Theory of the Simple Genetic Algorithm with α-Selection

André Neubauer
Information Processing Systems Lab
Electrical Engineering and Computer Science
Münster University of Applied Sciences
Stegerwaldstraße 39, D-48565 Steinfurt, Germany
andre.neubauer@fh-muenster.de

ABSTRACT

Genetic algorithms are random heuristic search (RHS) algorithms with a wide range of applications in adaptation and optimisation problems. The most advanced approach for a general theory of genetic algorithms is offered by the dynamical system model which describes the stochastic trajectory of a population under the dynamics of a genetic algorithm with the help of an underlying deterministic heuristic function and its fixed points. However, even for the simple genetic algorithm (SGA) with fitness-proportional selection, crossover and mutation the determination of the population trajectory and the fixed points of the heuristic function is unfeasible for practical problem sizes. In order to simplify the mathematical analysis α-selection is introduced in this paper. Based on this selection scheme it is possible to derive the dynamical system model and the fixed points in closed form. Although the heuristic function is not compatible with the equivalence relation imposed by schemata in the strict sense a simple coarse-grained system model with a single exogenous parameter is derivable for a given schemata family. In addition to the theoretical analysis experimental results are presented which confirm the theoretical predictions.

Categories and Subject Descriptors

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

General Terms

Algorithms

Keywords

Genetic algorithm, dynamical system model, random heuristic search, α-selection

1. INTRODUCTION

Genetic algorithms mimic the basic mechanisms of biological evolution and molecular genetics in simplified form.

These RHS algorithms are based on populations of individual solutions which evolve according to selection and genetic operators like crossover and mutation. Since they were proposed by HOLLAND [1] genetic algorithms have been used in a wide range of applications, e.g. for search as well as numerical and combinatorial optimisation problems [2]. Despite their successful applications genetic algorithms still lack a general and stringent theory which guides the practitioner in tailoring the algorithm to a specific application domain [3]. The first approach to theoretically describe the functioning of the SGA was the so-called *schema theorem* [1] which gives a lower bound for the expected number of individuals representing particular hyperplanes or *schemata* in the search space within the next generation. In recent years the schema theory and its implications like the so-called *building block hypothesis* have been criticised [3, 7]. As an alternative a dynamical system model was introduced for the SGA by VOSE et al. [8, 5, 7, 6]. In this model the SGA's stochastic dynamics is modeled according to an underlying deterministic heuristic function which describes the expected next population of the SGA given the current population. According to theoretical and empirical evidence the population trajectory is attracted by the fixed points of the heuristic function. However, even for moderate problem sizes the determination of the fixed points becomes unfeasible. Therefore, in order to simplify the mathematical analysis we introduce the genetic algorithm with α-selection in this paper.

The paper is organised as follows. After a brief description of the SGA over search spaces of binary ℓ-tuples in section 2, the dynamical system model is summarised in section 3. The underlying heuristic function is formulated for the SGA with fitness-proportional selection, 1-point or uniform crossover and bitwise mutation. The genetic algorithm with α-selection is introduced in section 4 based on the notion of the best individual randomly mating with other individuals in the current population. The dynamical system model of a genetic algorithm with α-selection is derived in section 5. The fixed points of the corresponding heuristic function are calculated analytically and compared to simulation results showing a close agreement between theory and experiment. As is well known the heuristic function is not compatible with the equivalence relation imposed by schemata due to the selection operation. Although this fact still holds for a genetic algorithm with α-selection a simplified model for a given schemata family with a single exogenous parameter can be derived as shown in section 6. Further, projections onto the space of schemata are discussed. A brief conclusion is given in section 7.

2. SIMPLE GENETIC ALGORITHM

In this section the SGA with fitness-proportional selection, 1-point or uniform crossover and bitwise mutation [7] is described which is used for the maximisation of a fitness function f. This fitness function f is defined over the search space $\Omega = \mathbb{Z}_2^\ell = \{0, 1\}^\ell$ consisting of binary ℓ-tuples $a = (a_0, a_1, \ldots, a_{\ell-1})$, i.e. $f : \Omega \to \mathbb{R}$.

2.1 Algorithm

The SGA works over populations $P(t)$ defined as multisets $P(t) = (a_0(t), a_1(t), \ldots, a_{r-1}(t))$ consisting of r individual binary ℓ-tuples $a_k(t) = (a_{k,0}(t), a_{k,1}(t), \ldots, a_{k,\ell-1}(t)) \in \Omega$ with fitness values $f(a_k(t))$. For the creation of offspring individuals in each generation t genetic operators like crossover χ_Ω and mutation μ_Ω are applied to parental individuals which are selected according to their fitness values as shown in Fig. 1. The population $P(0)$ is initialised appropriately, e.g. by randomly choosing individuals in Ω.

```
t := 0;
initialise population P(0);
while end of adaptation ≠ true do
    for k = 0 to r − 1 do
        select parental ℓ-tuples b(t) and c(t);
        apply crossover χΩ and mutation μΩ
            a_k(t + 1) := μΩ (χΩ (b(t), c(t)));
        evaluate fitness f (a_k(t + 1));
    end
    t := t + 1;
end
```

Figure 1: Simple genetic algorithm (SGA).

2.2 Selection

For fitness-proportional selection each individual $a_k(t)$ in population $P(t)$ is selected with a probability $p_k(t)$ which is directly proportional to its (positive) fitness $f(a_k(t)) > 0$. The individual selection probability is then given by

$$p_k(t) = \frac{f(a_k(t))}{\sum_{j=0}^{r-1} f(a_j(t))} \ . \tag{1}$$

2.3 Crossover

The crossover operator $\chi_\Omega : \Omega \times \Omega \to \Omega$ takes two selected individual ℓ-tuples $a = (a_0, a_1, \ldots, a_{\ell-1})$ and $b = (b_0, b_1, \ldots, b_{\ell-1})$ and randomly generates an offspring ℓ-tuple $c = (c_0, c_1, \ldots, c_{\ell-1})$ according to $c = \chi_\Omega(a, b)$.

For *1-point crossover* a position $1 \leq \lambda \leq \ell - 1$ is randomly chosen with uniform probability and the two ℓ-tuples $(a_0, \ldots, a_{\lambda-1}, b_\lambda, \ldots, b_{\ell-1})$ and $(b_0, \ldots, b_{\lambda-1}, a_\lambda, \ldots, a_{\ell-1})$ are generated one of which is chosen as offspring c with equal probabilities. The 1-point crossover operator is applied with crossover probability χ.

In case of *uniform crossover* each bit of the offspring c is chosen with probability χ from the ℓ-tuple a and with probability $1 - \chi$ from the ℓ-tuple b or vice versa.

2.4 Mutation

In the SGA with binary ℓ-tuples the bitwise mutation operator $\mu_\Omega : \Omega \to \Omega$ is defined by randomly flipping each bit of the individual ℓ-tuple $a = (a_0, a_1, \ldots, a_{\ell-1})$ with small mutation probability μ. A typical value is $\mu \sim \frac{1}{\ell}$.

