Threshold Selecting: Best Possible Probability Distribution for Crossover Selection in Genetic Algorithms

Jörg Lässig
Chemnitz University of Techn.
Reichenhainer Str. 70, R. A209
D-09126 Chemnitz, Germany
joerg.laessig@cs.tu-chemnitz.de

Karl Heinz Hoffmann
Chemnitz University of Techn.
Reichenhainer Str. 70, R. 356
D-09126 Chemnitz, Germany
hoffmann@physik.tu-chemnitz.de

Mihaela Enăchescu
Stanford University
Terman Eng. Center, R. 323
Stanford, CA 94305-4036,
United States
mihaela@cs.stanford.edu

ABSTRACT

The paper considers the problem of selecting individuals in the current population in Genetic Algorithms for crossover to find a solution of high fitness of a given combinatorial optimization problem.

Many different schemes have been considered in literature as possible crossover selection strategies, such as *Windowing*, *Exponential reduction*, *Linear transformation* or *normalization* and *Binary tournament selection*.

It is shown that if one wishes to maximize any linear function of the final state probabilities, e.g. the fitness of the best individual of the final population of the algorithm, then the best probability distribution for selecting individuals in each generation is a rectangular distribution over the individuals sorted by their fitness values.

This means uniform probabilities have to be assigned to a group of the best individuals of the population but probabilities equal to zero to individuals with fitness ranks higher than a fixed cutoff, which is equal to a certain rank in the sorted fitness vector. The considered strategy is called *Threshold Selecting*.

The proof applies basic arguments of MARKOV chains and linear optimization and makes only a few assumptions on the underlying principles and hence applies to a large class of Genetic Algorithms.

Track: Genetic Algorithms.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*heuristic methods*

General Terms

Algorithms, Theory

Keywords

Genetic Algorithms, Crossover Selection, MARKOV Process, Master Equation, Threshold Selecting

GECCO'08, July 12–16, 2008, Atlanta, Georgia, USA.
ACM 978-1-60558-131-6/08/07.

1. INTRODUCTION

Designing a *Genetic Algorithm* (GA) for a certain given problem, there are many degrees of freedom to be fixed but often the choice of certain parameters or operators relies on experimental tests and the experience of the programmer.

Such choices are e.g.:

- *representation of a solution* in the state space as an artificial genome,
- choice of a *crossover operator* to form a new population in each iteration,
- choice of a *mutation rate*,
- choice of a *selection scheme* over the individuals of a population for crossover.

Today GAs are in broad practical application to problems in many different fields as science, engineering or economics (see e.g. [2, 7, 12, 14, 16, 18]) and excellent experimental results have been obtained. Despite interesting theoretical progress in the last years [3, 5, 6, 4, 19, 22] exact proves for optimal choices of design criteria are still missing.

This paper focuses on the last of the design criteria above, also called *parent selection*. In all variants of GAs some form of the selection operator must be present [3]. A wide variety of selection strategies have been proposed in the literature. In general, m individuals of the current population of size n have to be selected for crossover into a *mating pool*. Individuals with higher fitness are more likely to receive more than one copy and less fit individuals are more likely to receive no copies. In different replacement schemes the size of the pool differs. After selecting the mating pool some crossover scheme takes individuals from that pool and produces new outcome, until the pool is exhausted. Regarding the crossover scheme no further restrictions are necessary for our considerations concerning the optimal choice of a selection strategy as discussed in the sections below.

The behavior of the GA very much depends on how individuals are chosen to go into the mating pool [20]. The simplest approach is that the reproduction probability of an individual of the population is proportional directly to its fitness (*roulette-wheel selection*). Other approaches are *windowing*, where first the fitness of the worst individual is subtracted from each individual fitness, *exponential*, where the square root of one plus the fitness is taken, *linear transformation*, where a linear function of the fitness is computed, e.g. $f' = \varrho \cdot f + \varphi$, *linear ranking selection*, where a linear function over a fitness ranking of the individuals is applied,

and *binary tournament selection*, where two individuals are selected with uniform probability in a preselection and the individual with better fitness is then submitted to the mating pool. See e.g. [3, 6, 20] for an overview on different selection schemes.

