Analysis of a Non-Generational Mutationless Evolutionary Algorithm for Separable Fitness Functions

Günter Rudolph
University of Dortmund
Department of Computer Science
44221 Dortmund / Germany
guenter.rudolph@uni-dortmund.de

Abstract- It is shown that the stochastic dynamics of non-generational evolutionary algorithms with binary tournament selection and gene pool recombination but without mutation is closely approximated by a stochastic process consisting of several de-coupled random walks, provided the fitness function is separable in a certain sense. This approach leads to a lower bound on the population size such that the evolutionary algorithm converges to a uniform population with globally optimal individuals for a given confidence level.

1 Introduction

Evolutionary algorithms (EAs) with finite search sets can be exactly modeled by means of finite state Markov chains. Although the derivation of the transition matrices does not pose an essential problem, it is rarely possible to obtain analytical expressions for the state distribution, limit distribution, absorption or first passage times et cetera, since the required symbolic mathematical operations with these matrices quickly become intractably complex. Exceptional cases require more or less strong assumptions regarding the population size, the participating evolutionary operators, and the problem under consideration. Those assumptions usually reduce the state space and therefore the size of the transition matrices considerably and/or turn the transition matrix into a sparse matrix with special properties. This is a common approach in the analysis of randomized algorithms [1]. In case of evolutionary algorithms this method led to upper bounds of the expected absorption time for simple EAs (single parent, only mutation and selection) and selected problem classes [2, 3, 4]. Further results on simple EAs can be found in [5, 6] and the references therein. First analyses of population-based EAs with crossover and mutation were presented in [7, 8].

The presence of mutation is an essential ingredience in the theoretical work mentioned so far. The idea of how to approach population-based EAs with recombination and selection but without mutation was introduced in [9]. It was (somewhat vaguely) argued that the dynamics of such EAs resemble the dynamics of a specific random walk. Here, we seize this suggestion again with the objective of underpinning this approach with a sound theoretical argumentation for the class of separable fitness functions and the class of non-generational evolutionary algorithms with binary tournament selection and gene pool recombination.

For this purpose, we first present a brief introduction to random walks in Section 2 before entering the theoretical analysis given in Section 3. Some auxiliary results and their proofs are deferred to the appendix in order to exempt the argumentation from technical details. Our concluding remarks are given in Section 4.

2 Random Walks with Absorbing Barriers

The basic definitions and results of this section are extracted from [10, p. 344f.]. Let the set $\{0, 1, \ldots, n\}$ with $n < \infty$ denote the set of states that may be visited by some stochastic process. Let p_i^+ be the probability of a transition from state i to state $i+1$, p_i^- be the probability of a transition from state i to state $i-1$, and p_i^0 be the probability of a transition from state i to state i. If $p_i^- > 0$, $p_i^0 \geq 0$, and $p_i^+ > 0$ such that $p_i^- + p_i^0 + p_i^+ = 1$ for $i = 1, \ldots, n-1$ while $p_i^- = p_i^+ = 0$ and $p_i^0 = 1$ for $i \in \{0, n\}$ then the stochastic process with state space $\{0, 1, \ldots, n\}$ is called a *random walk with absorbing barriers*. The states 0 and n are termed absorbing whereas the remaining ones are termed transient.

Let a_{i0} and a_{in} denote the probabilities that the random walk will be absorbed by state 0 resp. n provided it was started at state i. In general, the relationship $a_{in} = 1 - a_{i0}$ is valid for all $i = 0, 1, \ldots, n$. If the transition probabilities are independent from the states, i.e., $p_i^- = p^-$, $p_i^0 = p^0$ and $p_i^+ = p^+$ for all $i = 1, \ldots, n-1$, then

$$a_{in} = \begin{cases} \dfrac{1-\omega^i}{1-\omega^n} & \text{if } \omega \neq 1 \\ \dfrac{i}{n} & \text{if } \omega = 1 \end{cases} \tag{1}$$

where $\omega = p^-/p^+$. Evidently, the absorption probabilities are not affected by the value of p^0. This observation leads to the result shown next.

Lemma 1
Let $p_i^- > 0, p_i^0 > 0$, and $p_i^+ > 0$ with $p_i^- + p_i^0 + p_i^+ = 1$ be the transition probabilities of a random walk with absorbing states 0 and n. If the quotients $\tilde{p}^- = p_i^-/(1-p_i^0)$

and $\tilde{p}^+ = p_i^+/(1-p_i^0)$ are independent from index i for all $i \in \{1,\ldots,n-1\}$ then the absorption probability to state n is given by Eqn. (1) where $\omega = \tilde{p}^-/\tilde{p}^+$.

Proof: Assume that the random walk has just entered some transient state i. If $p_i^0 > 0$ for the transient states of the random walk, then the random walk stays on average $1/(1-p_i^0)$ time units at state i prior to a transition either to state $i-1$ or to state $i+1$. As long as the random walk stays at state i the probabilities of a transition to the left or right remain unaltered for each step. Since we are only interested in the absorption probabilities (and not in the number of steps until absorption takes place) we may skip the period of staying at state i, provided that the transition probabilities are appropriately adjusted. Notice that the probability of finally moving to state $i+1$ conditioned by the event that state i has been left is $\tilde{p}_i^+ = p_i^+/(1-p_i^0)$, and $\tilde{p}_i^- = p_i^-/(1-p_i^0)$ in case of a transition to state $i-1$. Since $\tilde{p}_i^+ + \tilde{p}_i^- = 1$ and $\tilde{p}_i^0 = 0$ for the transient states i, the originally aperiodic random walk has been converted to a periodic random walk possessing the same absorption probabilities but a smaller absorption time. Now insist that $\tilde{p}_i^+$ is identical for all transient states i. Under this additional assumption we may use Eqn. (1) in order to calculate the absorption probabilities of the new random walk. Since these absorption probabilities are identical to those of the original random walk the proof is completed. □

3 Evolutionary Algorithms as Random Walks

3.1 Assumptions and Goals

Let $x \in S^d$ be an element of the d–dimensional search space S^d where S is a non-empty finite set with cardinality c. The objective function $f : S^d \to \mathbb{R}$ is representable via

$$f(x) = \sum_{i=1}^{d} g(x_i)$$

where $g : S \to \mathbb{R}$ is a real-valued function. Without loss of generality it is assumed that $f(\cdot)$ is to be maximized. The evolutionary algorithm under consideration is characterized as follows:

(A1) Finite population size $n < \infty$.