3. DYNAMICAL SYSTEM MODEL OF THE SIMPLE GENETIC ALGORITHM

Following [7] the dynamical system model for the SGA is briefly summarised in this section. In view of the search space Ω of binary ℓ-tuples we identify each binary ℓ-tuple $(a_0, a_1, \ldots, a_{\ell-1}) = a_0 a_1 \ldots a_{\ell-1}$ with the integer $a = a_0 \cdot 2^{\ell-1} + a_1 \cdot 2^{\ell-2} + \ldots + a_{\ell-2} \cdot 2^1 + a_{\ell-1} \cdot 2^0$. The respective fitness is given by $f_a = f(a)$. This leads to the search space $\Omega = \{0, 1, \ldots, n-1\}$ with cardinality $|\Omega| = n = 2^\ell$. Based on this binary number representation we define the bitwise modulo-2 addition $a \oplus b$, bitwise modulo-2 multiplication $a \otimes b$ and bitwise binary complement $\overline{a}$, e.g. $3 \oplus 5 = 011 \oplus 101 = 110 = 6$, $3 \otimes 5 = 011 \otimes 101 = 001 = 1$ and $\overline{3} = \overline{011} = 100 = 4$ for ℓ-tuples of length $\ell = 3$. Vice versa, the integer $a \in \Omega$ is also viewed as a column vector $(a_0, a_1, \ldots, a_{\ell-1})^{\mathrm{T}}$ so that we can define the scalar product $a^{\mathrm{T}} b = a_0 \cdot b_0 + a_1 \cdot b_1 + \ldots + a_{\ell-2} \cdot b_{\ell-2} + a_{\ell-1} \cdot b_{\ell-1}$. In the following we will also make use of the all-one ℓ-tuple $\mathbf{1}$ which corresponds to the integer $n - 1 = 2^\ell - 1$. Furthermore, the indicator function $[i = j] = \delta_{i,j}$ is defined by $[i = j] = 1$ if $i = j$ and 0 if $i \neq j$.

In the dynamical system model [7] the dynamics of the SGA is compactly formulated by defining the population vector $\boldsymbol{p} = (p_0, p_1, \ldots, p_{n-1})^{\mathrm{T}}$. Each component

$$p_i = \frac{1}{r} \sum_{j \in P} [j = i]$$

gives the proportion of the element $i \in \Omega$ in the current population P. The population vector $\boldsymbol{p}$ is an element of the simplex $\Lambda = \left\{ \boldsymbol{p} \in \mathbb{R}^n : p_i \geq 0 \wedge \sum_{i \in \Omega} p_i = 1 \right\}$. For a population of size r the number of possible population vectors is given by $\binom{n+r-1}{r}$. In the limit of infinite populations with $r \to \infty$ the population vectors are dense in the simplex Λ. For simplicity we will take the simplex Λ as the defining region of the population vector $\boldsymbol{p}$ which is strictly valid only for large populations with $r \gg 1$.

The SGA is now described as an instance of RHS $\tau : \Lambda \to \Lambda$ according to $\boldsymbol{p}(t + 1) = \tau(\boldsymbol{p}(t))$ with τ depending on the random selection and genetic operators. As outlined in [7] τ can be equivalently represented by a suitable heuristic function $\mathcal{G} : \Lambda \to \Lambda$ which for a given population vector $\boldsymbol{p}$ yields the probability distribution $\mathcal{G}(\boldsymbol{p})$. This probability distribution

$$\mathcal{G}(\boldsymbol{p})_i = \Pr\{\text{individual } i \text{ is sampled from } \Omega\} \tag{2}$$

is used to generate the next population as illustrated in Fig. 2. For the SGA the heuristic function $\mathcal{G} = \mathcal{M} \circ \mathcal{F}$ is given by the composition of the selection operation $\mathcal{F}$ and the mixing operation $\mathcal{M}$ comprising crossover χ_Ω and mutation μ_Ω. The transition probabilities of the RHS τ are given by the formula [7]

$$\Pr\{\tau(\boldsymbol{p}) = \boldsymbol{q}\} = r! \prod_{i \in \Omega} \frac{\mathcal{G}(\boldsymbol{p})_i^{r q_i}}{(r q_i)!} \ . \tag{3}$$

The trajectory $\boldsymbol{p}, \tau(\boldsymbol{p}), \tau^2(\boldsymbol{p}), \ldots$ approximately follows the trajectory $\boldsymbol{p}, \mathcal{G}(\boldsymbol{p}), \mathcal{G}^2(\boldsymbol{p}), \ldots$ of the deterministic dynamical system defined by the heuristic function $\mathcal{G}$ with

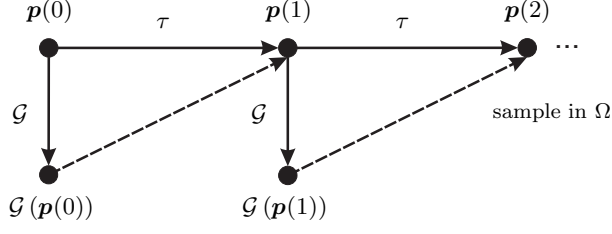$$\mathrm{E}\{\tau(\boldsymbol{p})\} = \mathcal{G}(\boldsymbol{p}) \ . \tag{4}$$

Figure 2: Genetic algorithm as RHS τ with heuristic function $\mathcal{G}$.

Because of the corresponding mean quadratic deviation

$$\mathrm{E}\{\|\tau\left(\boldsymbol{p}\right) - \mathcal{G}\left(\boldsymbol{p}\right)\|^2\} = \frac{1}{r} \cdot \left(1 - \|\mathcal{G}\left(\boldsymbol{p}\right)\|^2\right) \qquad (5)$$

the RHS τ behaves like the deterministic dynamical system model in the limit of infinite populations with $r \to \infty$. As illustrated by experimental evidence the RHS τ shows punctuated equilibria, i.e. phases of relative stability nearby a fixed point $\boldsymbol{\omega} = \mathcal{G}\left(\boldsymbol{\omega}\right)$ of the heuristic function $\mathcal{G}$ disrupted by sudden transitions to another dynamical equilibrium near another fixed point. We call this the *fixed point hypothesis* of genetic algorithms.

3.1 Selection

For fitness-proportional selection the respective heuristic function $\mathcal{F}\left(\boldsymbol{p}\right)_i = \Pr\{\text{individual } i \text{ selected}\}$ is given by [7]

$$\mathcal{F}\left(\boldsymbol{p}\right)_i = \frac{f_i \cdot p_i}{\sum_{j \in \Omega} f_j \cdot p_j} = \frac{f_i \cdot p_i}{\boldsymbol{f}^{\mathrm{T}} \boldsymbol{p}} \qquad (6)$$

with fitness vector $\boldsymbol{f} = (f_0, f_1, \ldots, f_{n-1})^{\mathrm{T}}$.

