That is, for $t \to \infty$ a residual distance to the optimum point remains. In other words, $(\,,\,)$ strategies ***without*** σ *control are not function optimizers*; they do not converge to the optimum.

For example, consider the (μ, λ) ES. From equation (B2.4.53) one obtains

$$(\mu, \lambda) \text{ ES, } \sigma = \text{constant} > 0: \qquad \sigma_0^\star = 2c_{\mu,\lambda} \qquad \Rightarrow \qquad r_\infty = \frac{\sigma n}{2c_{\mu,\lambda}}. \tag{B2.4.70}$$

Similar results can be obtained (or observed in simulations) for all $(\,,\,)$ strategies including bitstring optimizations (for $p_\mathrm{m} = \text{constant} > 0$, see e.g. the `OneMax` $\varphi_{1,\lambda} = \bar{Q}_{1,\lambda}$ (B2.4.41)) and combinatorial problems (e.g. ordering problems; see Section G4.2). G4.2

B2.4.2 Genetic algorithms

Günter Rudolph

Abstract

The expectation of the random time at which a genetic algorithm (GA) detects the global solution or some other element of a distinguished set for the first time represents a useful global performance measure for the GA. In this section it is shown how to deduce bounds on the global performance measure from local performance measures in the case of GAs with elitist selection, mutation, and crossover.

B2.4.2.1 Global performance measures

Let the tuple $P_t = (\boldsymbol{X}_t^{(1)}, \ldots, \boldsymbol{X}_t^{(\mu)})$ denote the random population of $\mu < \infty$ individuals at generation $t \geq 0$ with $\boldsymbol{X}_t^{(i)} \in \mathbb{B}^\ell = \{0,1\}^\ell$ for $i = 1, \ldots, \mu$. Assume that the genetic algorithm (GA) is used to find a global solution $\boldsymbol{x}^* \in \mathbb{B}^\ell$ at which the objective function $f : \mathbb{B}^\ell \to \mathbb{R}$ attains its global maximum $f(\boldsymbol{x}^*) = f^* = \max\{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{B}^\ell\}$. The best objective function value of a population P_t at generation $t \geq 0$ can be extracted via the mapping

$$f_\mathrm{b}(P_t) = \max\{f(\boldsymbol{X}_t^{(i)}) : i = 1, \ldots, \mu\}.$$

Then the random variable

$$T = \min\{t \geq 0 : f_\mathrm{b}(P_t) = f^*\}$$

denotes the *first hitting time* of the GA. Assume that the expectation of T can be bounded via $\mathbf{E}[\,T\,] \leq \widehat{T}$ B2.2.2 where $\widehat{T}$ is a polynomial in the problem dimension ℓ. If the GA is stopped after $c\,\widehat{T}$ steps with $c \geq 2$ one cannot be sure in general whether the best solution found is the global solution or not. The probability that the candidate solution is not the global solution is $\mathbf{P}\{T > c\,\widehat{T}\}$ which can be bounded via the Markov inequality yielding

$$\mathbf{P}\{T > c\,\widehat{T}\} \;\leq\; \frac{\mathbf{E}[\,T\,]}{c\,\widehat{T}} \;\leq\; \frac{\widehat{T}}{c\,\widehat{T}} \;=\; \frac{1}{c} \;\leq\; \frac{1}{2}.$$

After k independent runs (with different random seeds) the probability that the global solution has been found at least once is larger than or equal to $1 - c^{-k}$. For example, after 20 runs with $c = 2$ (possibly in parallel) the probability that the global solution has not been found is less than 10^{-6}.

If such a polynomial bound $\widehat{T}$ existed for some evolutionary algorithm and a class of optimization problems whose associated decision problem is nondeterministic polynomial-time (NP) complete, every optimization problem of this difficulty could be treated with this fictitious evolutionary algorithm in a similar manner. In fact, for the field of evolutionary algorithms this would be a pleasant result, but such a result is quite unlikely to hold.

B2.4.2.2 Deducing the global performance measure from the local performance measure

The moments of the first hitting time can be calculated from the transition matrix of the *Markov chain* B2.2.2 associated with the GA and the objective function under consideration. Unless the transition matrix is sparsely filled the practical application of the formulas given by Iosifescu (1980, pp 104, 133) is usually excluded.