(A2) Non-generational binary tournament selection.

(A3) Gene-pool recombination.

(A4) No mutation.

It is clear that an evolutionary algorithm with recombination and selection but without mutation will necessarily converge to a uniform population (i.e., all individuals are identical) with probability one in finite time [11]. Notice that each of the c^n uniform populations may be attained with some non-zero probability provided that the initial population is drawn at random. In [9] it was measured how many subfunctions $g(\cdot)$ have attained their global optimum as soon as the population has become uniform, or equivalently, how many correct building blocks have been collected by an individual of a uniform population. We show that this measure indeed depends on the absorption probability of a single specific random walk. Moreover, this quantity also can be used to obtain lower bounds on or at least a good approximation of the probability that the uniform population finally attained consists of globally optimal individuals. A rearrangement of the resulting inequality yields a bound for the population size required to obtain a globally optimal uniform population for a given confidence level.

3.2 Representation of the Evolutionary Algorithm

Regardless of the choice of the selection and variation operators, the population of an evolutionary algorithm may be represented by the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{pmatrix}$$

where the row vector $a_{i\bullet} = (a_{i1}\, a_{i2}\, \ldots\, a_{id}) \in S^d$ represents the ith individual ($i = 1,\ldots,n$) whereas the column vector $a_{\bullet j} = (a_{1j}\, a_{2j}\, \ldots\, a_{nj})'$ represents the *gene pool* of component $j = 1,\ldots,d$.

In a usual evolutionary algorithm with *gene pool recombination* an offspring b would be assembled by choosing a gene at random with uniform probability $1/n$ from each gene pool. Equivalently, this might be also achieved as follows: Calculate the relative frequencies $h_j(s)$ of elements $s \in S$ in each gene pool $j = 1,\ldots,d$. An offspring b is assembled by drawing component b_j from the discrete probability distribution with $\mathsf{P}\{\, b_j = s \,\} = h_j(s)$ for $s \in S$. Thus, the population is now equivalently represented by the probability distributions $h_1(\cdot),\ldots,h_d(\cdot)$ which might be gathered in a $d \times c$ matrix (actually, a $d \times (c-1)$ matrix would suffice since the probabilities must add to unity).

Since the binary tournament selection method is used in a non–generational manner the update rule for the probability distributions $h_j(\cdot)$ is very simple. Let X and Y be two offspring independently drawn via gene pool recombination from the probability distributions $h_j(\cdot)$. The new probability distributions $h'_j(\cdot)$ for $j = 1,\ldots,d$ are obtained by $h'_j(X_j) = h_j(X_j) + \Delta_j/n$ and $h'_j(Y_j) = h_j(Y_j) - \Delta_j/n$

where

$$\Delta_j = \begin{cases} 1 & \text{if } X_j \neq Y_j \wedge f(X) \geq f(Y) \\ 0 & \text{if } X_j = Y_j \\ -1 & \text{if } X_j \neq Y_j \wedge f(X) < f(Y)\,. \end{cases}$$

Thus, the frequencies of the alleles of the winner are increased whereas those of the looser are decreased for each gene pool. This formulation of the evolutionary algorithm does not facilitate the analysis *per se*. The update probability of each gene pool is of course still dependent on the frequency distributions of the other gene pools. But the point of view developed so far opens the door to a simpler yet not exact analysis that leads to surprising accurate results under certain circumstances. A demonstration of this fact is given in the next subsection before proceeding with the general case in the subsequent subsection.

3.3 The Random Walk Model: Instructive Example

Let $S = \{0, 1\}$, $d \geq 1$ and $g(s) = a\,s + b$ where $a \neq 0$ and $b \in \mathbb{R}$. Since binary tournament selection is an ordinal selection method it suffices to consider the case $(a, b) = (1, 0)$. The maximization of the resulting objective function

$$f(x) = \sum_{i=1}^{d} x_i$$

is also known as the "counting ones problem." Since the cardinality of S is $c = |S| = 2$ the population is representable by a single vector $p = (p_1, \ldots, p_d)$ where $p_j = h_j(1)$ and $1 - p_j = h_j(0)$ for $j = 1, \ldots, d$. Let X and Y be the two offspring generated by gene pool recombination. The update rule reduces to

$$p' = p + \frac{\gamma}{n}\,(X - Y)$$

where

$$\gamma = \begin{cases} 1 & \text{if } f(X) \geq f(Y), \\ -1 & \text{otherwise.} \end{cases}$$