3.2 Crossover

For the crossover operator χ_Ω applied to the individual ℓ-tuples a and b with crossover probability χ we make use of a random crossover mask $m \in \Omega$ according to

$$\chi_\Omega(a, b) = a \otimes m \oplus \overline{m} \otimes b$$

or

$$\chi_\Omega(a, b) = a \otimes \overline{m} \oplus m \otimes b$$

each with probability $\frac{1}{2}$. For 1-point crossover the crossover mask m is randomly chosen from Ω according to the probability distribution vector $\boldsymbol{\chi} = (\chi_0, \chi_1, \ldots, \chi_{n-1})^{\mathrm{T}}$ with [7]

$$\chi_m = \begin{cases} 1 - \chi\ , & m = 0 \\ \dfrac{\chi}{\ell - 1}\ , & m = 2^\lambda - 1 \text{ with } 1 \le \lambda \le \ell - 1 \\ 0\ , & \text{otherwise} \end{cases} \ . \quad (7)$$

In case of uniform crossover the probability distribution vector $\boldsymbol{\chi} = (\chi_0, \chi_1, \ldots, \chi_{n-1})^{\mathrm{T}}$ is given by [7]

$$\chi_m = \begin{cases} 1 - \chi + \chi \cdot 2^{-\ell}\ , & m = 0 \\ \chi \cdot 2^{-\ell}\ , & m > 0 \end{cases} \ . \quad (8)$$

3.3 Mutation

The bitwise mutation operator μ_Ω can be described with the help of the random mutation mask $m \in \Omega$ according to $\mu_\Omega(a) = a \oplus m$. The mutation mask m is randomly chosen from Ω according to the probability distribution vector $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_{n-1})^{\mathrm{T}}$ with [7]

$$\mu_m = \mu^{\mathbf{1}^{\mathrm{T}} m} \cdot (1 - \mu)^{\ell - \mathbf{1}^{\mathrm{T}} m} \ . \qquad (9)$$

3.4 Heuristic

The mixing operation comprises crossover χ_Ω and mutation μ_Ω. With the help of the respective probability distributions for crossover and mutation we obtain

$$\begin{aligned} &\Pr\{\mu_\Omega\left(\chi_\Omega(a, b)\right) = c\} \\ &= \sum_{j \in \Omega} \mu_j \cdot \Pr\{\chi_\Omega(a, b) = c \oplus j\} \\ &= \sum_{j \in \Omega} \mu_j \sum_{i \in \Omega} \frac{\chi_i + \chi_{\overline{i}}}{2} \cdot [a \otimes i \oplus \overline{i} \otimes b = c \oplus j] \ . \end{aligned}$$

Taking into account

$$\Pr\{\mu_\Omega\left(\chi_\Omega(a, b)\right) = c\} = \Pr\{\mu_\Omega\left(\chi_\Omega(a \oplus c, b \oplus c)\right) = 0\}$$

this yields

$$\Pr\{\mu_\Omega\left(\chi_\Omega(a, b)\right) = c\} = M_{a \oplus c, b \oplus c}$$

with the $n \times n$ mixing matrix

$$M_{i,j} = \sum_{u,v \in \Omega} \mu_v \cdot \frac{\chi_u + \chi_{\overline{u}}}{2} \cdot [i \otimes u \oplus \overline{u} \otimes j = v] \ . \qquad (10)$$

The heuristic function $\mathcal{G}$ follows according to

$$\begin{aligned} \mathcal{G}\left(\boldsymbol{p}\right)_i &= \Pr\{\text{individual } i \text{ is sampled from } \Omega\} \\ &= \sum_{u,v \in \Omega} \mathcal{F}\left(\boldsymbol{p}\right)_u \cdot \mathcal{F}\left(\boldsymbol{p}\right)_v \cdot \Pr\{\mu_\Omega\left(\chi_\Omega(u, v)\right) = i\} \\ &= \sum_{u,v \in \Omega} \mathcal{F}\left(\boldsymbol{p}\right)_u \cdot \mathcal{F}\left(\boldsymbol{p}\right)_v \cdot M_{u \oplus i, v \oplus i} \\ &= \sum_{u,v \in \Omega} \mathcal{F}\left(\boldsymbol{p}\right)_{u \oplus i} \cdot \mathcal{F}\left(\boldsymbol{p}\right)_{v \oplus i} \cdot M_{u,v} \ . \qquad (11) \end{aligned}$$

By defining the mixing operation $\mathcal{M} : \Lambda \to \Lambda$ according to

$$\mathcal{M}\left(\boldsymbol{p}\right)_i = \sum_{u,v \in \Omega} p_{u \oplus i} \cdot p_{v \oplus i} \cdot M_{u,v} \qquad (12)$$

the heuristic function is given by the composition $\mathcal{G} = \mathcal{M} \circ \mathcal{F}$.

In order to determine the heuristic function $\mathcal{G}$ the mixing matrix M has to be calculated. This can be done efficiently with the help of the WALSH transform [9, 10]. For a matrix A the WALSH transform is $\widehat{A} = W \cdot A \cdot W$ with the $n \times n$ WALSH matrix $W_{i,j} = n^{-1/2} \cdot (-1)^{i^{\mathrm{T}} j}$. Correspondingly, the WALSH transform of a vector $\boldsymbol{v}$ yields $\widehat{\boldsymbol{v}} = W \cdot \boldsymbol{v}$. The WALSH matrix W is symmetric and orthogonal, i.e. $W^{-1} = W^{\mathrm{T}} = W$. In Fig. 3 the WALSH matrix is illustrated for $n = 2^6 = 64$.

For crossover and mutation performed after crossover the WALSH transform of the mixing matrix M is given by

$$\widehat{M}_{i,j} = \frac{n^{1/2}}{2} \cdot [i \otimes j = 0] \cdot \widehat{\mu}_{i \oplus j} \cdot \sum_{k \in \Omega_{\overline{i \otimes j}}} (\chi_{k \oplus i} + \chi_{k \oplus j})$$

with $\Omega_k = \{i \in \Omega : i \otimes \overline{k} = 0\}$. Due to the factor $[i \otimes j = 0]$ the components $\widehat{M}_{i,j}$ are nonzero only if $i \otimes j = 0$ or $j \in \Omega_{\overline{i}}$, respectively. With the WALSH transform of the mutation mask distribution [7]
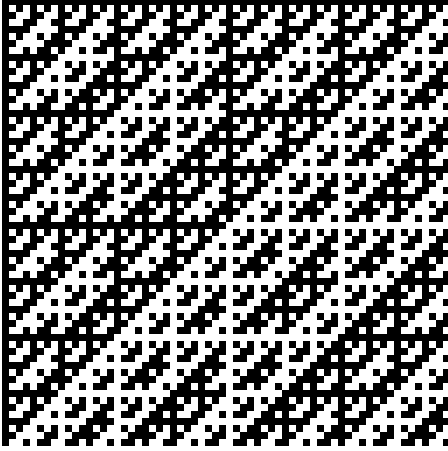
$$\widehat{\mu}_i = n^{-1/2} \cdot (1 - 2\mu)^{\mathbf{1}^{\mathrm{T}} i} \qquad (13)$$

Figure 3: Illustration of the WALSH **matrix** W **for** $n = 2^6 = 64$**.**

the transformed mixing matrix $\widehat{M}$ is obtained as follows.

$$\widehat{M}_{i,j} = [i \otimes j = 0] \cdot \frac{(1-2\mu)^{\mathbf{1}^{\mathrm{T}}(i \oplus j)}}{2} \sum_{k \in \Omega_{\overline{i} \otimes \overline{j}}} (\chi_{k \oplus i} + \chi_{k \oplus j}) \tag{14}$$

Although the mixing matrix M can be calculated efficiently (for 1-point crossover a closed-form expression for the WALSH transform $\widehat{M}$ is known [7]) the fixed point analysis of the heuristic function $\mathcal{G}$ becomes unfeasible even for moderate ℓ due to its dependence on the population vector $\boldsymbol{p}$ via $\mathcal{F}(\boldsymbol{p})$ and the cardinality $n = 2^\ell$ of the search space Ω. Therefore, in the next section we introduce the genetic algorithm with α-selection which yields a simpler heuristic function and allows to derive a closed-form solution for the fixed points.