The general idea to circumvent this problem is as follows. Let $S = \mathbb{B}^\ell \times \cdots \times \mathbb{B}^\ell = (\mathbb{B}^\ell)^\mu$ be the *state space* of the Markov chain. Then each element $\boldsymbol{x} \in S$ represents a potential population that can be attained by the GA in the course of the search. Notice that it is always possible to decompose the state space into n subsets via B2.2.2

$$S = \bigcup_{i=1}^{n} S_i \quad \text{with} \quad S_i \cap S_j = \emptyset \quad \text{for} \quad i \neq j$$

with the property

$$\forall \boldsymbol{x} \in S_i : \forall y \in S_j : i < j \Rightarrow f_{\mathrm{b}}(\boldsymbol{x}) < f_{\mathrm{b}}(\boldsymbol{y}).$$

If the GA employs *elitist selection* it is guaranteed that a population in subset S_j will never transition to a population represented by a state in some subset S_i with $i < j$. Thus, the Markov chain moves through the sets S_i with ascending index i. In general, this grouping of the states does not constitute a Markov chain whose states are represented by the sets S_i (see Iosifescu 1980, pp 166–70). In this case one has to determine a lower bound on the probabilities to transition from some arbitrary element in S_i to an arbitrary element in S_j. These lower bounded probabilities represent the transition probabilities p_{ij} for the grouped Markov chain to transition from set S_i to S_j. After the probabilities p_{ij} have been determined for $j = i+1, \ldots, n$ the setting C2.7.4

$$p_{ii} = 1 - \sum_{j=i+1}^{n} p_{ij}$$

ensures that the row sums of the transition matrix of the grouped Markov chain are unity.

If the mutation of an individual is realized by inverting each bit with some *mutation probability* $p \in (0, 1)$ then there exist nonzero transition probabilities to move from set S_i to S_j for all indices i, j with $1 \leq i < j \leq n$. This Markov chain can be simplified by setting C3.2.1

$$\begin{aligned} q_{ii} &= p_{ii} + \sum_{j=i+2}^{n} p_{ij} \\ q_{i,i+1} &= p_{i,i+1} \\ q_{i,i+k} &= 0 \qquad \text{for } k \geq 2 \text{ and } i + k \leq n \\ q_{ij} &= 0 \qquad \text{for } j < i. \end{aligned}$$

Thus, only transitions from the set S_i to S_{i+1} for $i = 1, \ldots, n-1$ are considered—the remaining improving transitions are ignored by bending them back to state S_i. Evidently, this simplified Markov chain must have a worse performance than the original Markov chain, but its simple structure allows an easy determination of the first hitting time representing an upper bound on the first hitting time of the original chain. To this end, let T_{ij} denote the random time that is necessary to transition from set S_i to S_j. Then the expectation of T is bounded by

$$\mathbf{E}[\, T \,] \leq \sum_{i=1}^{n-1} \mathbf{E}[\, T_{i,i+1} \,]. \tag{B2.4.71}$$

Evidently, the probability distribution of random variable $T_{i,i+1}$ is geometric with probability density function

$$\mathbf{P}\{T_{i,i+1} = \tau\} = q_{i,i+1}\,(1 - q_{i,i+1})^{\tau - 1}$$

and expectation $\mathbf{E}[\, T_{i,i+1} \,] = 1/q_{i,i+1}$. Consequently, the expectation of the first hitting time T of the GA can bounded by

$$\mathbf{E}[\, T \,] \leq \sum_{i=1}^{n-1} \frac{1}{q_{i,i+1}}. \tag{B2.4.72}$$

It is not guaranteed that this approach will always lead to sharp bounds. The manner in which the state space is decomposed determines the quality of the bounds. Unfortunately, there is currently no guideline helping to decide which partitioning will be appropriate. The following examples will offer the opportunity to gain some experience, but before beginning the examples notice that it can be sufficient to analyze the $(1+1)$ GA with mutation and elitist selection to obtain an upper bound of the first hitting time: an ordinary GA with elitist selection, mutation, and crossover is at least as fast as a $(1+1)$ GA with the same mutation probability. Thus, the potential improving effects of crossover will be ignored. This can lead to weak bounds—but as long as the bounds are polynomially bounded in ℓ this approach is reasonable.