Let p_k^+ be the probability that gene pool k will be increased $1/n$, p_k^- the probability that it will be decreased by $1/n$ and p_k^0 the probability that it remains unaltered. These probabilities are given by

$$\begin{aligned} p_k^0 &= \mathsf{P}\{\, X_k = 1, Y_k = 1 \,\} + \mathsf{P}\{\, X_k = 0, Y_k = 0 \,\} \\ &= \mathsf{P}\{\, X_k = 1 \,\} \cdot \mathsf{P}\{\, Y_k = 1 \,\} + \\ &\quad \mathsf{P}\{\, X_k = 0 \,\} \cdot \mathsf{P}\{\, Y_k = 0 \,\} \\ &= p_k^2 + (1 - p_k)^2 \\ &= 1 - 2\,p_k\,(1 - p_k), \end{aligned} \tag{2}$$

$$\begin{aligned} p_k^+ &= \mathsf{P}\{\, X_k = 1, Y_k = 0, f(X) \geq f(Y) \,\} + \\ &\quad \mathsf{P}\{\, X_k = 0, Y_k = 1, f(X) < f(Y) \,\} \\ &= \mathsf{P}\{\, X_k = 1, Y_k = 0, D_{d,k} \geq -1 \,\} + \\ &\quad \mathsf{P}\{\, X_k = 0, Y_k = 1, D_{d,k} < 1 \,\} \\ &= \mathsf{P}\{\, X_k = 1 \,\} \cdot \mathsf{P}\{\, Y_k = 0 \,\} \cdot \mathsf{P}\{\, D_{d,k} \geq -1 \,\} + \\ &\quad \mathsf{P}\{\, X_k = 0 \,\} \cdot \mathsf{P}\{\, Y_k = 1 \,\} \cdot \mathsf{P}\{\, D_{d,k} < 1 \,\} \\ &= p_k\,(1 - p_k)\,[\,\mathsf{P}\{\, D_{d,k} \geq -1 \,\} + \mathsf{P}\{\, D_{d,k} < 1 \,\}\,] \\ &= 2\,p_k\,(1 - p_k)\,\alpha_{d,k} \end{aligned} \tag{3}$$

$$\begin{aligned} p_k^- &= 1 - p_k^0 - p_k^+ \\ &= p_k\,(1 - p_k)\,[\,2 - (\mathsf{P}\{\, D_{d,k} \geq -1 \,\} + \\ &\quad \mathsf{P}\{\, D_{d,k} < 1 \,\})\,] \\ &= 2\,p_k\,(1 - p_k)\,(1 - \alpha_{d,k}). \end{aligned} \tag{4}$$

where

$$D_{d,k} = \sum_{\substack{i=1 \\ i \neq k}}^{d} (X_i - Y_i)$$

and $\alpha_{d,k} = (\mathsf{P}\{\, D_{d,k} \geq -1 \,\} + \mathsf{P}\{\, D_{d,k} < 1 \,\})/2$. Suppose there exists an α such that $\alpha_{k,d} \geq \alpha > 0$ for all $k = 1, \ldots, d$. In this case one obtains $p_k^+ \geq 2\,p_k\,(1 - p_k)\,\alpha$ and $p_k^- \leq 2\,p_k\,(1 - p_k)\,(1 - \alpha)$ for *every* gene pool $k = 1, \ldots, d$ regardless of the true state of the other gene pools. As a consequence, the entire stochastic process over the d nonlinearly coupled gene pools may be seen as d *independent* random walks. Notice that the transition probabilities p_k^0, p_k^+, and p_k^- are state–dependent even if we use α and regard the inequalities as equalities. But Lemma 1 offers a remedy: Since

$$\begin{aligned} \tilde{p}_k^+ &= \frac{p_k^+}{1 - p_k^0} \;=\; \alpha_{d,k} \;\geq\; \alpha \\ \tilde{p}_k^- &= \frac{p_k^-}{1 - p_k^0} \;=\; 1 - \alpha_{d,k} \;\leq\; 1 - \alpha \end{aligned}$$

the absorption probabilities of the state–dependent random walk are identical to those of the associated simple random walk without state–dependent transition probabilities. Consequently, $\omega \leq \tilde{\omega} = (1 - \alpha)/\alpha < 1$ if $\alpha > 1/2$. In this case Lemma 2 ensures that the probability of absorption at state n is lower bounded by

$$a_{in} \geq \frac{1 - \tilde{\omega}^i}{1 - \tilde{\omega}^n}\,.$$

Let B^* be the random variable representing the number of independent random walks starting at state i and finally being absorbed at state n. Evidently, B^* is the sum of d independent Bernoulli random variables with success probabil-

ity a_{in}. Therefore the expectation of B^* is given by

$$\begin{aligned}\mathsf{E}[\,B^*\,] &= d\,a_{in} \;\geq\; d\cdot\frac{1-\tilde{\omega}^i}{1-\tilde{\omega}^n}\\ &= d\,\alpha^{n-i}\cdot\frac{\alpha^i-(1-\alpha)^i}{\alpha^n-(1-\alpha)^n}\,. \qquad (5)\end{aligned}$$

Notice that random variable B^* represents the number of optimized subfunctions $g(\cdot)$ within a converged population, or equivalently, the number of correctly compiled building blocks. The bound on the expectation of B^* only depends on the absorption probability a_{in} of a specific random walk and the number d of random walks executed in parallel. This partially explains the close match between the theoretical considerations and numerical experiments presented in [9].

Needless to say, it remains to guarantee that such an $\alpha > 1/2$ actually exists. Lemma 4 in conjunction with Lemma 5 yields the tight lower bound

$$\alpha \;=\; \frac{1}{2}+\frac{1}{2}\binom{2\,d-1}{d}4^{-(d-1)} \;=\; \frac{1}{2}+\binom{2\,d}{d}4^{-d} \quad (6)$$

revealing that $\alpha > 1/2$ as required. Since Stirlings's formula (see entry 6.1.38 in [12]) leads to

$$\exp\left(-\frac{1}{6\,d}\right) \;<\; \binom{2\,d}{d}4^{-d}\,(d\,\pi)^{1/2} \;<\; \exp\left(\frac{1}{24\,d}\right)$$

equation (6) may be replaced by the more convenient but remarkably accurate bound

$$\alpha \;>\; \frac{1}{2}+\exp\left(-\frac{1}{6\,d}\right)(\pi\,d)^{-1/2} \;\sim\; \frac{1}{2}+\frac{1}{\sqrt{\pi\,d}}\,. \quad (7)$$