4. SIMPLE GENETIC ALGORITHM WITH α-SELECTION

Let the current population be described by the population vector $\boldsymbol{p}$ and let

$$b = \operatorname{argmax} \{ f_i : i \in \Omega \wedge p_i > 0 \} \tag{15}$$

be the best individual in the current population. Clearly, this best individual or α-*individual b* depends on the population vector $\boldsymbol{p}$ (for clarity of notation we omit this dependence in the following). In the genetic algorithm with α-selection the α-individual b is mated with individuals randomly chosen from the current population with uniform probability. The genetic algorithm with α-selection is shown in Fig. 4.

5. DYNAMICAL SYSTEM MODEL OF THE SIMPLE GENETIC ALGORITHM WITH α-SELECTION

In the following we determine the dynamical system model for the genetic algorithm with α-selection by deriving the heuristic function $\mathcal{G}(\boldsymbol{p})$ for a given population vector $\boldsymbol{p}$.

5.1 Heuristic

The α-individual b is selected as the first parent for creation of a new offspring, whereas the second parent is chosen

```
t := 0;
initialise population P(0);
while end of adaptation ≠ true do
    select α-individual b(t) as first parent;
    for k = 0 to r − 1 do
        select second parent c(t) randomly;
        apply crossover χ_Ω and mutation μ_Ω
            a_k(t + 1) := μ_Ω (χ_Ω (b(t), c(t)));
        evaluate fitness f (a_k(t + 1));
    end
    t := t + 1;
end
```

Figure 4: Genetic algorithm with α-selection.

uniformly at random from the current population according to the probability distribution

$$\Pr\{\text{individual } j \text{ is selected}\} = p_j$$

with $j \in \Omega$. The heuristic function $\mathcal{G}$ is then given by

$$\begin{aligned} \mathcal{G}(\boldsymbol{p})_i &= \Pr\{\text{individual } i \text{ is sampled from } \Omega\} \\ &= \sum_{j \in \Omega} p_j \cdot \Pr\{\mu_\Omega (\chi_\Omega(b, j)) = i\} \\ &= \sum_{j \in \Omega} p_j \cdot M_{b \oplus i, j \oplus i} \ . \end{aligned} \tag{16}$$

By defining the $n \times n$ system matrix

$$A_{i,j} = M_{b \oplus i, j \oplus i} \tag{17}$$

we obtain the linear system of equations for the new population vector

$$\boldsymbol{q} = \mathcal{G}(\boldsymbol{p}) = A \cdot \boldsymbol{p} \ . \tag{18}$$

With the permutation matrix $(\sigma_b)_{i,j} = [i \oplus j = b]$ and the twist $(M^*)_{i,j} = M_{i \oplus j, i}$ of the symmetric mixing matrix $M = M^{\mathrm{T}}$ the system matrix A can also be expressed as

$$A = \sigma_b \cdot M^* \cdot \sigma_b \ . \tag{19}$$

Due to the α-selection scheme we gain a simpler heuristic function $\mathcal{G}$ which is completely described by the α-individual b and the mixing matrix M or system matrix A, respectively.

5.2 Fixed Points

Because of the linear relationship $\mathcal{G}(\boldsymbol{p}) = A \cdot \boldsymbol{p}$ in case of a given α-individual b the fixed points $\boldsymbol{\omega} = \mathcal{G}(\boldsymbol{\omega})$ of the heuristic function $\mathcal{G}$ are obtained from the eigenvectors of the system matrix A to eigenvalue $\lambda = 1$, i.e. $A \cdot \boldsymbol{\omega} = \boldsymbol{\omega}$. To this end, we consider the WALSH transform of both sides of the equation $\boldsymbol{q} = A \cdot \boldsymbol{p}$ according to

$$\widehat{\boldsymbol{q}} = W \cdot \boldsymbol{q} = W \cdot A \cdot \boldsymbol{p} = W \cdot A \cdot W \cdot W \cdot \boldsymbol{p} = \widehat{A} \cdot \widehat{\boldsymbol{p}}$$

where we have made use of the orthogonality of the WALSH matrix W. For an eigenvector $\boldsymbol{v}$ with eigenvalue λ we obtain

$$A \cdot \boldsymbol{v} = \lambda \cdot \boldsymbol{v} \quad \Leftrightarrow \quad \widehat{A} \cdot \widehat{\boldsymbol{v}} = \lambda \cdot \widehat{\boldsymbol{v}} \ .$$

The system matrix A and its WALSH transform $\widehat{A}$ have the same eigenvalues with eigenvectors which are also related by the WALSH transform. The WALSH transform of the system matrix can be calculated as follows

$$\widehat{A}_{i,j} = (-1)^{b^{\mathrm{T}}(i \oplus j)} \cdot \widehat{M}_{i \oplus j, j} \ . \tag{20}$$

The system matrix A as well as its WALSH transform $\widehat{A}$ depend on the α-individual b.

As an example we consider the genetic algorithm with α-selection and mutation without crossover, i.e. $\chi = 0$. For the WALSH transform of the mixing matrix we obtain

$$\widehat{M}_{i,j} = \frac{n^{1/2}}{2} \cdot (\widehat{\mu}_i \cdot [j = 0] + \widehat{\mu}_j \cdot [i = 0]) \ .$$

This yields

$$\widehat{A}_{i,j} = \frac{n^{1/2}}{2} \cdot \widehat{\mu}_i \cdot \left((-1)^{b^{\mathrm{T}}i} \cdot [j = 0] + [i = j] \right) \ .$$

For crossover χ_Ω and mutation μ_Ω the WALSH transform of the mixing matrix fulfills $\widehat{M}_{i,j} \propto [i \otimes j = 0]$. With this relation we obtain $\widehat{A}_{i,j} \propto \widehat{M}_{i\oplus j,j} \propto [(i\oplus j)\otimes j = 0]$. Because $(i\oplus j)\otimes j = 0$ is equivalent to $j\otimes\overline{i} = 0$ or $j \in \Omega_i$, respectively, the components $\widehat{A}_{i,j}$ of the WALSH transform of the system matrix are nonzero only if $j \in \Omega_i$ which is possible only for $j \leq i$. It follows that the WALSH transform $\widehat{A}$ is a lower triangular matrix with eigenvalues λ_i given by the diagonal elements $\lambda_i = \widehat{A}_{i,i} = \widehat{M}_{0,i}$. This yields the eigenvalues

$$\lambda_i = \frac{(1 - 2\mu)^{\mathbf{1}^{\mathrm{T}}i}}{2} \cdot \sum_{k\in\Omega_{\overline{i}}} (\chi_k + \chi_{k\oplus i}) \qquad (21)$$

of the system matrix A. It is a remarkable fact that the eigenvalues do not depend on the α-individual b. Taking into account $\sum_{i\in\Omega} \chi_i = 1$ and $0 < \mu < \frac{1}{2}$ it is evident from this expression that the eigenvalues are bounded by $0 \leq \lambda_i \leq 1$. The system matrix A and its WALSH transform $\widehat{A}$ are therefore positive semidefinite. Because of $\lambda_0 = 1$ and $0 \leq \lambda_i \leq 1 - 2\mu$ for $1 \leq i \leq n - 1$ there exists a single eigenvector $\boldsymbol{\omega}$ which is a fixed point of the heuristic function

$$\boldsymbol{\omega} = \mathcal{G}(\boldsymbol{\omega}) = A \cdot \boldsymbol{\omega} \ .$$

Both the system matrix A and the fixed point $\boldsymbol{\omega}$ depend on the α-individual b. According to the *fixed point hypothesis* the population will stay near this fixed point $\boldsymbol{\omega}$. If a new best or α-individual $b \in \Omega$ is found the genetic algorithm will converge to the new fixed point thus showing punctuated equilibria, i.e. phases of relative stability nearby a fixed point disrupted by sudden transitions to another fixed point.