Each of these choices is reported to have strengths and weaknesses. The selection strategy has to be chosen such that the population evolves towards "better" overall fitness. For example, the fitness of the fittest individual in the final population might be required to be as high as possible.

In the following it is proven that *Threshold Selecting* is optimal in a certain sense defined below. In *Threshold Selecting* the selection is based on fitness ranks, and the selection probability on the ranks is rectangular, i.e. it includes one or more individual(s) with the highest fitness value(s) with the same non-vanishing probability but introduces a cutoff rank γ so that all individuals with higher ranks are selected with probability zero.

Table 1 [20] gives an example on the different methods for a population of four individuals with exemplary fitness values 50, 25, 15, 10.

Table 1: Comparison of different selection strategies

Rank of the individuals	1	2	3	4
Rough fitness	50.0	25.0	15.0	10.0
Roulette-wheel	0.5	0.25	0.15	0.1
Windowing	$0.\overline{6}$	0.25	$0.08\overline{3}$	0.0
Exponential	0.365	0.261	0.205	0.169
Linear transformation $(2f+1)$	0.495	0.25	0.152	0.103
Linear ranking selection	0.4	0.3	0.2	0.1
Binary tournament selection	0.438	0.312	0.188	0.062
Threshold Selecting ($\gamma = 3$)	$0.\overline{3}$	$0.\overline{3}$	$0.\overline{3}$	0.0

2. TECHNIQUE

The proof technique is based on the fact that the selection probability distributions assigning probabilities to n ordered objects can be seen as vectors in a n dimensional space. Special assumptions of the problem structure restrict the space of possible solutions to a simplex. Then, due do the linearity of applicable objective functions on the selection probabilities, the problem reduces to the task to find the minimum of a linear function on a simplex, which must be a vertex in the general case by the fundamental theorem of linear programming. As will be shown, the vertices are exactly equivalent to the rectangular distributions mentioned above.

This technique has been already applied to show for the acceptance rule in Monte Carlo methods as *Simulated Annealing* [17], *Threshold Accepting* [8] or TSALLIS *Statistics* [9, 10, 23] that Threshold Accepting is provably a best possible choice [11].

Further the stochastic optimization algorithm *Extremal Optimization* [1] has been investigated [13, 15]. Extremal Optimization also works by simulating random walkers as the methods mentioned before, but needs a special structure of the problem under consideration: every state is specified by several degrees of freedom, each of which can be assigned a fitness. Each iteration chooses one degree of freedom for change based on fitness values. It has been shown that a rectangular distribution is the best choice in each iteration of Extremal Optimization.

3. DEFINITIONS

We consider combinatorial optimization problems with a finite state space Ω of states $\alpha \in \Omega$, which are the possible solutions for the problem. A fitness function $f(\alpha)$ describes the quality of the solution α and has to be maximized, i.e. the states with a higher fitness are better. Note that there is only a finite number of possible values for $f(\alpha)$.

GAs consider populations (or pools) of states. If there are n states in a population, then each generation of the GA is equivalent to a generalized state $\boldsymbol{\alpha} := (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \Omega^n = \boldsymbol{\Omega}$ with n finite. A generalized fitness function $f(\boldsymbol{\alpha})$ has to be defined as well, which is usually done by $f(\boldsymbol{\alpha}) := \max\{f(\alpha_i) \mid i = 1, 2, \ldots, n\}$.

To obtain good solutions GAs proceed by randomly selecting a start population, and then evolving it by a selection and subsequent crossover operation. Mutations are also possible, but are of no importance here. We here confine ourselves to selection steps, where the probability to enter the mating pool is based on the fitness ranks of the population members. The possible mating pools are again described by generalized states $\bar{\boldsymbol{\gamma}} := (\gamma_1, \gamma_2, \ldots, \gamma_m)$, albeit not of size n but of size m. The bar notation is used to differentiate between the population and the mating pool.