If the evolutionary algorithm is initialized uniformly at random then the initial state of each random walk is $i = \lfloor n/2 \rfloor$. As a consequence, we obtain

$$a_{n/2,n} = \frac{1}{1+\tilde{\omega}^{n/2}}\,.$$

If each random walk is absorbed in state n then the population converges to a uniform population with optimal individuals. The probability of this event is at least

$$a_{n/2,n}^{d} \geq \left(\frac{1}{1+\tilde{\omega}^{n/2}}\right)^{d} \geq \beta$$

where $\beta \in (0,1)$ is the desired minimum probability of convergence to the optimal uniform population. Elementary transformations of the rightmost inequality above lead to

$$n \geq 2\,\log\left(\frac{1-\beta^{1/d}}{\beta^{1/d}}\right)\Big/\log\left(\frac{1-\alpha}{\alpha}\right).$$

Since α depends on the dimension d it is possible to develop an asymptotic expression for the population size n. Taking into account equation (7) we obtain

$$\log\left(\frac{1-\alpha}{\alpha}\right) \sim \log\left(\frac{1-2/\sqrt{\pi\,d}}{1+2/\sqrt{\pi\,d}}\right) \sim -\frac{4}{(\pi\,d)^{1/2}}$$

whereas

$$\log\left(\frac{1-\beta^{1/d}}{\beta^{1/d}}\right) = -\log d + \log\log\beta^{-1} + O(d^{-1})$$

for large d. Thus, if the dimension d of the problem increases the population size should be set to

$$n = \frac{\sqrt{\pi\,d}}{2}\,(\log d - \log\log\beta^{-1}) + O(d^{-1/2})$$

in order to guarantee that the EA will converge to the optimal population at least with probability $\beta > 0$.

3.4 The Random Walk Model: General Case

Let $|S| = c \geq 2$ and $\mathcal{V} = \{g(x) : x \in \mathcal{S}\} = \{v_1, \ldots, v_m\}$ with $2 \leq m \leq c$ be the set of values attainable by means of the subfunction $g(\cdot)$. Again, the random vectors X and Y denote individuals that are independently generated via gene pool recombination. Here, random variable $D_{d,k}$ is defined as

$$D_{d,k} = \sum_{\substack{i=1\\ i\neq k}}^{d}(g(X_i) - g(Y_i)).$$

Let p_k^+ denote the probability that the relative frequency of the best building block $s^* \in S$ with $g(s^*) = v_m$ of gene pool k is increased by $1/n$. This probability can be bounded by

$$\begin{aligned}p_k^+ = \;& \mathsf{P}\{\,g(X_k) = v_m, g(Y_k) \neq v_m, f(X) \geq f(Y)\,\} +\\ & \mathsf{P}\{\,g(X_k) \neq v_m, g(Y_k) = v_m, f(X) < f(Y)\,\}\\ =\; & \sum_{i=1}^{m-1}\mathsf{P}\{\,g(X_k) = v_m, g(Y_k) = v_i, f(X) \geq f(Y)\,\} +\\ & \sum_{i=1}^{m-1}\mathsf{P}\{\,g(X_k) = v_i, g(Y_k) = v_m, f(X) < f(Y)\,\}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{m-1} \mathsf{P}\{\, g(X_k) = v_m \,\} \times \mathsf{P}\{\, g(Y_k) = v_i \,\} \\
&\qquad \times \mathsf{P}\{\, D_{d,k} \geq v_i - v_m \,\} + \\
&\quad \sum_{i=1}^{m-1} \mathsf{P}\{\, g(X_k) = v_i \,\} \times \mathsf{P}\{\, g(Y_k) = v_m \,\} \\
&\qquad \times \mathsf{P}\{\, D_{d,k} < v_m - v_i \,\} \\
&\geq \sum_{i=1}^{m-1} \mathsf{P}\{\, g(X_k) = v_m \,\} \times \mathsf{P}\{\, g(Y_k) = v_i \,\} \\
&\qquad \times \mathsf{P}\{\, D_{d,k} \geq v_{m-1} - v_m \,\} + \\
&\quad \sum_{i=1}^{m-1} \mathsf{P}\{\, g(X_k) = v_i \,\} \times \mathsf{P}\{\, g(Y_k) = v_m \,\} \\
&\qquad \times \mathsf{P}\{\, D_{d,k} < v_m - v_{m-1} \,\} \\
&= \mathsf{P}\{\, g(X_k) = v_m \,\} \times (1 - \mathsf{P}\{\, g(Y_k) = v_m \,\}) \\
&\qquad \times \mathsf{P}\{\, D_{d,k} \geq -\delta \,\} + \\
&\quad \mathsf{P}\{\, g(Y_k) = v_m \,\} \times (1 - \mathsf{P}\{\, g(X_k) = v_m \,\}) \\
&\qquad \times \mathsf{P}\{\, D_{d,k} < \delta \,\} \\
&= p_k \,(1 - p_k) \,(\mathsf{P}\{\, D_{d,k} \geq -\delta \,\} + \mathsf{P}\{\, D_{d,k} < \delta \,\}) \\
&= 2\, p_k \,(1 - p_k) \,\alpha_{d,k} \qquad (8)
\end{aligned}
$$

where $\delta = v_m - v_{m-1} > 0$ is the smallest difference between the maximum and any other value of the subfunction $g(\cdot)$, $\alpha_{d,k} = (\mathsf{P}\{\, D_{d,k} \geq -\delta \,\} + \mathsf{P}\{\, D_{d,k} < \delta \,\})/2$, and $p_k = \mathsf{P}\{\, g(X_k) = v_m \,\} = \mathsf{P}\{\, g(Y_k) = v_m \,\}$ is the probability of drawing the optimal building block, i.e., p_k is the relative frequency of optimal building blocks in gene pool k. Analogously, we obtain