The fixed point $\boldsymbol{\omega}$ can be determined explicitly with the help of the WALSH transform. First we note that due to the relation $\widehat{\boldsymbol{\omega}} = \widehat{A} \cdot \widehat{\boldsymbol{\omega}}$ and the lower triangular matrix $\widehat{A}$ we can recursively calculate the WALSH transform of the fixed point according to

$$\widehat{\omega}_i = \frac{1}{1 - \widehat{A}_{i,i}} \cdot \sum_{j=0}^{i-1} \widehat{A}_{i,j} \cdot \widehat{\omega}_j \qquad (22)$$

for $1 \leq i \leq n - 1$ starting with $\widehat{\omega}_0 = n^{-1/2}$ which ensures that $\sum_{i\in\Omega} \omega_i = 1$. The fixed point is then obtained via the inverse WALSH transform $\boldsymbol{\omega} = W \cdot \widehat{\boldsymbol{\omega}}$.

5.3 Experimental Results

For the shifted and inverted ONEMAX problem with fitness function $f_i = \mathbf{1}^{\mathrm{T}}(\overline{i \oplus i_0})$ and $i_0 = 128$ (corresponding to the number of 0's in the binary representation of $i \oplus i_0 \in \Omega$) Fig. 5 shows the theoretical and estimated population vectors $\boldsymbol{p}$ in generation $t = 25$ for a genetic algorithm with α-selection for $\ell = 10$ and $n = 2^{10} = 1024$ using 1-point

crossover with crossover probability $\chi = 1$, bitwise mutation with mutation probability $\mu = \frac{1}{\ell}$, population size $r = 100$ and random initial population. The EUCLIDean distance of the simulated population vector $\boldsymbol{p}$ to the fixed point $\boldsymbol{\omega}$ is

$$\sqrt{\sum_{i\in\Omega} |p_i - \omega_i|^2} = 0.0921 \ .$$

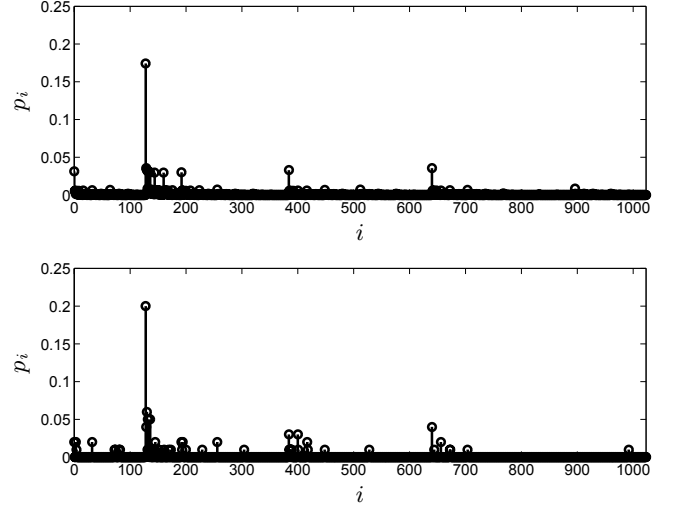There is a close match between the theoretical prediction and the experimental result.

Figure 5: **Population vector p according to the dynamical system model with heuristic function $\mathcal{G}$ (top) and its estimate from a single run of the genetic algorithm with α-selection as RHS τ (bottom).**

6. COARSE-GRAINED SYSTEM MODEL

With the help of the dynamical system model for the genetic algorithm with α-selection we will now explore coarse-grained system models based especially on schemata.

6.1 Equivalence Relations and Schemata

First we consider equivalence relations $\equiv$ according to [7]. Two equivalent individuals $i \equiv j$ in the search space Ω belong to the same equivalence class $[i] = \{j \in \Omega : j \equiv i\} \in \Omega/\equiv$. This can be expressed with the help of the linear operator (also called the quotient map)

$$\Xi_{[i],j} = [i \equiv j] = \begin{cases} 1 \ , & j \in [i] \\ 0 \ , & j \notin [i] \end{cases} \ ,$$

i.e. $i \equiv j$ if $\Xi_{[i],j} = 1$. As discussed in [7] the equivalence of individuals can be extended to the equivalence of populations. Two populations are equivalent if the proportions of individuals $j \in \Omega$ in each of the equivalence classes $[i]$ with $i \in \Omega$ are the same in both populations. By using the population vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ in the simplex Λ this yields

$$\boldsymbol{p} \equiv \boldsymbol{q} \quad \Leftrightarrow \quad \forall i \in \Omega : \sum_{j\in\Omega} [i \equiv j] \cdot p_j = \sum_{j\in\Omega} [i \equiv j] \cdot q_j \ .$$

Taking into account $\Xi_{[i],j} = [i \equiv j]$ we obtain the expression $\sum_{j\in\Omega}[i \equiv j] \cdot p_j = \sum_{j\in\Omega} \Xi_{[i],j} \cdot p_j = (\Xi\boldsymbol{p})_{[i]}$ and therefore

$$\boldsymbol{p} \equiv \boldsymbol{q} \quad \Leftrightarrow \quad \Xi\boldsymbol{p} = \Xi\boldsymbol{q} \ .$$

A dynamical system defined by the heuristic function $\mathcal{G}$ is consistently modeled with the help of a simplified coarse-grained system model implied by the equivalence relation $\equiv$ if the diagram in Fig. 6 commutes, i.e. $\Xi\mathcal{G}(\boldsymbol{p}) \overset{!}{=} \widetilde{\mathcal{G}}(\Xi\boldsymbol{p}) = \widetilde{\mathcal{G}}(\widetilde{\boldsymbol{p}})$ with $\widetilde{\boldsymbol{p}} = \Xi\boldsymbol{p}$. This amounts to the requirement that for two equivalent population vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ the population vectors in the next generation $\mathcal{G}(\boldsymbol{p})$ and $\mathcal{G}(\boldsymbol{q})$ must also be equivalent: $\boldsymbol{p} \equiv \boldsymbol{q} \Rightarrow \mathcal{G}(\boldsymbol{p}) \equiv \mathcal{G}(\boldsymbol{q})$.
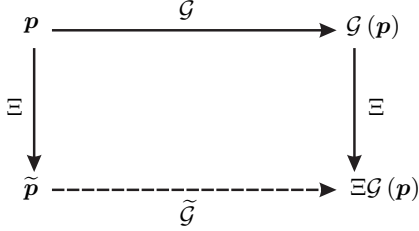

Figure 6: Commutativity diagram for the equivalence relation $\equiv$ with quotient map Ξ.