For the choice of the m individuals for the crossover step in the GA, m time dependent probability distributions $d^{i,t}(k)$, $i = 1, 2, \ldots, m$ are defined over the ranks k. Given this structure at time t exactly m ranks $k_{l_1}, k_{l_2}, \ldots, k_{l_m}$ are chosen by the GA and hence m individuals from the current population according to $d^{i,t}, i = 1, 2, \ldots, m$. Technically, each of the individual members β_i of the current population $\boldsymbol{\beta}$ is assigned a rank k_i based on its fitness: The individuals of a population can be ordered according to their fitness in a ranking $k_i \in \mathbb{N}_n^\star = \{1, 2, \ldots, n\}$:

$$k_i \leq k_j \Longleftrightarrow f(\alpha_i) \geq f(\alpha_j) \;\; \forall \text{ pairs } (i, j) \; .$$

The following assumptions are adopted for the selection probabilities $d^{i,t}(k)$:

(A1) Each step of the algorithm is independent of the former steps.

(A2) In each step t, $1 \geq d^{i,t}(1) \geq d^{i,t}(2) \geq \cdots \geq d^{i,t}(n) \geq 0$ holds for $i = 1, 2, \ldots, m$, i.e. it is more probable to recombine individuals with lower rank (higher fitness) than individuals with a higher rank (lower fitness).

(A3) $\sum_{k=1}^{n} d^{i,t}(k) = 1$ for $i = 1, 2, \ldots m$, i.e. the distributions are normalized.

Due to the random nature of the selection process there is a transition probability

$$\Lambda^S_{\bar{\boldsymbol{\gamma}}\boldsymbol{\beta}} = d^{1,t}(k_{l_1}) d^{2,t}(k_{l_2}) \cdots d^{m,t}(k_{l_m}) \tag{1}$$

to obtain the mating pool $\bar{\boldsymbol{\gamma}} = (\beta_{l_1}, \beta_{l_2}, \ldots, \beta_{l_m})$ from the population $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n)$.

In the crossover step an operator $\mathcal{C}_{\bar{\boldsymbol{\gamma}}}$ is applied to the current population $\boldsymbol{\beta}$. The operator $\mathcal{C}_{\bar{\boldsymbol{\gamma}}}$ is not deterministic but determines the fixed probabilities $\Lambda^C_{\boldsymbol{\alpha}\bar{\boldsymbol{\gamma}}\boldsymbol{\beta}}$ to obtain a new population $\boldsymbol{\alpha} \in \boldsymbol{\Omega}$ from $\boldsymbol{\beta} \in \boldsymbol{\Omega}$ and $\bar{\boldsymbol{\gamma}}$ as intermediate step. For each fixed pair $\bar{\boldsymbol{\gamma}}$ and $\boldsymbol{\beta}$ we have

$$\sum_{\boldsymbol{\alpha} \in \boldsymbol{\Omega}} \Lambda^C_{\boldsymbol{\alpha}\bar{\boldsymbol{\gamma}}\boldsymbol{\beta}} = 1.$$

The dependence on both, $\bar{\boldsymbol{\gamma}}$ and $\boldsymbol{\beta}$, can be explained by the fact that not each crossover operator creates the new

population $\boldsymbol{\alpha}$ by solely utilizing the mating pool $\bar{\gamma}$, which is only the case for the *Generation Replacement Model*. In the *Steady-State Replacement Model* it is also possible that other states β_i from the current population $\boldsymbol{\beta}$ are kept.

An exemplary procedure could work as follows: After getting a mating pool $\bar{\gamma}$ of m states m new states are created by recombination from $\bar{\gamma}$. In addition to the current n states then there are $n+m$ states available and n states are kept for the new generation $\boldsymbol{\alpha}$ applying some standard procedure (e.g. keep the best n of all $n+m$ states). In the special case of generation replacement we have $n = m$ and $\boldsymbol{\beta}$ is replaced completely by the states from recombination.

Remark 1. Think of the recombination itself that states from the mating pool $\bar{\gamma}$ are taken one after another. Each possible tuple of states for one crossover operation is chosen with the same probability, i.e. probability is uniform distributed among all possible tuples of desired size in $\bar{\gamma}$. Most commonly pairs are chosen and for each pair a split position for one point crossover or more than one split position for multi point crossover procedures are determined - again uniform distributed (the proof is general enough that also other distributions or procedures are possible here).