B2.4.2.3 Linear binary problems

Definition B2.4.1. A function $f : \mathbb{B}^\ell \to \mathbb{R}$ is called *linear* if it is representable via

$$f(\boldsymbol{x}) = a_0 + \sum_{i=1}^{\ell} a_i \, x_i$$

with $\boldsymbol{x} \in \mathbb{B}^\ell$ and $a_i \in \mathbb{R}$ for $i = 0, 1, \ldots, \ell$. □

The so-called *counting ones problem* consists of the task of finding the maximum of the linear function

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

that is attained if all entries in vector $\boldsymbol{x}$ are set to 1. Bäck (1992) investigated this problem for a $(1+1)$ GA with mutations as described previously and derived the transition probabilities for the Markov chain while Mühlenbein (1992) succeeded in calculating an approximation of the expected number of function evaluations needed to reach the optimum.

The first step of the analysis is to reduce the state space of the Markov chain by an appropriate grouping of states: to this end note that there are $\binom{\ell}{i}$ states with i ones that can be grouped into one state because the specific instantiation of the vector with exactly i ones is not important—the probability of transition to any other state only depends on the number of ones (or zeros). Thus, the states of the grouped Markov chain represent the number of ones. This reduces the cardinality of the state space from 2^ℓ to $\ell + 1$: the Markov chain is in state $i \in \{0, 1, \ldots, \ell\}$ if there are exactly i ones in the current vector. Consequently, one would like to know the expected time to reach state ℓ.

The next step consists of the determination of the transition probabilities. Since the algorithm only accepts improvements it is sufficient to know the transition probabilities from some state i to some state $j > i$. Let A_{ik} be the event that 'k ones out of i flip to zero and $i - k$ ones are not flipped' and B_{ijk} the event that '$k + j - i$ zeros out of $\ell - i$ flip to one and $\ell - j - k$ zeros are not flipped'. Note that both events are independent. The probabilities of these events are

$$\mathbf{P}\{A_{ik}\} = \binom{i}{k} p^k (1-p)^{i-k} \quad \text{and} \quad \mathbf{P}\{B_{ijk}\} = \binom{\ell - i}{k + j - i} p^{k+j-i} (1-p)^{\ell-j-k}.$$

Thus, the probability to transition from state i to j is the sum over k of the product of the probabilities of both events ($0 \le i < j \le \ell$):

$$\begin{aligned} p_{ij} &= \sum_{k=0}^{\ell} \mathbf{P}\{A_{ik}\} \cdot \mathbf{P}\{B_{ijk}\} \\ &= \sum_{k=0}^{\ell} \binom{i}{k} p^k (1-p)^{i-k} \cdot \binom{\ell - i}{k + j - i} p^{k+j-i} (1-p)^{\ell-j-k} \\ &= \sum_{k=0}^{\ell} \binom{i}{k} \binom{\ell - i}{k + j - i} p^{2k+j-i} (1-p)^{\ell+i-j-2k} \\ &= p^{j-i} (1-p)^{\ell-(j-i)} \sum_{k=0}^{\ell} \binom{i}{k} \binom{\ell - i}{k + j - i} \left(\frac{p}{1-p}\right)^{2k}. \end{aligned} \tag{B2.4.73}$$

This formula is equivalent to that of Bäck (1992, p 88). The last nonzero term of the series in (B2.4.73) is that with index $k = \min\{i, \ell - j\}$. For larger indices at least one of the binomial coefficients becomes zero. This reflects the fact that some of the events are impossible: for example, the event A_{ik} can not occur if $k > i$ because one cannot flip k ones when there are only i. Since the Markov chain stays in its current state if mutation has generated a state of worse or equal quality the probabilities of staying are

$$p_{ii} = 1 - \sum_{j=i+1}^{\ell} p_{ij}$$

for $0 \leq i < \ell$. Clearly, $p_{\ell\ell} = 1$. Since all other entries are zero the transition matrix $\mathbf{P} = (p_{ij})$ has been derived completely.

Now we are in the position to use the technique described previously. Mühlenbein (1992) used a similar method to attack this problem: in principle, he also converted the exact Markov chain to another one, that always performs less well than the original one but which is much simpler to analyze. Actually, his analysis was a pure approximation without taking into account whether the approximations yielded a lower or upper bound of the expected absorption time. However it will be shown in the following that this approach leads to an upper bound of the expectation of the first hitting time.