According to [7] *schemata* can be considered as specific equivalence relations. We define a *schemata family* with the help of the ℓ-tuple $\xi \in \Omega$ via the quotient map

$$\Xi_{[i],j} = [j \otimes \xi = i] \ . \tag{23}$$

Here, $i \in \Omega_\xi = \{i \in \Omega : i \otimes \overline{\xi} = 0\}$ and $j \in \Omega$. Two individuals $j, k \in \Omega$ are equivalent if they agree on the defining positions defined by ξ according to $j \equiv k \Leftrightarrow j \otimes \xi = k \otimes \xi$. The number of the defining positions is $\mathbf{1}^\mathrm{T}\xi$ which leads to the cardinality $|\Omega_\xi| = 2^{\mathbf{1}^\mathrm{T}\xi}$ of Ω_ξ. With the $2^{\mathbf{1}^\mathrm{T}\xi} \times 2^\ell$ matrix Ξ the proportion of the population representing the schema or equivalence class $[i] = i \oplus \Omega_{\overline{\xi}}$, respectively, is given by

$$(\Xi\boldsymbol{p})_{[i]} = \sum_{j \in \Omega} \Xi_{[i],j} \cdot p_j = \sum_{j \in \Omega} [j \otimes \xi = i] \cdot p_j = \sum_{j \in \Omega_{\overline{\xi}}} p_{i \oplus j} \ .$$

6.2   Schema Heuristics

As discussed in [7] the mixing operation $\mathcal{M}$ of the SGA with crossover and mutation is compatible with the equivalence relation defined by the schemata family ξ, i.e. the diagram shown in Fig. 7 commutes. This property of the mixing operation $\mathcal{M}$ and the schemata family ξ has been argued in [7, 11] as the reason for *implicit parallelism*. However, the selection operation $\mathcal{F}$ in general is not compatible with the equivalence relation $\equiv$ and, therefore, the heuristic function $\mathcal{G} = \mathcal{M} \circ \mathcal{F}$ is not compatible either [4, 7].
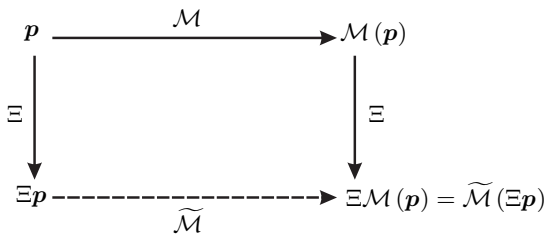

Figure 7: Commutativity diagram of the mixing operation $\mathcal{M}$ for the equivalence relation imposed by the schemata family ξ.

In the following we will analyse this general observation in case of α-selection. For the genetic algorithm with α-selection, crossover and mutation we have derived in Eq. 16 the heuristic function

$$\mathcal{G}(\boldsymbol{p})_i = \sum_{j \in \Omega} p_j \cdot M_{b \oplus i, j \oplus i}$$

with α-individual $b = \mathrm{argmax}\,\{f_i : i \in \Omega \wedge p_i > 0\}$ and mixing matrix (see Eq. 10)

$$M_{i,j} = \sum_{u,v \in \Omega} \mu_v \cdot \frac{\chi_u + \chi_{\overline{u}}}{2} \cdot [i \otimes u \oplus \overline{u} \otimes j = v] \ .$$

The proportion of the expected next population representing the schema $[i] = i \oplus \Omega_{\overline{\xi}}$ with $i \in \Omega_\xi$ can be calculated according to

$$
\begin{aligned}
(\Xi\mathcal{G}(\boldsymbol{p}))_{[i]} &= \sum_{j \in \Omega_{\overline{\xi}}} \mathcal{G}(\boldsymbol{p})_{i \oplus j} \\
&= \sum_{j \in \Omega_{\overline{\xi}}} \sum_{k \in \Omega} p_k \cdot M_{b \oplus i \oplus j, k \oplus i \oplus j} \ .
\end{aligned}
$$

After a lengthy calculation we obtain

$$
\begin{aligned}
(\Xi\mathcal{G}(\boldsymbol{p}))_{[i]} &= \sum_{j \in \Omega_\xi} (\Xi\boldsymbol{p})_{[j]} \cdot (M_\xi)_{[b \oplus i],[j \oplus i]} \\
&= \sum_{j \in \Omega_\xi} \widetilde{p}_{[j]} \cdot (M_\xi)_{[b \oplus i],[j \oplus i]} \tag{24}
\end{aligned}
$$

with $\widetilde{\boldsymbol{p}} = \Xi\boldsymbol{p}$ and the $2^{\mathbf{1}^\mathrm{T}\xi} \times 2^{\mathbf{1}^\mathrm{T}\xi}$ schema mixing matrix

$(M_\xi)_{[i],[j]}$

$$
\begin{aligned}
&= \sum_{u,v \in \Omega_\xi} (\Xi\mu)_{[v]} \cdot \frac{(\Xi\chi)_{[u]} + (\Xi\chi)_{[\overline{u}]}}{2} \cdot [i \otimes u \oplus \overline{u} \otimes j = v] \\
&= \sum_{u,v \in \Omega_\xi} \widetilde{\mu}_{[v]} \cdot \frac{\widetilde{\chi}_{[u]} + \widetilde{\chi}_{[\overline{u}]}}{2} \cdot [i \otimes u \oplus \overline{u} \otimes j = v] \tag{25}
\end{aligned}
$$

with $\widetilde{\boldsymbol{\chi}} = \Xi\boldsymbol{\chi}$ and $\widetilde{\boldsymbol{\mu}} = \Xi\boldsymbol{\mu}$. By defining the $2^{\mathbf{1}^\mathrm{T}\xi} \times 2^{\mathbf{1}^\mathrm{T}\xi}$ system matrix

$$(A_\xi)_{[i],[j]} = (M_\xi)_{[b \oplus i],[j \oplus i]} \tag{26}$$

with $i, j \in \Omega_\xi$ and a given α-individual $b \in \Omega$ we obtain the linear system of equations for the new schema population vector $\widetilde{\boldsymbol{q}} = \widetilde{\mathcal{G}}(\Xi\boldsymbol{p}) = A_\xi \cdot \Xi\boldsymbol{p}$ or equivalently

$$\widetilde{\boldsymbol{q}} = A_\xi \cdot \widetilde{\boldsymbol{p}} \ . \tag{27}$$

The schema heuristic function $\widetilde{\mathcal{G}}$ and the schema system matrix A_ξ depend on the α-individual b. Therefore, the heuristic function $\mathcal{G}$ is not compatible with the equivalence relation imposed by schemata in the strict sense. If in the next generation the α-individual b is lost or a better individual is sampled from the search space Ω the schema heuristic function $\widetilde{\mathcal{G}}$ and the schema system matrix A_ξ change. The α-individual b (or its corresponding equivalence class $[b]$) thus acts like an exogenous parameter to the coarse-grained system model as illustrated in Fig. 8.

Similar to the system matrix A the WALSH transform $\widehat{A}_\xi = W_\xi \cdot A_\xi \cdot W_\xi$ of the schema system matrix A_ξ with $2^{\mathbf{1}^\mathrm{T}\xi} \times 2^{\mathbf{1}^\mathrm{T}\xi}$ WALSH matrix W_ξ over Ω_ξ can be calculated according to

$$(\widehat{A}_\xi)_{[i],[j]} = (-1)^{b^\mathrm{T}(i \oplus j)} \cdot (\widehat{M}_\xi)_{[i \oplus j],[j]} \tag{28}$$
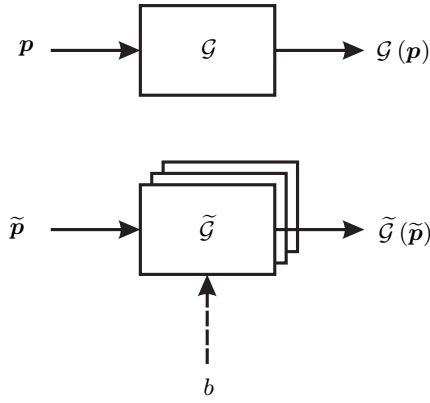
Figure 8: Dynamical system model with heuristic function $\mathcal{G}$ (top) and its coarse-grained system model with schema heuristic function $\widetilde{\mathcal{G}}$ (bottom) depending on the α-individual b.

with $i, j \in \Omega_\xi$. The schema system matrix A_ξ as well as its WALSH transform $\widehat{A}_\xi$ depend on the α-individual b. As for the system matrix A it can be shown that the WALSH transform $\widehat{A}_\xi$ of the schema system matrix A_ξ is a lower triangular matrix with an eigenvalue $\lambda_{[0]} = (\widehat{A}_\xi)_{[0],[0]} = 1$ leading to the schema fixed point $\widetilde{\boldsymbol{\omega}} = A_\xi \cdot \widetilde{\boldsymbol{\omega}}$.