Combining selection and crossover leads to a transition probability $\Gamma^t_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ from one population $\boldsymbol{\beta}$ to the next population $\boldsymbol{\alpha}$.

In summary the dynamics of GAs can be described as a MARKOVian random walk in state space. For the development of the probability $p^t_{\boldsymbol{\alpha}}$ to be in state $\boldsymbol{\alpha}$ (which means to have a certain population in the GA) the master equation

$$p^t_{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\beta}\in\boldsymbol{\Omega}} \Gamma^t_{\boldsymbol{\alpha}\boldsymbol{\beta}} \cdot p^{t-1}_{\boldsymbol{\beta}} \tag{2}$$

is applicable. Here $\Gamma^t_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ is defined to be

$$\Gamma^t_{\boldsymbol{\alpha}\boldsymbol{\beta}} = \sum_{\bar{\gamma}\in\bar{\Omega}} \Lambda^C_{\boldsymbol{\alpha}\bar{\gamma}\boldsymbol{\beta}} \Lambda^S_{\bar{\gamma}\boldsymbol{\beta}}. = \sum_{\bar{\gamma}\in\bar{\Omega}} \Lambda^C_{\boldsymbol{\alpha}\bar{\gamma}\boldsymbol{\beta}} \prod_{i=1}^{m} d^{i,t}. \tag{3}$$

In the next step the dependence of the performance of the GA on the probability distributions $d^{1,t}, d^{2,t}, \ldots, d^{m,t}$ over the ranks in the population is investigated and which choice is an optimal one for these distributions considering an optimization run with S steps.

Most commonly one of the following objectives is used [11] (here slightly adapted in the notation for GAs):

(O1) The mean of the fitness of the best individual in the final population should be as large as possible.

(O2) The probability of having a final population containing a member of optimal fitness should be as large as possible.

To optimize according to (O1) one chooses

$$g_1(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}) = \max\{f(\alpha_i) \mid i = 1, 2, \ldots, n\}$$

which means essentially that the quality of a population is assumed to be equivalent to the quality of the best individual in the population, and to optimize according to (O2) one chooses

$$g_2(\boldsymbol{\alpha}) \quad = \quad \begin{cases} 1 & \mid \text{ if } \boldsymbol{\alpha} \text{ contains a state with fitness } f_{\max} \\ 0 & \mid \text{ otherwise,} \end{cases}$$

i.e. only optimal states with fitness $f_{\max}$ have objective values different from zero. Other objectives are possible and the only important fact for the proof is that they are linear in the final state probabilities, as we will see.

The optimization process consists of a finite number of S steps ($t := 1, 2, \ldots, S$). Note, that $\Gamma^t_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ is linear in $d^{i,t}(k)$ for i fixed. The arguments below apply in general to any objective function which is linear in the final state probabilities $p^S_{\boldsymbol{\alpha}}$ as e.g. (O1) and (O2). The state probabilities at time t are considered as vector $\boldsymbol{p}^t$ and the linear objective function with values $g(\boldsymbol{\alpha})$ for each state $\boldsymbol{\alpha}$ as vector $\boldsymbol{g}$. If $(\cdot)^{\mathrm{tr}}$ denotes the transpose, the measure of performance is equivalent to

$$g(\boldsymbol{p}^S) = \boldsymbol{g}^{\mathrm{tr}} \cdot \boldsymbol{p}^S = \sum_{\boldsymbol{\alpha}\in\boldsymbol{\Omega}} g(\boldsymbol{\alpha}) \cdot p^S_{\boldsymbol{\alpha}} \longrightarrow \max \ . \tag{4}$$