$$
p_k^- \leq 2\, p_k \,(1 - p_k) \,(1 - \alpha_{d,k}) \qquad (9)
$$

and

$$
p_k^0 = \sum_{i=1}^{m} \mathsf{P}\{\, g(X_k) = v_i \,\}^2 > 0\,.
$$

In principle, we consider all non-optimal building blocks as being equally bad (no distinction) and assume the worst case regarding the signal differences, i.e., the smallest value of $v_m - v_i$ where $i \neq m$. As a consequence, we only distinguish between optimal and non-optimal building blocks whose associated subfunction values are v_m resp. v_{m-1}. In other words, we have reduced the general case to the random walk model of the preceding subsection, provided we are able to find an α such that $\alpha_{d,k} \geq \alpha > 1/2$ for all $k = 1, \ldots, d$. The problem of finding tight bounds for α in the general case seems intractable. But it is easy to develop asymptotic expressions via a version of the central limit theorem.

Theorem 1 ([10], pp. 253–255)
Let $(Z_i)_{i=1}^{\infty}$ be a sequence of independent random variables which need not be identically distributed and set

$$
a_d = \sum_{i=1}^{d} \mathsf{E}[\, Z_i \,], \quad b_d^2 = \sum_{i=1}^{d} \mathsf{V}[\, Z_i \,], \quad \text{and} \quad S_d = \sum_{i=1}^{d} Z_i\,.
$$

If $|Z_i| \leq C < \infty$ for all $i \geq 1$ and $b_d/d = o(1)$ then the random variable $(S_d - a_d)/b_d$ converges in distribution to a standard normal random variable as $d \to \infty$. □

Proposition 1
Let $\{\tilde{X}_i : i = 1, \ldots, d\}$ and $\{\tilde{Y}_i : i = 1, \ldots, d\}$ be two collections of discrete, mutually independent random variables with identical finite support $\{v_1, \ldots, v_m\}$. If the distributions of $\tilde{X}_i$ and $\tilde{Y}_i$ are identical for $i = 1, \ldots, d$ then

$$
\begin{aligned}
\mathsf{P}\left\{ \sum_{i=1}^{d} (\tilde{X}_i - \tilde{Y}_i) \leq z \right\} &\approx \Phi\left(\frac{z}{\sigma_d}\right) \\
&\geq \Phi\left(\frac{z}{v_m - v_1}\sqrt{\frac{2}{d}}\right)
\end{aligned}
$$

where $\sigma_d^2 = \sum_{i=1}^{d} \mathsf{V}[\, \tilde{X}_i - \tilde{Y}_i \,]$ and $\Phi(\cdot)$ is the distribution function of the standard normal distribution.

Proof: At first it is verified that the preconditions of Theorem 1 are fulfilled. Let $D_i = \tilde{X}_i - \tilde{Y}_i$ for $i = 1, \ldots, d$. Since $|D_i| \leq C := v_m - v_1 < \infty$ the first condition is verified. As for the second condition, notice that neither the variance of $\tilde{X}_i$ nor the variance of $\tilde{Y}_i$ can exceed the value $\eta_{\max}^2 = (v_m - v_1)^2/4$ for all $i = 1, \ldots, d$. Since all random variables are mutually independent one obtains

$$
\begin{aligned}
\sigma_d^2 &= \sum_{i=1}^{d} \mathsf{V}[\, D_i \,] = \sum_{i=1}^{d} \mathsf{V}[\, \tilde{X}_i - \tilde{Y}_i \,] = 2 \sum_{i=1}^{d} \mathsf{V}[\, \tilde{X}_i \,] \\
&\leq 2\, d\, \eta_{\max}^2 = \frac{d\,(v_m - v_1)^2}{2}\,. \qquad (10)
\end{aligned}
$$

Using this inequality it immediately follows that $0 \leq \sigma_d/d \leq C/(2\,d)^{1/2} \to 0$ as $d \to \infty$. Thus, the second precondition of Theorem 1 is also fulfilled. As a consequence, the true probability distribution of the sum of differences may be approximated by a normal distribution. Finally, the inequality given in the proposition follows from inequality (10) and the fact that $\Phi(\cdot)$ is a distribution function of a continuous random variable. □

With $\tilde{X}_i = g(X_i)$ and $\tilde{Y}_i = g(Y_i)$ we may use Proposition 1 to obtain

$$
\begin{aligned}
\alpha_{d,k} &= \frac{\mathsf{P}\{\, D_{d,k} \geq -\delta \,\} + \mathsf{P}\{\, D_{d,k} < \delta \,\}}{2} \\
&\approx \Phi\left(\frac{\delta}{\sigma_{d-1}}\right) \\
&\geq \Phi\left(\frac{v_m - v_{m-1}}{v_m - v_1}\sqrt{\frac{2}{d-1}}\right) = \alpha > \frac{1}{2}\,.
\end{aligned}
$$

A Taylor expansion of $\Phi(x)$ at $x = 0$ leads to the approximation

$$\alpha \approx \frac{1}{2} + \frac{v_m - v_{m-1}}{v_m - v_1} \sqrt{\frac{1}{\pi\,(d-1)}} \tag{11}$$

which is in a noteworthy accordance with the bound (7) of the preceding example where $v_m = 1$ and $v_{m-1} = v_1 = 0$. To proceed we have to calculate the new transition probabilities of the modified random walk. Taking into account inequalities (8) and (9) we obtain

$$\begin{aligned} \tilde{p}_k^+ &= \frac{p_k^+}{1-p_k^0} = \frac{p_k^+}{p_k^+ + p_k^-} = \left(1 + \frac{p_k^-}{p_k^+}\right)^{-1} \\ &\geq \left(1 + \frac{p_k\,(1-p_k)\,(1-\alpha_{d,k})}{p_k\,(1-p_k)\,\alpha_{d,k}}\right)^{-1} = \alpha_{d,k} \geq \alpha \end{aligned}$$