6.3 Experimental Results

For numerical calculations the equivalence class $[i] \in \Omega/\equiv$ is identified with $i \in \Omega_\xi$. The elements of Ω_ξ are numbered in ordinary binary fashion, e.g. for $\ell = 5$ and $\xi = 13$ we identify $\{0\underline{0}00\underline{0}, 0\underline{0}00\underline{1}, 0\underline{0}1\underline{0}0, 0\underline{0}1\underline{0}1, 0\underline{1}00\underline{0}, 0\underline{1}00\underline{1}, 0\underline{1}1\underline{0}0, 0\underline{1}1\underline{0}1\}$ with the set $\{000, 001, 010, 011, 100, 101, 110, 111\}$ of binary 3-tuples or index numbers $\{0, 1, 2, 3, 4, 5, 6, 7\}$, respectively.

We consider the shifted and inverted ONEMAX problem with fitness function $f_i = \mathbf{1}^{\mathrm{T}}(\overline{i \oplus i_0})$ for $\ell = 10$ and $n = 2^{10} = 1024$ with the optimal binary ℓ-tuple $i_0 = 0010000000$ or $i_0 = 128$, respectively. Fig. 9 shows the theoretical and estimated schema population vectors $\widetilde{\boldsymbol{p}}$ in generation $t = 25$ for a genetic algorithm with α-selection using 1-point crossover with crossover probability $\chi = 1$, bitwise mutation with mutation probability $\mu = \frac{1}{\ell}$, population size $r = 100$ and randomly initialised population.

The schemata family is defined by the binary ℓ-tuple $\xi = 00\underline{111}00000$ or $\xi = 224$, respectively. For the calculation of the corresponding $2^3 \times 2^3$ schema system matrix

$$A_\xi =$$

$$\begin{pmatrix}
0.405 & 0.145 & 0.113 & 0.081 & 0.081 & 0.045 & 0.045 & 0.037 \\
0.045 & 0.305 & 0.013 & 0.045 & 0.009 & 0.045 & 0.005 & 0.013 \\
0.045 & 0.016 & 0.337 & 0.081 & 0.009 & 0.005 & 0.045 & 0.013 \\
0.005 & 0.034 & 0.037 & 0.294 & 0.001 & 0.005 & 0.005 & 0.037 \\
0.405 & 0.305 & 0.337 & 0.294 & 0.729 & 0.405 & 0.405 & 0.337 \\
0.045 & 0.145 & 0.037 & 0.081 & 0.081 & 0.405 & 0.045 & 0.113 \\
0.045 & 0.034 & 0.113 & 0.045 & 0.081 & 0.045 & 0.405 & 0.113 \\
0.005 & 0.016 & 0.013 & 0.081 & 0.009 & 0.045 & 0.045 & 0.337
\end{pmatrix}$$

we have used the equivalence class $[b] = [i_0] = [00\underline{1}0000000]$ according to the index number 4 of the best individual $b = i_0$ for all generations t of the dynamic system model. The

corresponding WALSH transform

$$\widehat{A}_\xi =$$

$$\begin{pmatrix}
1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.400 & 0.400 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.400 & 0.000 & 0.400 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.284 & 0.036 & 0.036 & 0.284 & 0.000 & 0.000 & 0.000 & 0.000 \\
-0.400 & 0.000 & 0.000 & 0.000 & 0.400 & 0.000 & 0.000 & 0.000 \\
-0.249 & -0.071 & 0.000 & 0.000 & 0.071 & 0.249 & 0.000 & 0.000 \\
-0.284 & 0.000 & -0.036 & 0.000 & 0.036 & 0.000 & 0.284 & 0.000 \\
-0.199 & -0.028 & 0.000 & -0.028 & 0.028 & 0.000 & 0.028 & 0.199
\end{pmatrix}$$

of the schema system matrix is a lower triangular matrix. The EUCLIDean distance of the simulated schema population vector $\widetilde{\boldsymbol{p}}$ to the schema fixed point $\widetilde{\boldsymbol{\omega}}$ is

$$\sqrt{\sum_{[i] \in \Omega/\equiv} \left| \widetilde{p}_{[i]} - \widetilde{\omega}_{[i]} \right|^2} = 0.0972 \ .$$

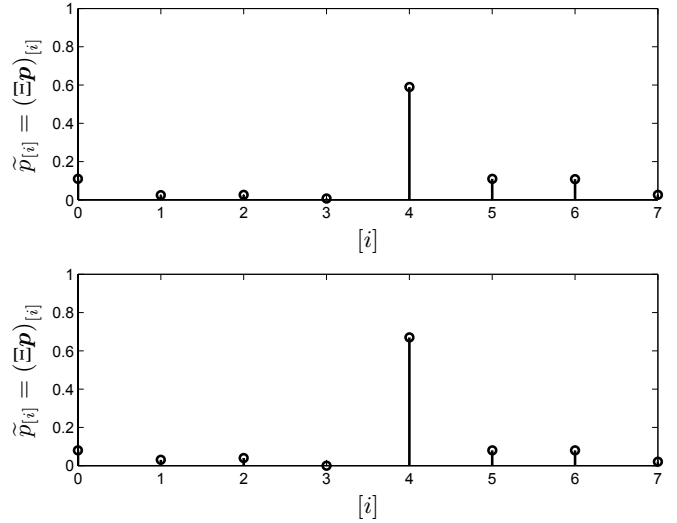There is again a close match between the theoretical prediction and the experimental result.

Figure 9: Schema population vector $\widetilde{\boldsymbol{p}} = \Xi \boldsymbol{p}$ according to the coarse-grained system model with schema heuristic function $\widetilde{\mathcal{G}}$ (top) and its estimate from a single run of the genetic algorithm with α-selection as RHS τ (bottom).

6.4 Schema Projections

There exists an interesting relationship between the projection onto the space Ω_ξ and the WALSH transform as already observed in [7]. With the help of the unit vector $\boldsymbol{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)^{\mathrm{T}}$ the j-th component of which is equal to 1 the population vector $\boldsymbol{p}$ can be written as $\boldsymbol{p} = \sum_{j \in \Omega} p_j \cdot \boldsymbol{e}_j$. Let ρ_ξ denote the projection onto the space Ω_ξ. Then the projection $\rho_\xi \boldsymbol{p}$ of the population vector $\boldsymbol{p}$ onto Ω_ξ is given by

$$\rho_\xi \boldsymbol{p} = \sum_{j \in \Omega_\xi} p_j \cdot \boldsymbol{e}_j = \sum_{j \in \Omega} [j \otimes \overline{\xi} = 0] \cdot p_j \cdot \boldsymbol{e}_j \ .$$