4. SETUP OF A VECTOR SPACE

In the following the distributions $d^{i,t}(k), k = 1, 2, \ldots, n$, are considered to be n dimensional vectors $\boldsymbol{d}^{i,t}$ with entries $\boldsymbol{d}^{i,t}_k \in [0, 1]$. Consider without loss of generality $n-1$ of these distributions $\boldsymbol{d}^{i,t}, i \in 1, 2, \ldots, n$ to be fixed. Only one remaining distribution denoted by $\boldsymbol{d}^{r,t}$ is open to optimize. The question is then how to choose $\boldsymbol{d}^{r,t}$ to maximize the objective function. As a consequence of the assumptions (A2) and (A3), the region $\mathcal{F}$ of feasible vectors $\boldsymbol{d}^{r,t}$ is defined by the $n+1$ linear inequations in (A2) and one linear equation in (A3) where the first inequation $1 \geq \boldsymbol{d}^{r,t}_1$ follows from the others. Of the remaining n inequations $n-1$ must be set to equations to find extreme points (vertices) in the region $\mathcal{F}$. Letting V denote the set of extreme points of $\mathcal{F}$, the elements of V are exactly those vectors $\boldsymbol{d}^{r,t}$ which have the initial sequence of i entries equal to $1/i$ followed by a sequence of $n-i$ entries equal to zeros. Explicitly, $V = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$, where $\boldsymbol{v}_1 = (1, 0, 0, \ldots, 0)^{\mathrm{tr}}$, $\boldsymbol{v}_2 = (1/2, 1/2, 0, 0, \ldots, 0)^{\mathrm{tr}}$, $\boldsymbol{v}_i = (1/i, 1/i, \ldots, 1/i, 0, 0, \ldots, 0)^{\mathrm{tr}}$, and $\boldsymbol{v}_n = (1/n, 1/n, \ldots, 1/n)^{\mathrm{tr}}$. Note that the elements of V are linearly independent. Then $\mathcal{F}$ is exactly the convex hull $C(V)$ of V, which is a simplex.

This equivalence of $\mathcal{F}$ and $C(V)$ can be shown by standard calculations, see e.g. [13] as reference.

5. PROOF

Now the Bellman principle of dynamic programming is applied, starting with the last step in the optimization process. The output of the last step is $\boldsymbol{p}^S$ and used to determine the optimality criterion (4). In the last step S one has to solve the optimization problem (4) for the given input $\boldsymbol{p}^{S-1}$. Using (2) one gets

$$g(\boldsymbol{p}^S) = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\boldsymbol{\Omega}} g(\boldsymbol{\alpha}) \cdot \Gamma^S_{\boldsymbol{\alpha}\boldsymbol{\beta}} \cdot p^{S-1}_{\boldsymbol{\beta}} \longrightarrow \max$$

with $\Gamma^S_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ given by (3). Hence, the maximum of a linear function on a simplex has to be found. To do so choose the distribution $\boldsymbol{d}^{r,S}$, which selects one of the m individuals for crossover, equal to one of the vertices $v_i \in V$. The corresponding transition probabilities are denoted by $\boldsymbol{\Gamma}^S$, because then all m distributions are fixed in this stage. Considering now the step before, i.e. step $S-1$, one gets

$$\boldsymbol{g}^{\mathrm{tr}} \cdot \boldsymbol{p}^S = (\boldsymbol{g}^{\mathrm{tr}} \cdot \boldsymbol{\Gamma}^S) \cdot \left(\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\boldsymbol{\Omega}} \Gamma^{S-1}_{\boldsymbol{\alpha}\boldsymbol{\beta}} \cdot p^{S-2}_{\boldsymbol{\beta}} \right) \longrightarrow \max \ .$$

Defining $g^{S-1} = \boldsymbol{g}^{\mathrm{tr}} \cdot \boldsymbol{\Gamma}^S$ as new objective function the same arguments can be applied to choose $\boldsymbol{d}^{r,S-1}$. The resulting matrix is then denoted with $\boldsymbol{\Gamma}^{S-1}$ where the optimal transition probabilities are again found by taking $\boldsymbol{d}^{r,S-1}$ to be an element of V. For all other steps the same argument holds as well, i.e. $\boldsymbol{d}^{r,t}, t = 1, 2, \ldots, S$, are all elements of the vertex set V.

