection schedule $\{\Gamma_{(n)}\}$ defined as

$$\Gamma_{(n)}\varphi(x) = \frac{\varphi(x)\exp(x\sum_{k=1}^{n}g_k)}{\sum_{y\in E_\varphi}\varphi(y)\exp(y\sum_{k=1}^{n}g_k)} \ , \forall x \in \mathbb{R} \ , \ (22)$$

for arbitrary $\{g_k\} \subset \mathbb{R}^+ \cup \{0\}$ satisfies (CB1). Now we derive annealing schedule for $\{\gamma_n = \sum_{k=1}^{n} g_k\}_n$ for the selection schedule $\{\Gamma_{(n)}\}$ to satisfy (CB2). First we prove following inequality.

Lemma 3.3: Let $\{\Gamma_{(n)}\}$ be a sequence of Boltzmann selection operators defined as in (22), then for any NFD $\varphi \in \mathcal{O}$, we have

$$d(\Gamma_{(n)}(\varphi), \Gamma_{(m)}(\varphi)) \leq \sum_{x\in E_\varphi}\left(\exp(x\sum_{k=m+1}^{n}g_k) - 1\right)$$

whenever $n > m$ and $n, m \in \mathbb{Z}^+$.

Proof: Denote

$$C_n(x) = \varphi(x)\exp(x\sum_{k=1}^{n}g_k) \ \ \forall x \in E_\varphi \ .$$

Then,

$$d(\Gamma_{(n)}(\varphi), \Gamma_{(m)}(\varphi)) =$$
$$\sum_{x\in E_{\Gamma_{(n)}(\varphi)}\cup E_{\Gamma_{(m)}(\varphi)}}\left|\frac{C_n(x)}{\sum_{y\in E_\varphi}C_n(y)} - \frac{C_m(x)}{\sum_{y\in E_\varphi}C_m(y)}\right| \ .$$

Since $supp(\varphi) \supseteq supp(\Gamma_n(\varphi)) \cup supp(\Gamma_m(\varphi))$ and $supp(C_n) = supp(\varphi) \ \forall n$ we can write

$$d(\Gamma_{(n)}(\varphi), \Gamma_{(m)}(\varphi)) =$$
$$\sum_{x\in E_\varphi}\left|\frac{C_n(x)}{\sum_{y\in E_\varphi}C_n(y)} - \frac{C_m(x)}{\sum_{y\in E_\varphi}C_m(y)}\right|$$

$$\leq \left(\frac{1}{\sum_{x\in E_\varphi}C_m(x)}\right)\left(\sum_{x\in E_\varphi}|C_n(x) - C_m(x)|\right) \ , \ (23)$$

since for $n > m$, $C_n(x) \geq C_m(x), \ \ \forall x > 0$.

We have,

$$C_n(x) = \varphi(x)\exp(x\sum_{k=1}^{n}g_k)$$
$$= \varphi(x)\exp(x\sum_{k=1}^{m}g_k)\exp(x\sum_{k=m+1}^{n}g_k)$$
$$= C_m(x)\exp(x\sum_{k=m+1}^{n}g_k) \ , \ \forall x \in E_\varphi \ . \ (24)$$

Hence we can write (23) as

$$d(\Gamma_{(n)}(\varphi), \Gamma_{(m)}(\varphi)) = \left(\frac{1}{\sum_{x\in E_\varphi}C_m(x)}\right)$$
$$\left(\sum_{x\in E_\varphi}\left|C_m(x)\left(\exp(x\sum_{k=m+1}^{n}g_k) - 1\right)\right|\right)$$

$$\leq \left(\frac{1}{\sum_{x\in E_\varphi}C_m(x)}\right)$$
$$\sqrt{\sum_{x\in E_\varphi}\{C_m(x)\}^2\sum_{x\in E_\varphi}\left\{\exp(x\sum_{k=m+1}^{n}g_k) - 1\right\}^2} \ ,$$

by Cauchy-Schwartz-Bunyakovsky inequality.