In the third step the original Markov chain is approximated by a simpler one that has (provable) worse performance. Recall that the key idea is to ignore all those paths that take shortcuts to state ℓ by jumping over some states in between. If the original Markov chain takes such a shortcut this move is considered deteriorating in the approximating Markov chain and it stays at its current state. Consequently, the approximating Markov chain needs more time to reach the absorbing state ℓ on average. Moreover, the approximating chain must pass all states greater than or equal to i to arrive at state ℓ when being started in state $i < \ell$.

Thus, one needs to know the transition probabilities q_{ij} of the simplified Markov chain. Actually, it is sufficient to know the values for $q_{i,i+1}$ with $i = 0, \ldots, \ell - 1$. In this case (B2.4.73) reduces to

$$q_{i,i+1} = p\,(1-p)^{\ell-1} \sum_{k=0}^{\ell} \binom{i}{k} \binom{\ell-i}{k+1} \left(\frac{p}{1-p}\right)^{2k} \tag{B2.4.74}$$

$$\geq p\,(1-p)^{\ell-1}(\ell - i). \tag{B2.4.75}$$

Expression (B2.4.74) is still too complicated. Therefore it was bounded by (B2.4.75). In principle, the approximating Markov chain was approximated again by a Markov chain with even worse performance: the probabilities to transition to the next state were decreased so that this (third) Markov chain will take an even longer time to reach state ℓ. To determine the expected time until absorption insert (B2.4.75) into (B2.4.72). This leads to

$$\mathbf{E}[\,T\,] \leq \sum_{i=0}^{\ell-1} \frac{1}{p\,(1-p)^{\ell-1}(\ell-i)} = \frac{1}{p\,(1-p)^{\ell-1}} \sum_{i=1}^{\ell} \frac{1}{i} \leq \frac{\log \ell + 1}{p\,(1-p)^{\ell-1}}. \tag{B2.4.76}$$

Evidently, the absorption time depends on the mutation probability $p \in (0, 1)$ and attains its minimum for $p^* = 1/\ell$. Then (B2.4.76) becomes (also see Mühlenbein 1992, p 19)

$$\mathbf{E}[\,T\,] \leq \ell\,(\log \ell + 1)\left(1 - \frac{1}{\ell}\right)^{1-\ell} \leq \ell\,(\log \ell + 1)\,\exp(1). \tag{B2.4.77}$$

The bound (B2.4.77) is very close to the absorption time of the original Markov chain with $p = 1/\ell$. It is clear that the optimal mutation probability for the original Markov chain will differ from $1/\ell$, but the difference is remarkably small as the numerical investigations of Bäck (1993) reveal.

B2.4.2.4 Unimodal binary functions

The notion of *unimodal* functions usually appears in probability theory (to describe the shape of probability density functions), nonlinear one-dimensional dynamics (to characterize the shapes of return maps) and in the theory of optimization of one-dimensional functions with domain $\mathbb{R}$. Since a commonly accepted definition for unimodal functions in $\mathbb{R}^\ell$ does not seem to exist, it comes as no surprise that the definition of unimodality of function with domain $\mathbb{B}^\ell$ is not unique in the literature either. Here, the following definition will be used.

Definition B2.4.2. Let f be a real-valued function with domain $\mathbb{B}^\ell$. A point $\boldsymbol{x}^* \in \mathbb{B}^\ell$ is called a *local solution* of f if

$$f(\boldsymbol{x}^*) \geq f(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in \{\boldsymbol{y} \in \mathbb{B}^\ell : \|\boldsymbol{y} - \boldsymbol{x}^*\|_1 = 1\}. \tag{B2.4.78}$$

If the inequality in (B2.4.78) is strict, then $\boldsymbol{x}^*$ is termed a *strictly local solution*. The value $f(\boldsymbol{x}^*)$ at a (strictly) local solution is called a *(strictly) local maximum* of f. A function $f : \mathbb{B}^\ell \to \mathbb{R}$ is said to be *unimodal*, if there exists exactly one local solution. □

Before determining the expected absorption time of a $(1+1)$-EA for this problem, it is useful to know whether such problems are solvable in polynomial time at all. Johnson *et al* (1988, p 86) have shown that this problem cannot be NP hard unless NP = co-NP, an event which is commonly considered very unlikely.