Because an arbitrary distribution $\boldsymbol{d}^{r,t}$ has been chosen for the proof to be variable, the same arguments hold for all distributions $\boldsymbol{d}^{r,t}, r = 1, 2, \ldots, n$ as well. Hence, the proof shows that a rectangular distribution over the fittest individuals in each generation in the iterations $t = 1, 2, \ldots, S$ in GAs gives the best implementation of the selection step for each individual used for the crossover step in iteration t. For the mutation operator the same as for crossover holds, because this is an operator with equivalent characteristics with regard to our proof but only one input state.

6. CONCLUSIONS

In this paper the problem of selecting individuals from the population of a GA for crossover based on a fitness function has been considered. The master equation was the means of choice to describe the corresponding dynamics as a random walk in state space and some straightforward assumptions on the probability distributions for selecting the individuals in a certain generation have been formulated.

The goal was to find transition probabilities assuring the optimum control of the evolutionary development in the GA. A rectangular distribution of selection probabilities is provably optimal, provided the performance is measured by a linear function in the state probabilities, which includes many reasonable choices as for instance maximizing the mean fitness of the best individual in the final population.

The proof above is based on the fundamental theorem of linear programming, which states that a linear function defined on a simplex assumes its minimum at a vertex. The proof does not state that all optimal crossover selection strategies in GAs are rectangular. Other strategies may do equally well, but not better. If there exists an optimal strategy other than *Threshold Selecting*, it follows that an edge or a face of the described simplex does equally well. Thus, it seems unlikely that a strictly monotonic distribution can be optimal [11], which would imply that all the vertices in V do equally well.

As presented the proof can be applied for any crossover procedure in GAs with independent probability distributions for the selection of the crossover individuals and both for the *Generation Replacement Model*, where the mating pool has size n for populations of size n, and also for the *Steady-State Replacement Model*, where only some individuals are replaced [21].

Currently the knowledge that best performance can be achieved using *Threshold Selecting* is only of limited use, since the cutoff ranks γ to be used are not known a priori. Therefore it would be interesting to perform numerical experiments comparing different possible distributions empirically. Further, it is reasonable to introduce a schedule on the cutoff rank γ, narrowing the rectangular distribution during the optimization process and thus increasing the evolutionary pressure gradually. Moreover, it would be interesting to obtain also theoretical progress concerning the choice of one of the possible rectangular distributions or to reduce the choice to a certain assortment.

Our proof was based on the assumption that the objective measuring the performance of the GA is a linear function of the state probabilities. While this includes very common measures, it does not include them all.

As an example, the *best-so-far fitness* over individuals of all previous generations as a measure is beyond the scope of the proof presented here. From a practical point of view this can be fixed easily, adding one individual to the population and adapting the crossover operator in a way to keep the best individual in each iteration, if not one with better or equal fitness is found.

Clearly this adaption is possible for each given crossover operator $\mathcal{C}$ to obtain an operator $\mathcal{C}'$. Applying $\mathcal{C}'$ the objective value of the individual with best fitness in the final population is equivalent to the *best-so-far fitness* but the distributions are as well rectangular because the proof applies as before.

Further, the proof above had to assume a finite state space. The exploration of continuous state spaces would be interesting as well, but considering the discrete arithmetic of digital computers any state space in practice is effectively finite [13].

Finally, the arguments presented here establish the structure of a provably optimal strategy which could be applied to study also other heuristic approaches to global optimization.

7. ACKNOWLEDGMENTS

The authors would like to thank the *German Research Foundation* (DFG) for partially funding this research.

8. REFERENCES

[1] S. Boettcher and A. G. Percus. Extremal Optimization: Methods derived from Co-Evolution. In *GECCO-99, Proceedings of the Genetic and Evolutionary Computation Conference*, pages 825–832, Orlando, Florida, July 1999.

[2] P. G. Busacca, M. Marseguerra, and E. Zio. Multiobjective Optimization by Genetic Algorithms: Application to Safety Systems. *Reliability Engineering & System Safety*, 72(1):59–74, April 2001.

[3] M. Chakraborty and U. K. Chakraborty. An Analysis of Linear Ranking and Binary Tournament Selection in Genetic Algorithms. In *Proceedings of the International Conference on Information, Communications and Signal Processing*, pages 407–411, Singapore, September 1997.