Since $C_m(x)$ and $\exp(x\sum_{k=m+1}^{n}g_k) - 1$ are positive, we have

$$d(\Gamma_{(n)}(\varphi), \Gamma_{(m)}(\varphi)) \leq \left(\frac{1}{\sum_{x\in E_\varphi}C_m(x)}\right)$$
$$\sqrt{\left\{\sum_{x\in E_\varphi}C_m(x)\right\}^2\left\{\sum_{x\in E_\varphi}\exp(x\sum_{k=m+1}^{n}g_k) - 1\right\}^2} \ ,$$

$$d(\Gamma_{(n)}(\varphi), \Gamma_{(m)}(\varphi)) \leq \sum_{x\in E_\varphi}\left(\exp(x\sum_{k=m+1}^{n}g_k) - 1\right) \ .$$
$$(25)$$
∎

We now give our main result which gives condition on annealing schedule $\{\gamma_n\}$ for Boltzmann selection to satisfy Cauchy criteria.

Theorem 3.4: Let $\{\Gamma_{(n)}\}$ be a sequence of Boltzmann selection operators defined as in (22). Then,

$$\left\{\sum_{k=1}^{n}g_k\right\}_{(n)} \ \text{is Cauchy} \Longrightarrow \{\Gamma_{(n)}\varphi\} \ \text{is Cauchy}$$

$\forall \varphi \in \mathcal{O}$ and for any $\{g_k\} \subset \mathbb{R}^+ \cup \{0\}$.

Proof: $\{\Gamma_{(n)}\varphi\}$ is Cauchy for any $\varphi \in \mathcal{O}$ if

$$\forall \epsilon > 0, \exists N = N(\epsilon) \in \mathbb{Z}^+ \ni$$

$$n, m \geq N \Rightarrow d(\Gamma_{(n)}(\varphi), \Gamma_{(m)}(\varphi)) < \epsilon \ .$$

Now consider $d(\Gamma_{(n)}(\varphi), \Gamma_{(m)}(\varphi))$. With out loss of generality assume that $n > m$. From Lemma 3.3 we have

$$d(\Gamma_{(n)}(\varphi), \Gamma_{(m)}(\varphi)) \leq \sum_{x \in E_\varphi} \left(\exp(x \sum_{k=m+1}^{n} g_k) - 1 \right) .$$

Let $\epsilon > 0$ arbitrary. So,

$$\sum_{x \in E_\varphi} \left(\exp(x \sum_{k=m+1}^{n} g_k) - 1 \right) < \epsilon \implies$$
$$d(\Gamma_{(n)}(\varphi), \Gamma_{(m)}(\varphi)) < \epsilon . \quad (26)$$

Hence it is enough to prove that

$$\exists N = N(\epsilon) \in \mathbb{Z}^+ \ni n, m \geq N \Rightarrow$$
$$\sum_{x \in E_\varphi} \left(\exp(x \sum_{k=m+1}^{n} g_k) - 1 \right) < \epsilon . \quad (27)$$

Now let $E_\varphi = \{x_i\}_{i=1}^r$. $r < \infty$ since E_φ is finite. We thus have to prove that

$$\exists N = N(\epsilon) \in \mathbb{Z}^+ \ni n, m \geq N \implies$$
$$\sum_{i=1}^{r} \left(\exp(x_i \sum_{k=m+1}^{n} g_k) - 1 \right) < \epsilon . \quad (28)$$

Now it is enough if we show that

$$\exists N_i = N_i(\frac{\epsilon}{r}) \in \mathbb{Z}^+ \ni n, m \geq N_i \implies \quad (29)$$
$$\exp(x_i \sum_{k=m+1}^{n} g_k) - 1 \leq \frac{\epsilon}{r} , \quad \forall i = 1 \ldots r .$$

For $N = \max\{N_i : i = 1 \ldots r\}$

$$n, m \geq N \Rightarrow \exp(x_i \sum_{k=m+1}^{n} g_k) - 1 \leq \frac{\epsilon}{r} , \quad \forall i = 1 \ldots r ,$$
$$(30)$$

which gives us

$$n, m \geq N \Rightarrow \sum_{i=1}^{r} \exp(x_i \sum_{k=m+1}^{n} g_k) - 1 \leq \sum_{i=1}^{r} \frac{\epsilon}{r} = \epsilon . \quad (31)$$