and analogously $\tilde{p}_k^- \leq 1 - \alpha$. Thus, $\tilde{\omega} = (1-\alpha)/\alpha < 1$ and hence

$$a_{in} \geq \frac{1-\tilde{\omega}^i}{1-\tilde{\omega}^n} > 1 - \tilde{\omega}^i \tag{12}$$

where i denotes the initial state of the random walk for each gene pool. Suppose that $|S| = c = 2^r$ with $r \in \mathbb{N}$ and that the optimal building block is unique. In this case the elements of the original search set S^d may be encoded by binary strings of length $\ell = d \cdot r$. If the population of bit strings is initialized uniformly at random then there are on average $n/c = n \cdot 2^{-r}$ optimal building blocks in each gene pool. As a consequence, the initial state of each random walk is $i = \lfloor n \cdot 2^{-r} \rfloor$.

Now we are in the position to determine a bound for the minimum population size such that the population converges to the optimal uniform population. Owing to (12) the probability of this event is at least $a_{in}^d \geq (1 - \tilde{\omega}^i)^d$. Since the probability is required to exceed the confidence level $\beta \in (0,1)$ we obtain the inequality $(1 - \tilde{\omega}^{n/c})^d \geq \beta$ which can be rearranged to

$$n \geq r \cdot \frac{\log(1-\beta^{1/d})}{\log\left(\dfrac{1-\alpha}{\alpha}\right)}\,.$$

Notice that

$$\log(1-\beta^{1/d}) = -\log d + \log\log\beta^{-1} + O(\log(\beta)/d)$$

for large d. Taking into account the asymptotic expression (11) we finally arrive at

$$n \sim 2^{r-2}\,\pi^{1/2}\,\frac{v_m - v_1}{v_m - v_{m-1}}\,\sqrt{d-1}\,(\log d - \log\log\beta^{-1})\,. \tag{13}$$

An analysis of this expression under a *ceteris paribus* scenario yields the following results: The estimated value for the population size is not affected by an affine transformation of the fitness values. But the population size should increase

1. exponentially for linearly increasing building block size r,
2. sublinearly for linearly increasing problem dimension d, and
3. logarithmically for linearly increasing confidence level β.

Moreover, it is important to keep in mind that our estimate (13) is conservative, i.e., we consistently overestimate the actually required population size.

4 Concluding Remarks

We are certainly aware of the fact that the usefulness of this approach in analyzing evolutionary algorithms is limited. Nevertheless it offers the opportunity of investigating subclasses of evolutionary algorithms and objective functions whose analysis was intractable previously. Nevertheless, there are several directions for an extension of this approach. For example, the subfunctions and the building block sizes may differ. Even the introduction of noise [9] or weakly nonlinear interactions might be accessible by this approach.

The most interesting question, however, is associated with the phenomenon that the results derived by this approach are in close accordance with experimental results [9] obtained by *generational* evolutionary algorithms (with uniform instead of gene pool recombination). We conjecture that the answer is closely related to the reasons for the similarity of the generational Wright and non-generational Moran model [13] in genetics. A treatise of this mainly technical topic, however, is beyond the scope of this paper.

Bibliography

[1] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, New York, 1995.

[2] G. Rudolph. How mutation and selection solve long path-problems in polynomial expected time. *Evolutionary Computation*, 4(2):195–205, 1996.

[3] G. Rudolph. *Convergence Properties of Evolutionary Algorithms*. Kovač, Hamburg, 1997.

[4] S. Droste, T. Jansen, and I. Wegener. A rigorous complexity analysis of the (1+1) evolutionary algorithm for separable functions with Boolean inputs. *Evolutionary Computation*, 6(2):185–196, 1998.

[5] S. Droste, T. Jansen, and I. Wegener. On the analysis of the (1+1) evolutionary algorithm. *Theoretical Computer Science*, 276:51–81, 2002.

[6] I. Wegener and C. Witt. On the analysis of a simple evolutionary algorithm on quadratic pseudo-boolean functions. *Journal of Discrete Algorithms*, 3:61–78, 2005.

[7] T. Jansen and I. Wegener. The analysis of evolutionary algorithms – a proof that crossover really can help. *Algorithmica*, 34:47–66, 2002.

[8] T. Jansen and I. Wegener. Real royal road functions - where crossover provably is essential. *Discrete Applied Mathematics*, 149:111–125, 2005.

[9] G. Harik, E. Cantú-Paz, D. E. Goldberg, and B. L. Miller. The gambler's ruin problem, genetic algorithms, and the sizing of populations. In *Proceedings of the 1997 IEEE International Conference on Evolutionary Computation*, pages 7–12. IEEE Press, Piscataway (NJ), 1997.

[10] W. Feller. *An Introduction to Probability Theory and Its Applications Vol. 1*. Wiley, Singapore, 3rd (revised) edition, 1970.

[11] D. B. Fogel. Asymptotic convergence properties of genetic algorithms and evolutionary programming: Analysis and experiments. *Cybernetics and Systems*, 25(3):389–407, 1994.

[12] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Dover Publications, New York, 1965.

[13] M. Iosifescu. *Finite Markov Processes and Their Applications*. Wiley, Chichester, 1980.