The *ladder problem* consists of the task of finding the maximum of the unimodal binary function

$$f(\boldsymbol{x}) = \sum_{i=1}^{\ell} \prod_{j=1}^{i} x_j$$

which is attained if all entries in vector $\boldsymbol{x}$ are set to 1. The objective function counts the number of consecutive 1s in $\boldsymbol{x}$ from left to right. Note that this function is unimodal: choose $\boldsymbol{x} \in \mathbb{B}^\ell$ such that $\boldsymbol{x} \neq \boldsymbol{x}^*$. It suffices to show that there exists a point $y \in \mathbb{B}^\ell$ with $\|\boldsymbol{x}-\boldsymbol{y}\|_1 = 1$ and $f(\boldsymbol{y}) > f(\boldsymbol{x})$. In fact, this is true: since $\boldsymbol{x} \neq \boldsymbol{x}^*$ there exists an index $k = \min\{i : x_i = 0\} \leq \ell$. Choose $\boldsymbol{y} \in \mathbb{B}^\ell$ such that $y_i = x_i$ for all $i \in \{1, \ldots, \ell\} \setminus \{k\}$ and $y_k = 1$. By construction one obtains $\|\boldsymbol{x}-\boldsymbol{y}\|_1 = 1$ and finally $f(\boldsymbol{y}) > f(\boldsymbol{x})$ since the number of consecutive 1s in $\boldsymbol{y}$ is larger than the number of consecutive 1s in $\boldsymbol{x}$. Consequently, $\boldsymbol{x}^* = (1 \ldots 1)'$ is the only point at which f attains a local maximum. Therefore f is unimodal.

To derive an upper bound on the expected number of steps to reach the global maximum consider the following decomposition of the search space: define $S_i := \{\boldsymbol{x} \in \mathbb{B}^\ell : f(\boldsymbol{x}) = i\}$ for $i = 0, 1, \ldots, \ell$. For example, for $\ell = 4$ one obtains

$$\begin{aligned} S_0 &= \{0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111\} \\ S_1 &= \{1000, 1001, 1010, 1011\} \\ S_2 &= \{1100, 1101\} \\ S_3 &= \{1110\} \\ S_4 &= \{1111\}. \end{aligned}$$

Thus, if $\boldsymbol{x} \in S_i$ then the first i bits are set correctly. Note that this grouping of states is not suited to formulate a Markov chain model with $\ell+1$ states that is equivalent to a model with 2^ℓ states. But it is possible to formulate a simplified Markov chain model with $\ell+1$ states that has worse performance than the true model. To this end assume $\boldsymbol{x} \in S_0$. Subset S_0 only can be left if the first entry mutates from zero to one. If this event occurs the Markov chain is at least in subset S_1. But it may also happen that the Markov chain transitions to any other subset S_i with $i > 1$. In the simplified model these events are not allowed: all transitions from S_0 to S_i with $i > 1$ are considered as transitions to S_1. Subset S_1 can be left only if the first entry does not mutate and the second entry flips from zero to one. In this case the Markov chain transitions at least to subset S_2. All transitions to subset S_i with $i > 2$ are considered as transitions to S_2. Analogous simplifications apply to the other subsets S_i. Since all shortcuts on the path to S_ℓ are bent back to a transition of the type S_i to S_{i+1} the expected number of trials of the simplified Markov chain is larger than the expected number of trials of the original Markov chain. The state space of the simplified Markov chain is $S = \{0, 1, \ldots, \ell\}$ where state $i \in S$ represents subset S_i. The only possible path from state 0 to state ℓ must visit all states in between in ascending order. Thus, the probability $p_{i,i+1}$ to transition from state i to $i+1$ for $i < \ell$ is the probability to flip entry $i+1$ multiplied by the (independent) probability that the first i entries remain unchanged. Thus,

$$p_{i,i+1} = p(1-p)^i$$

where $p \in (0, 1)$ denotes the probability to flip from 0 to 1 and vice versa. The expected number of steps to reach the optimum is

$$\mathbf{E}[\, T_{0,\ell} \,] = \sum_{i=0}^{\ell-1} \mathbf{E}[\, T_{i,i+1} \,] = \sum_{i=0}^{\ell-1} \frac{1}{p_{i,i+1}} = \frac{1}{p} \sum_{i=0}^{\ell-1} \left(\frac{1}{1-p} \right)^i = \frac{1-p}{p^2} [\, (1-p)^{-\ell} - 1 \,]. \qquad \text{(B2.4.79)}$$