Now to assert (29) it is enough, for a fixed $x \in E_\varphi$, if we have following

$$\forall \epsilon' > 0, \exists N' = N'(\epsilon') \in \mathbb{Z}^+ \ni n, m \geq N' \implies$$
$$\exp \left(x \sum_{k=m+1}^{n} g_k \right) - 1 \leq \epsilon' .$$

Note that ϵ' can be chosen as $\epsilon' = \frac{\epsilon}{r}$, and ϵ' is arbitrary since ϵ arbitrary. Since

$$\exp(x \sum_{k=m+1}^{n} g_k) - 1 \leq \epsilon' \implies \sum_{k=m+1}^{n} g_k \leq \frac{\ln(\epsilon' + 1)}{x}$$

it is enough if

$$\forall \epsilon'' > 0, \exists N'' = N''(\epsilon'') \in \mathbb{Z}^+ \ni n, m \geq N'' \implies$$
$$\sum_{k=m+1}^{n} g_k \leq \epsilon'' . \quad (32)$$

Note that ϵ'' can be chosen as $\epsilon'' = \frac{\ln(\epsilon' + 1)}{x}$ for a fixed $x \in E$ and ϵ'' is arbitrary since ϵ' is arbitrary.

Since ϵ'' is arbitrary (32) can be asserted if the sequence

$$\left\{ \sum_{k=1}^{n} g_k \right\}_{(n)}$$

is Cauchy by the definition of Cauchy sequence.

∎

IV. SIMULATION RESULTS

A. Choice of $\{g_k\}$

As a specific case, for $\{g_k\}$ to satisfy (21), we choose

$$g_k = g_0 \frac{1}{k^\alpha}, \quad (33)$$

where g_0 is any constant and $\alpha > 1$. Since $\left\{ \sum_{k=1}^{n} \frac{1}{k^\alpha} \right\}_n$ is a Cauchy sequence for any $\alpha > 1$ [17], $\left\{ g_0 \sum_{k=1}^{n} \frac{1}{k^\alpha} \right\}_n$ is also a Cauchy sequence. In this specific choice of sequence, α plays an important role in the annealing schedule (see Figure 1).
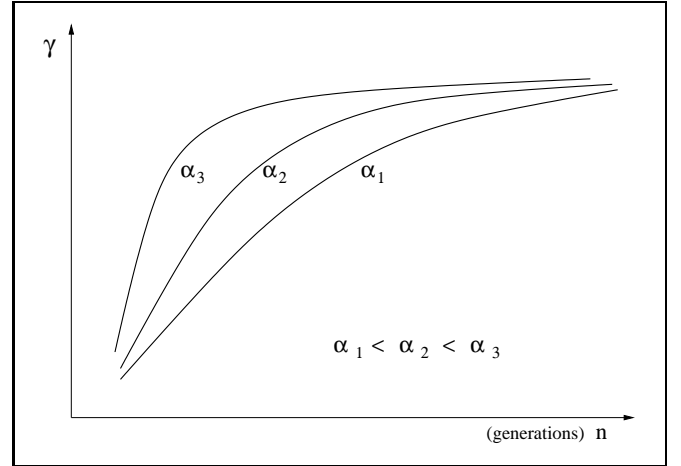

Fig. 1. Cauchy Annealing Schedules for Different Values of α where γ is defined according to (34)

Here we give simulation results using the annealing schedule $\{\gamma_n\}$ defined as

$$\gamma_n = \sum_{k=1}^{n} g_k = g_0 \sum_{k=1}^{n} \frac{1}{k^\alpha}. \quad (34)$$

B. Results

We discuss the simulations conducted to study the annealing schedule for Boltzmann selection proposed in this paper. We compare three selection mechanisms viz., proportionate selection (*proportionate*), Boltzmann selection with constant γ (*Boltzmann*) and Boltzmann selection with proposed Cauchy annealing schedule $\{\gamma_n\}$ (*Cauchy-Boltzmann*). We study multi-variable function optimization in the framework of genetic algorithms. Specifically, we use the following functions [16]:

- Rastrigin's function:
  $f_6(\vec{x}) = lA + \sum_{i=1}^{l} x_i^2 - A\cos(2\pi x_i),$
  where $A = 10$; $-5.12 \leq x_i \leq 5.12$
- Griewangk's function:
  $f_8(\vec{x}) = \sum_{i=1}^{l} \frac{x_i^2}{4000} - \prod_{i=1}^{l} \cos(\frac{x_i}{\sqrt{i}}) + 1,$
  where $-600 \leq x_i \leq 600$
- Ackley's function:
  $f_9(\vec{x}) = -20\exp(-0.2\sqrt{\frac{1}{l}\sum_{i=1}^{l} x_i^2})$
  $- \exp(\frac{1}{l}\sum_{i=1}^{l}\cos(2\pi x_i)) + 20 + e,$
  where $-30 \leq x_i \leq 30$
- Schwefel's function:
  $f_7(\vec{x}) = \sum_{i=1}^{l} -x_i\sin(\sqrt{|x_i|}),$
  where $-500 \leq x_i \leq 500$

The following parameter values have been used in all the experiments:

- Each x_i is encoded with 5 bits and $l = 15$ i.e search space is of size 2^{75}
- Population size $n_P = 150$
- For Boltzmann selection the inverse temperature $\gamma = 300$. For Boltzmann selection with annealing, we vary $\alpha = 1.0001, 1.1, 1.5, 2$ and we chose g_0 for each value of α in such a way that, $\gamma_{100} = 300$ where 100 is the total number of generations for each process. Figure 2 shows the plots of values of γ_n for $\alpha = 1.0001, 1.1, 1.5, 2$.
- For all the experiments probability of uniform crossover is 0.8 and probability of mutation is below 0.1
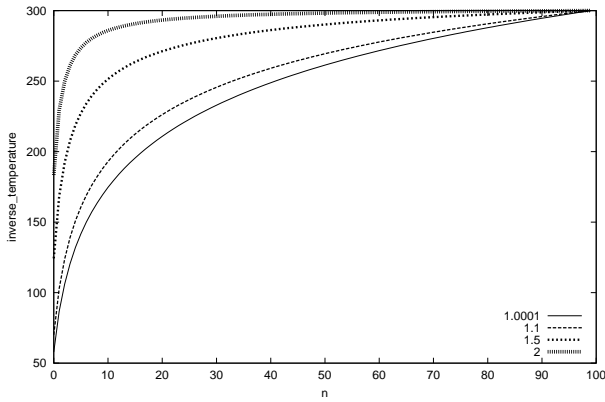- Each simulation is performed 17 times to get the average behavior of the process

Fig. 2. γ_n for $\alpha = 1.0001, 1.1, 1.5, 2$ where γ_n is defined according to (34)

From various simulations we observed that when the problem size is small (for example smaller values of l) all the selection mechanisms perform equally well. Boltzmann selection is effective when we increase the problem size. In the case of Boltzmann selection with constant γ, one has to increase the value of γ when the problem size is large. Note that choice of parameter α is very important for Cauchy annealing schedule and it depends on the specific problem. Here we have given results corresponding to the best values of α. Figures 3, 4, 5, 6, show the plots for behavior of the process when averaged over multiple runs. Figures 7 and 8 show plots for single run. Our simulations showed that Boltzmann selection with the Cauchy annealing schedule performs better than other mechanisms.

V. Conclusions

In this paper we derived an annealing schedule for inverse temperature parameter in the Boltzmann selection scheme, which is based on Cauchy criteria for Boltzmann selection schedule. Usage of Cauchy criteria for Boltzmann selection schedule is justified by the hypothesis: as process goes on

- selection strength should increase,
- difference between the selection strengths should decrease.

We have given alternative formalism for selection mechanisms based on the fitness distributions. We have also given a new measure for selection strength which is suitable for theoretical analysis.

Using the above formalism, we presented an important mathematical result for Boltzmann selection schedule; using which we derived the annealing schedule. Cauchy annealing schedule is a generalized mechanism from which one can choose different specific sequences for annealing based on the problem at hand.