If we neglect all components which are projected onto 0 the projection ρ_ξ can be formulated as the $2^{\mathbf{1}^{\mathrm{T}}\xi} \times 2^\ell$ matrix

$$(\rho_\xi)_{[i],j} = [i = j] \cdot [j \otimes \overline{\xi} = 0]$$

with $j \in \Omega$. Taking into account $i \in \Omega_\xi$ and $[i \otimes \overline{\xi} = 0]$ the projection map $(\rho_\xi)_{[i],j} = [j \otimes \xi = i] \cdot [j \otimes \overline{\xi} = 0]$ is obtained. We observe a close relationship between the projection map and the quotient map $\Xi_{[i],j} = [j \otimes \xi = i]$ defined for a schemata family ξ. In fact, the WALSH transform $\widehat{\rho}_\xi = W_\xi \cdot \rho_\xi \cdot W$ with $2^\ell \times 2^\ell$ WALSH matrix W over Ω and $2^{\mathbf{1}^T \xi} \times 2^{\mathbf{1}^T \xi}$ WALSH matrix W_ξ over Ω_ξ is obtained from

$$2^{\ell/2} \cdot \widehat{\rho}_\xi = 2^{\mathbf{1}^T \xi / 2} \cdot \Xi \quad \Leftrightarrow \quad \widehat{\rho}_\xi = 2^{-\mathbf{1}^T \overline{\xi}/2} \cdot \Xi \ .$$

This yields the schema population vector

$$\begin{aligned}
\Xi \boldsymbol{p} &= 2^{\mathbf{1}^T \overline{\xi}/2} \cdot \widehat{\rho}_\xi \cdot \boldsymbol{p} \\
&= 2^{\mathbf{1}^T \overline{\xi}/2} \cdot W_\xi \cdot \rho_\xi \cdot W \cdot \boldsymbol{p} \\
&= 2^{\mathbf{1}^T \overline{\xi}/2} \cdot W_\xi \cdot \rho_\xi \widehat{\boldsymbol{p}} \ .
\end{aligned}$$

Making use of the schema heuristic $\Xi \boldsymbol{q} = A_\xi \cdot \Xi \boldsymbol{p}$ we have

$$2^{\mathbf{1}^T \overline{\xi}/2} \cdot W_\xi \cdot \rho_\xi \widehat{\boldsymbol{q}} = A_\xi \cdot 2^{\mathbf{1}^T \overline{\xi}/2} \cdot W_\xi \cdot \rho_\xi \widehat{\boldsymbol{p}}$$

leading to

$$\rho_\xi \widehat{\boldsymbol{q}} = W_\xi \cdot A_\xi \cdot W_\xi \cdot \rho_\xi \widehat{\boldsymbol{p}} = \widehat{A}_\xi \cdot \rho_\xi \widehat{\boldsymbol{p}} \ . \tag{29}$$

The vector $\rho_\xi \widehat{\boldsymbol{p}}$ corresponds to the WALSH transformed population vector $\boldsymbol{p}$ which is projected onto Ω_ξ; it evolves according to the WALSH transformed schema system matrix $\widehat{A}_\xi$ as shown in Fig. 10. As before, the matrix $\widehat{A}_\xi$ depends on the α-individual b.
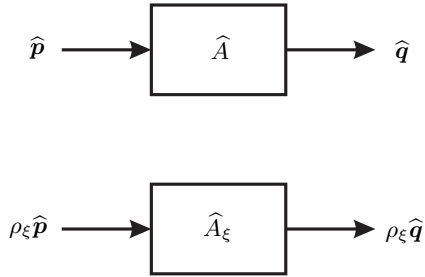

Figure 10: Projection of the WALSH **transformed population vectors** $\widehat{\boldsymbol{p}}$ **and** $\widehat{\boldsymbol{q}} = \widehat{A} \cdot \widehat{\boldsymbol{p}}$ **according to the dynamical system model with** WALSH **transformed system matrix** $\widehat{A}$ **(top) onto the space** Ω_ξ **of schemata (bottom) for a genetic algorithm with** α**-selection.**

7. CONCLUSION

The dynamical system model represents the current state of the art of genetic algorithm theory [3, 7]. For practical problem sizes, however, the determination of the population trajectory and the fixed points of the underlying heuristic function $\mathcal{G}$ becomes unfeasible.

In this paper α-selection has been introduced in order to simplify the mathematical analysis and to enable the derivation of the dynamical system model and the fixed points of the heuristic function in closed form. It turned out that for a given α-individual b the heuristic function $\mathcal{G}$ of the RHS is defined by a linear system of equations with system matrix A. There exists a single fixed point $\boldsymbol{\omega}$ which can be calculated analytically from the WALSH transformed system matrix $\widehat{A}$. Experimental results showed close agreement to the theoretical predictions obtained from the derived dynamical

system model. It has to be observed that the dynamical system model for the genetic algorithm with α-selection depends on the α-individual b. If in the next generation this individual is lost or a better individual is sampled from the search space Ω the heuristic function $\mathcal{G}$ changes due to the dependence of the system matrix A on the α-individual b, thus showing punctuated equilibria.

The introduction of the α-selection scheme further allowed to explicitly determine the coarse-grained system model for a schemata family defined by the ℓ-tuple ξ. For a given α-individual the corresponding RHS is defined by a linear system of equations with schema system matrix A_ξ with similar properties as the system matrix A.

Finally, it was shown that population vectors which are WALSH transformed and subsequently projected onto Ω_ξ evolve according to the WALSH transform of the schema system matrix $\widehat{A}_\xi$. Again, the schema system matrix $\widehat{A}_\xi$ depends on the α-individual b.

The reason for the analytical simplifications gained for the genetic algorithm with α-selection lies in the fact that the fitness function f is hidden from the mathematical formulation by introducing the α-individual b. Since α-selection is a strong selection scheme which might be too strong for practical applications future research will be oriented towards genetic algorithms with weaker selection schemes.

8. REFERENCES

[1] J. H. Holland. *Adaptation in Natural and Artificial Systems – An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence.* First MIT Press Edition, Cambridge, 1992.

[2] M. Mitchell. *An Introduction to Genetic Algorithms.* MIT Press, Cambridge, 1996.

[3] C. R. Reeves and J. E. Rowe. *Genetic Algorithms – Principles and Perspectives, A Guide to GA Theory.* Kluwer Academic Publishers, Boston, 2003.

[4] J. E. Rowe, M. D. Vose, and A. H. Wright. Coarse graining selection and mutation. In *Foundations of Genetic Algorithms*, Lecture Notes in Computer Science, pages 176–191. Springer, 2005.

[5] M. D. Vose. Modeling simple genetic algorithms. *Evolutionary Computation*, 3(4):453–472, 1995.

[6] M. D. Vose. Random heuristic search. *Theoretical Computer Science*, 229(1-2):103–142, 1999.

[7] M. D. Vose. *The Simple Genetic Algorithm – Foundations and Theory.* MIT Press, Cambridge, 1999.

[8] M. D. Vose and A. H. Wright. Simple genetic algorithms with linear fitness. *Evolutionary Computation*, 2(4):347–368, 1994.

[9] M. D. Vose and A. H. Wright. The simple genetic algorithm and the walsh transform – part i, theory. *Evolutionary Computation*, 6(3):253–273, 1998.

[10] M. D. Vose and A. H. Wright. The simple genetic algorithm and the walsh transform – part ii, the inverse. *Evolutionary Computation*, 6(3):275–289, 1998.

[11] M. D. Vose, A. H. Wright, and J. E. Rowe. Implicit parallelism. In *Genetic and Evolutionary Computation Conference GECCO 2003*, Lecture Notes in Computer Science, pages 1003–1014. Springer, 2003.