Appendix

Lemma 2
Let $0 < i < n$ with $n \geq 2$. If $0 < x < y < 1$ then

$$\frac{1-x^i}{1-x^n} \;>\; \frac{1-y^i}{1-y^n}\,.$$

Proof:
Let $f(x) = (1-x^i)/(1-x^n)$. Notice that $f(x) > f(y)$ for $0 < x < y < 1$ if and only if $f(\cdot)$ is strictly monotonous decreasing on $(0,1)$, i.e., $f'(x) < 0$ on $(0,1)$. Since

$$\begin{aligned} f'(x) \;&=\; \frac{n\,x^{n-1}\,(1-x^i) - i\,x^{i-1}\,(1-x^n)}{(1-x^n)^2} \;<\; 0 \\ \Leftrightarrow \qquad & n\,x^{n-1}\,(1-x^i) \;<\; i\,x^{i-1}\,(1-x^n) \\ \Leftrightarrow \qquad & g(x) \;=\; x^n - \frac{n}{n-i}\,x^{n-i} + \frac{i}{n-i} \;>\; 0 \end{aligned}$$

for $x \in (0,1)$ it would suffice to show that $g(\cdot)$ is strictly monotonous on $(0,1)$ and that its infimum is larger than or equal to zero. Differentiation of $g(\cdot)$ with respect to x yields

$$g'(x) \;=\; n\,x^{n-1} - n\,x^{n-i-1} \;=\; -n\,x^{n-i-1}\,(1-x^i) \;<\; 0$$

revealing that $g(\cdot)$ is strictly monotonous decreasing on $(0,1)$. Since $g(\cdot)$ is continuous on $\mathbb{R}$ the infimum of $g(\cdot)$ restricted to $(0,1)$ is given by $g(1) = 0$. □

Lemma 3
If X and Y are independent discrete random variables with support $\{0,1,\ldots,n\}$ and $X \stackrel{d}{=} Y$ then $\mathsf{P}\{\, X - Y = 0 \,\} > \mathsf{P}\{\, X - Y = 2 \,\}$.
Proof: Let $p_k = \mathsf{P}\{\, X = k \,\} = \mathsf{P}\{\, Y = k \,\}$. Since X and Y are independent one obtains

$$\mathsf{P}\{\, X - Y = 0 \,\} = \sum_{k=0}^{n} p_k^2 \quad \text{and}$$

$$\mathsf{P}\{\, X - Y = 2 \,\} = \sum_{k=0}^{n-2} p_k\,p_{k+2}\,.$$

Notice that

$$\sum_{k=0}^{n} p_k^2 - \sum_{k=0}^{n-2} p_k\,p_{k+2} \;=$$

$$\frac{1}{2}\sum_{k=0}^{n-2}(p_k - p_{k+2})^2 + \frac{1}{2}\,(p_0^2 + p_1^2 + p_{n-1}^2 + p_n^2) \geq 0 \quad (14)$$

is zero if and only if $p_k = 0$ for all $k = 0,1,\ldots,n$. But this case is excluded since necessarily $p_0 + p_1 + \cdots + p_n = 1$. As a consequence, the expression in eqn. (14) is always larger than zero and the proof is completed. □

Lemma 4
Let $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_n$ be mutually independent Bernoulli random variables with $0 \leq \mathsf{P}\{\, X_k = 1 \,\} = p_k = \mathsf{P}\{\, Y_k = 1 \,\} \leq 1$ for $k = 1,\ldots,n$ and let

$$D_n = \sum_{k=1}^{n}(X_k - Y_k)\,.$$

The sum of probabilities $\mathsf{P}\{\, D_n = 0 \,\} + \mathsf{P}\{\, D_n = 1 \,\}$ is minimal if and only if $p_k = 1/2$ for all $k = 1,\ldots,n$.
Proof: At first notice that the distribution of D_n is symmetrical with respect to zero, i.e., for all $j = 1,\ldots,n$ holds $\mathsf{P}\{\, D_n = -j \,\} = \mathsf{P}\{\, D_n = j \,\}$. Let $Z_k = X_k - Y_k$ and

$$D_{n,k} = \sum_{\substack{i=1 \\ i \neq k}}^{n}(X_i - Y_i)$$

for an arbitrary $k \in \{1,\ldots,n\}$. It is clear that $D_{n,k}$ and Z_k are independent, $D_n = D_{n,k} + Z_k$, and that $\mathsf{P}\{\, Z_k =$

$-1\,\} = \mathsf{P}\{\, Z_k = 1\,\} = p_k\,(1-p_k) = (1-\mathsf{P}\{\, Z_k = 0\,\})/2$. After these preparations it is easily seen that

$$\begin{aligned} f(p) &= \mathsf{P}\{\, D_n = 0\,\} + \mathsf{P}\{\, D_n = 1\,\} \\ &= \mathsf{P}\{\, D_{n,k} + Z_k = 0\,\} + \mathsf{P}\{\, D_{n,k} + Z_k = 1\,\} \\ &= \sum_{i=-1}^{1} \mathsf{P}\{\, Z_k = i\,\} \cdot \mathsf{P}\{\, D_{n,k} = -i\,\} + \\ &\quad \sum_{i=-1}^{1} \mathsf{P}\{\, Z_k = i\,\} \cdot \mathsf{P}\{\, D_{n,k} = -i+1\,\} \\ &= \mathsf{P}\{\, D_{n,k} = 0\,\} + \mathsf{P}\{\, D_{n,k} = 1\,\} - \\ &\quad \mathsf{P}\{\, Z_k = 1\,\}\,(\mathsf{P}\{\, Z_{n,k} = 0\,\} - \mathsf{P}\{\, Z_{n,k} = 2\,\}) \\ &= g_k(p) - p_k\,(1-p_k)\,h_k(p) \end{aligned} \tag{15}$$

with $p = (p_1, \ldots, p_n)'$ and where the functions $g_k(p) = \mathsf{P}\{\, D_{n,k} = 0\,\} + \mathsf{P}\{\, D_{n,k} = 1\,\}$ and $h_k(p) = \mathsf{P}\{\, Z_{n,k} = 0\,\} - \mathsf{P}\{\, Z_{n,k} = 2\,\}$ do not depend on p_k. Notice that Lemma 3 ensures that $h_k(p) > 0$ for every $k \in \{1, \ldots, n\}$. Partial differentiation in eqn. (15) with respect to p_k reveals that

$$\frac{\partial f(p)}{\partial p_k} = (2\,p_k - 1)\,h_k(p) = 0 \quad \Leftrightarrow \quad p_k = \frac{1}{2}\,.$$

Since

$$\frac{\partial^2 f(p)}{\partial p_k^2} = 2\,h_k(p) > 0 \quad \text{and}$$

$$\left.\frac{\partial^2 f(p)}{\partial p_k\,\partial p_j}\right|_{p_k=1/2} = (2\,p_k - 1)\left.\frac{\partial h_k(p)}{\partial p_j}\right|_{p_k=1/2} = 0$$

for $j \neq k$ it follows that the Hessian matrix $\nabla^2 f(p)$ is positive definite at the stationary point $p^* = (1/2, \ldots, 1/2)'$. As a consequence, the value $f(p^*)$ is a local minimum of $f(\cdot)$ for $p \in [\,0,1]^n$.