Our simulation results justify the hypothesis we presented and the utility of techniques we used; they also support usage of the mathematical results we presented, in practice. We conducted experiments using specific annealing schedule, where one can choose the speed of (inverse) annealing. We compared our results with algorithms with proportionate selection, Boltzmann selection without annealing schedule and Boltzmann selection with the proposed annealing schedule. We found that with an appropriate choice of speed of annealing, algorithms with annealing schedule outperform other methods.

This analysis does not consider any of the genetic operators. Our future work would involve comprehensive analysis which leads to more generalized selection schedules based on the techniques presented in this paper.

One important consequence of techniques we developed in this paper would be proving convergence of the process. If one can show that the underlying space, for example spaces of NFDs, is *complete* (see Appendix for the definition of complete metric space), one can conclude the convergence of evolutionary process, based on the *Cauchy criteria*.
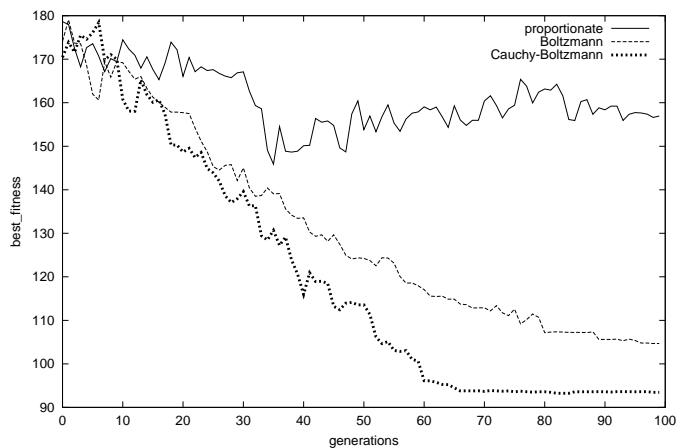
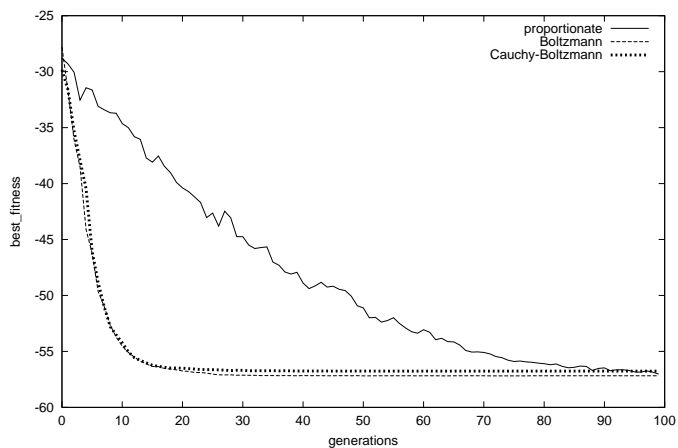Fig. 3. Rastrigin: $\alpha = 2$ (Multiple Runs)