[4] U. K. Chakraborty. A Simpler Derivation of Schema Hazard in Genetic Algorithms. *Information Processing Letters*, 56(2):77–78, 1995.

[5] U. K. Chakraborty. Abranching Process Model for Genetic Algorithms. *Information Processing Letters*, 56(5):281–292, December 1995.

[6] U. K. Chakraborty, K. Deb, and M. Chakraborty. Analysis of Selection Algorithms: A Markov Chain Approach. *Evolutionary Computation*, 4(2):133–167, 1997.

[7] L. D. Chambers. *The Practical Handbook of Genetic Algorithms: Applications*. CRC Press Inc., 2000.

[8] G. Dueck and T. Scheuer. Threshold Accepting: A General Purpose Optimization Algorithm Appearing

Superior to Simulated Annealing. *Journal of Computational Physics*, 90:161–175, 1990.

[9] A. Franz and K. H. Hoffmann. Optimal Annealing Schedules for a Modified Tsallis Statistics. *Jounal of Computational Physics*, 176(1):196–204, Feb. 2002.

[10] A. Franz and K. H. Hoffmann. Threshold Accepting as Limit Case for a Modified Tsallis Statistics. *Applied Mathematics Letters*, 16(1):27–31, January 2003.

[11] A. Franz, K. H. Hoffmann, and P. Salamon. Best Possible Strategy for Finding Ground States. *Physical Review Letters*, 86(23):5219–5222, June 2001.

[12] P. Godefroid and S. Khurshid. Exploring Very Large State Spaces Using Genetic Algorithms. In J.-P. Katoen and P. Stevens, editors, *LNCS 2280*, pages 266–280. Springer-Verlag Berlin Heidelberg, 2002.

[13] F. Heilmann, K. H. Hoffmann, and P. Salamon. Best Possible Probability Distribution over Extremal Optimization Ranks. *Europhysics Letters*, 66(3):305–310, March 2004.

[14] B. Hemmateenejad, M. Akhond, R. Miri, and M. Shamsipur. Genetic Algorithm Applied to the Selection of Factors in Principal Component-Artificial Neural Networks: Application to QSAR Study of Calcium Channel Antagonist Activity of 1,4-Dihydropyridines (Nifedipine Analogous). *Journal of Chemical Information and Modeling*, 43(4):1328 –1334, June 2003.

[15] K. H. Hoffmann, F. Heilmann, and P. Salamon. Fitness Threshold Accepting over Extremal Optimization Ranks. *Physical Review E*, 70(4):046704, October 2004.

[16] S. Kikuchi, D. Tominaga, M. Arita, K. Takahashi, and M. Tomita. Dynamic Modeling of Genetic Networks Using Genetic Algorithm and S-System. *Bioinformatics*, 19(5):643–650, 2003.

[17] S. Kirkpatrick, C. Galatt, and M. Vecchi. Optimization by simulated annealing. *Science*, 4598, 1983.

[18] A. Kolen. A Genetic Algorithm for the Partial Binary Constraint Satisfaction Problem: an Application to a Frequency Assignment Problem. *Statistica Neerlandica*, 61(1):4–15, February 2007.

[19] H. Mühlenbein and D. Schlierkamp-Voosen. The Science of Breeding and its Application to the Breeder Genetic Algorithm (BGA). *Evolutionary Computation*, 1(4):335–360, 1993.

[20] J.-P. Rennard. Introduction to Genetic Algorithms. http://www.rennard.org/alife/english/gavgb.pdf, 2000. [Online; accessed 29-February-2008].

[21] D. Srinivasan and L. Rachmawati. An Efficient Multi-objective Evolutionary Algorithm with Steady-State Replacement Model. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pages 715–722, Seattle, Washington, USA, July 2006.

[22] C. R. Stephens, M. Toussaint, D. Whitley, and P. F. Stadler, editors. *Foundations of Genetic Algorithms: 9th International Workshop, FOGA 2007*, Mexico City, Mexico, 2007. Springer Berlin.

[23] C. Tsallis and D. A. Stariolo. Generalized Simulated Annealing. *Physica A*, 233:395–406, 1996.