To ensure that $f(p^*)$ is the global minimum it is necessary to investigate the values of $f(p)$ at the boundary of $[\,0,1]^n$. Since $f(\cdot)$ is continuous on the compact set $[0,1]^n$ it is guaranteed that the global minimum and maximum will be attained over $[0,1]^n$. In fact, the global maximum is attained at $p \in \{0,1\}^n$ with $f(p) = 1$. As for a proof that $f(p^*)$ is indeed the global minimum, first notice that repeated application of eqn. (15) leads to

$$\mathsf{P}\{\, D_n = 0 \vee 1\,\} \;=\; \mathsf{P}\{\, D_{n-m} = 0 \vee 1\,\} -$$

$$\sum_{i=0}^{m-1} p_{n-i}\,(1-p_{n-i})\,(\mathsf{P}\{\, D_{n-i+1} = 0\,\} - \mathsf{P}\{\, D_{n-i+1} = 2\,\}) \tag{16}$$

for every $p \in [\,0,1]^n$ and $m \in \{1, \ldots, n-1\}$. Now assume that the global minimum of $f(\cdot)$ will be attained at $\hat{p}$ with $f(\hat{p}) < f(p^*)$ where w.l.o.g. $0 < \hat{p}_1, \ldots, \hat{p}_{n-m} < 1$ and $\hat{p}_{n-m+1}, \ldots, \hat{p}_n \in \{0,1\}$. To indicate that random variable D_n is parameterized by $\hat{p}$ or p^* we shall write $\hat{D}_n$ and D_n^*, respectively. Owing to eqn. (16) we obtain

$$\begin{aligned} \mathsf{P}\{\, D_n^* = 0 \vee 1\,\} &< \mathsf{P}\{\, D_{n-m}^* = 0 \vee 1\,\} \quad \text{and} \\ \mathsf{P}\{\, \hat{D}_n = 0 \vee 1\,\} &= \mathsf{P}\{\, \hat{D}_{n-m} = 0 \vee 1\,\}\,. \end{aligned}$$

Since the parameters $\hat{p}_1, \ldots, \hat{p}_{n-m}$ of random variable $\hat{D}_{n-m}$ are in the open set $(0,1)^{n-m}$ it follows that $\mathsf{P}\{\, D_{n-m}^* = 0 \vee 1\,\} \leq \mathsf{P}\{\, \hat{D}_{n-m} = 0 \vee 1\,\}$ and finally

$$\begin{aligned} f(p^*) &= \mathsf{P}\{D_n^* = 0 \vee 1\,\} < \mathsf{P}\{D_{n-m}^* = 0 \vee 1\,\} \\ &\leq \mathsf{P}\{\hat{D}_{n-m} = 0 \vee 1\,\} \\ &= \mathsf{P}\{\hat{D}_n = 0 \vee 1\,\} \;=\; f(\hat{p}) \end{aligned}$$

in contradiction to the assumption that $f(\hat{p}) < f(p^*)$. As a consequence, the global minimum of $f(\cdot)$ over $[\,0,1]^n$ is attained at $p^* = (1/2, \ldots, 1/2)'$. □

Lemma 5
Let $X \stackrel{d}{=} Y \sim B(n, 1/2)$ be independent binomial random variables. The probability distribution function of $Z = X - Y$ is

$$\mathsf{P}\{\, Z = k\,\} = \binom{2\,n}{n+k} 2^{-2\,n}$$

for $k \in \{-n, \ldots, n\}$ and zero otherwise. In particular,

$$\mathsf{P}\{\, Z \geq -1\,\} + \mathsf{P}\{\, Z < 1\,\} = 1 + \binom{2\,n+1}{n+1} 2^{-2n}\,. \tag{17}$$

Proof: Notice that $X + Y \sim B(2\,n, 1/2)$. Since Y is symmetric one obtains $Y - n \stackrel{d}{=} -Y$. Thus, $Z = X - Y \stackrel{d}{=} X + Y - n$ which proves the first part of the lemma. As for the second part, the exploitation of the identity

$$\mathsf{P}\{\, Z = 1\,\} = \mathsf{P}\{\, Z = -1\,\}$$

leads to

$$\begin{aligned} &\mathsf{P}\{\, Z \geq -1\,\} + \mathsf{P}\{\, Z \leq 0\,\} \\ = \;& \mathsf{P}\{\, Z = -1\,\} + \mathsf{P}\{\, Z = 0\,\} \;+ \\ & \underbrace{\mathsf{P}\{\, Z > 0\,\} + \mathsf{P}\{\, Z \leq 0\,\}}_{=1} \\ = \;& 1 + \mathsf{P}\{\, Z = 0\,\} + \mathsf{P}\{\, Z = 1\,\} \\ = \;& 1 + 2^{-2\,n}\left[\binom{2\,n}{n} + \binom{2\,n}{n+1}\right] \\ = \;& 1 + 2^{-2\,n}\binom{2\,n+1}{n+1} \end{aligned}$$

which is the desired result. □