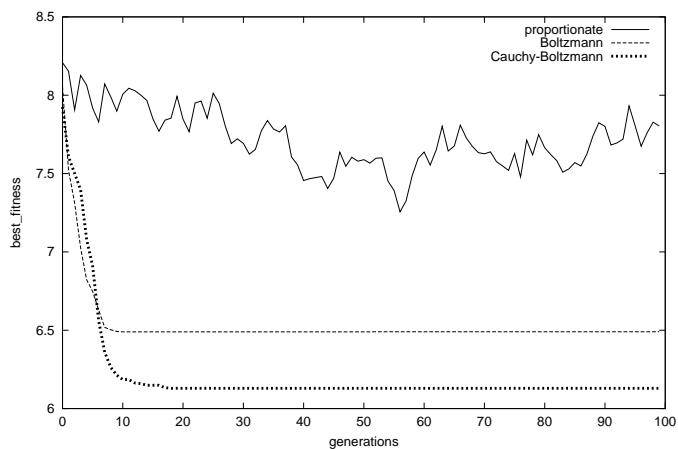Fig. 6. Schwefel: $\alpha = 1.5$ (Multiple Runs)
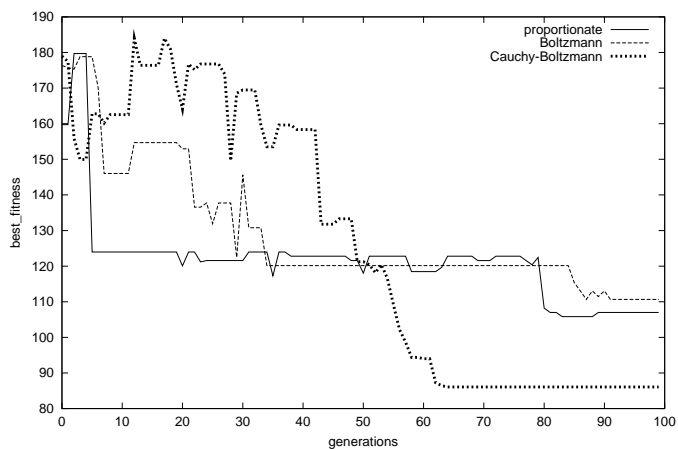

Fig. 4. Ackley: $\alpha = 1.1$ (Multiple Runs)
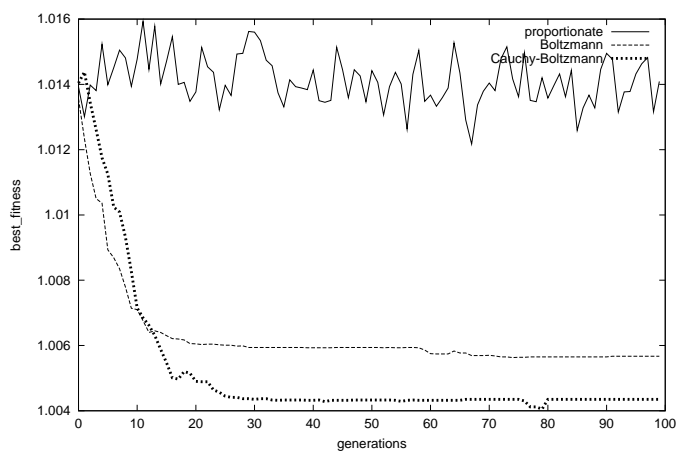

Fig. 7. Rastrigin: $\alpha = 2$ (Single Run)
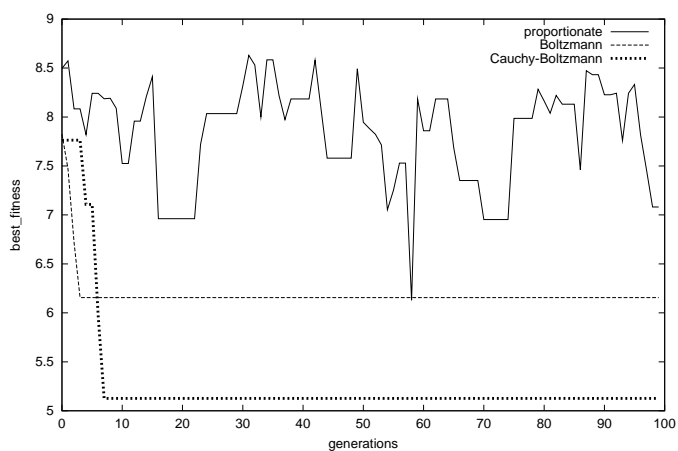

Fig. 5. Griewangk: $\alpha = 1.1$ (Multiple Runs)

Fig. 8. Ackley: $\alpha = 1.1$ (Single Run)

ACKNOWLEDGMENT

APPENDIX
METRIC SPACES

Here we present some basic concepts of metric spaces used in this paper.

Let X be any set. A function $d : X \times X \to \mathbb{R}$ is said to be metric on X if

1) $d(x,y) \geq 0$ and $d(x,y) = 0 \Leftrightarrow x = y$, $\forall x, y \in X$
2) $d(x,y) = d(y,x)$, $\forall x, y \in X$
3) $d(x,y) \leq d(x,z) + d(z,y)$, $\forall x, y, z \in X$ (Triangular inequality)

Example of metric space is $\mathbb{R}$ with $|\,.\,|$ as a metric.

A sequence $\{x_n\}$ is said to be Cauchy sequence if

$$\forall \epsilon > 0, \exists N = N(\epsilon) \in \mathbb{Z}^+ \ni$$

$$n, m \geq N \Rightarrow d(x_n, x_m) < \epsilon \ .$$

We say metric space (X, d) is complete if every Cauchy sequence in X converges.

REFERENCES

[1] Tobias Blickle and Lothar Thiele, "A comparison of selection schemes used in evolutionary algorithms," *Evolutionary Computation*, vol. 4, no. 4, pp. 361–394, 1996.

[2] Thilo Mahnig and Heinz Mühlenbein, "A new adaptive boltzmann selection schedule sds," in *Proceedings of the Congress on Evolutionary Computation (CEC'2001)*. 2001, pp. 183–190, IEEE Press.

[3] Adam Prügel-Bennett and Jonathan Shapiro, "Analysis of genetic algorithms using statistical mechanics," *Physical Review Letters*, vol. 9, no. 72, February 1994.

[4] Adam Prügel-Bennett, "Modelling evolving populations," *Journal of Theoretical Biology*, vol. 185, pp. 81–95, 1997.

[5] Adam Prügel-Bennett and Alex Rogers, "Modelling ga dynamics," in *Theoretical Aspects of Evolutionary Computing*, L. Kallel, B. Naudts, and A. Rogers, Eds., pp. 59–86. Springer-Verlag, Berlin Heidelberg, 2001.

[6] Lars Magnus Rattray, "The dynamics of a genetic algorithm under stabilizing selection," *Complex Systems*, vol. 9, pp. 213–234, 1995.

[7] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines– A Stochastic Approach to Combinatorial Optimization and Neural Computing*, Wiley, New York, 1989.

[8] Stefan Bornholdt, "Annealing schedule from population dynamics," *Physica Review E*, vol. 59, no. 4, pp. 3942–3946, 1999.

[9] Thomas Back, "Selective pressure in evolutionary algorithms: A characterization of selection mechanisms," in *Proceedings of the First IEEE Conference on Evolutionary Computation*, Piscataway, NJ, 1994, pp. 57–62, IEEE Press.

[10] J. H. Holland, *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, MI, 1975.

[11] M. de la Maza and B. Tidor, "An analysis of selection procedures with particular attention paid to proportional and boltzmann selection," in *Proceedings of the Fifth International Conference on Genetic Algorithms*, S. Forrest, Ed., San Mateo, CA, 1993, pp. 124–131, Morgan Kaufmann Publishers.

[12] Wojciech Wieczorek and Zbigniew J. Czech, "Selection schemes in evolutionary algorithms," in *Intelligent Information Systems*, Mieczyslaw A. Klopotek, Slawomir T. Wierzchon, and Maciej Michalewicz, Eds. 2002, Advances in Soft Computing, pp. 185–194, Physica-Verlag.

[13] Ambedkar Dukkipati, M. Narasimha Murty, and Shalabh Bhatnagar, "Quotient evolutionary space: Abstraction of evolutionary process w.r.t macroscopic properties," in *Proceedings of the Congress on Evolutionary Computation(CEC'2003)*. 2003, IEEE Press.

[14] H. K. Kesavan and J. N. Kapur, *Entropy Optimization Principles with Applications*, Academic Press, 1997.

[15] David E. Goldberg and Kalyanmoy Deb, "A comparative analysis of selection shemes used in genetic algorithms," in *Foundations of Genetic Algorithms*, G. Rawlins, Ed., San Mateo, California, 1991, pp. 69–93, Morgan Kaufmann Publishers.

[16] Heinz Mühlenbein and Dirk Schlierkamp-Voosen, "Predictive models for the breeder genetic algorithm," *Evolutionary Computation*, vol. 1, no. 1, pp. 25–49, 1993.

[17] Walter Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 1964.