Now insist that $p = c/\ell$ with $0 < c < \ell$. Insertion into (B2.4.79) leads to

$$\mathbf{E}[\, T_{0,\ell} \,] = \frac{\ell^2}{c^2} \left(1 - \frac{c}{\ell}\right) \left[\left(1 - \frac{c}{\ell}\right)^{-\ell} - 1 \right] \leq \ell^2 \, \frac{e^c - 1}{c^2} \qquad \text{(B2.4.80)}$$

where the rightmost expression attains its minimum for $c \approx 1.6$. In summary, it has been shown that the expected number of steps of the $(1+1)$-EA can be bounded by $O(\ell^2)$.

Let $F = \{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{B}^\ell\}$ be the set of function values of a unimodal function f. If the cardinality of F is bounded by a polynomial in ℓ, then it is guaranteed that the (1+1)-EA will be absorbed at the local/global solution after polynomially many trials on average, because only polynomially many improvements via one-bit mutations are possible and sufficient to reach the optimum. Such a problem was considered in the preceding example. Therefore, these problems can be excluded from further considerations. Rather, unimodal problems with $|F| = \Omega(2^\ell)$ are the interesting candidates.

By definition, each unimodal problem has at least one path to the optimum with strictly increasing function values, where consecutive points on the path differ in one bit only. Since the expected time to change a single specific bit is less than $e\ell$, an upper bound on the absorption time is the length of the path times $e\ell$. Horn *et al* (1994) succeeded in constructing paths that grow exponentially in ℓ and can be used to build unimodal problems. Consequently, the upper bound derived by the above reasoning either is too rough or indicates that polynomial bounds do not exist. It is clear that such a 'long path' must possess much structure, because the one-bit path has to be folded several times to fit into 'box' $\mathbb{B}^\ell$. One might suspect that there exist many shortcuts, by appropriate two-bit mutations, that decrease the order of the upper bound considerably. In fact, this is true. Since the analysis is quite involved only the result will be reported: the exponentially long `root2`-path is maximized after $O(\ell^3)$ function evaluations on average (see Rudolph 1996).

B2.4.2.5 Supermodular functions

Definition B2.4.3. A function $f : \mathbb{B}^\ell \to \mathbb{R}$ is said to be *supermodular* if

$$f(\boldsymbol{x} \wedge \boldsymbol{y}) + f(\boldsymbol{x} \vee \boldsymbol{y}) \geq f(\boldsymbol{x}) + f(\boldsymbol{y}) \tag{B2.4.81}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{B}^\ell$. If the inequality in (B2.4.81) is reversed, then f is called *submodular*. □

Evidently, if $f(\boldsymbol{x})$ is supermodular then $g(\boldsymbol{x}) := a + bf(\boldsymbol{x})$ with $a \in \mathbb{R}$ and $b \in \mathbb{R}\backslash\{0\}$ is supermodular for $b > 0$ and submodular for $b < 0$. Thus, maximization of supermodular functions is of the same difficulty as the minimization of submodular functions. For this problem class there exists a strong result.

Theorem B2.4.1 (Grötschel et al 1993, pp 310–11). Each supermodular function $f : \mathbb{B}^\ell \to \mathbb{Q}$ can be globally maximized in strongly polynomial time. □

As will be shown, it is impossible to obtain an upper bound $\widehat{T}$ on the expectation of the first hitting time that is polynomial in ℓ.

Theorem B2.4.2. There exist supermodular functions that cannot be maximized by a $(1 + 1)$-EA with a number of mutations that is upper bounded by a polynomial in ℓ.

Proof. Consider the objective function

$$f(\boldsymbol{x}) = \begin{cases} \ell + 1 & \text{if } \|\boldsymbol{x}\|_1 = \ell \\ \ell - \|\boldsymbol{x}\|_1 & \text{if } \|\boldsymbol{x}\|_1 < \ell \end{cases} \tag{B2.4.82}$$

that is easily shown to be supermodular. The state space of the $(1 + 1)$-EA can be represented by $S = \{0, 1, \ldots, \ell\}$ where each state $i \in S$ represents the number of 1s in vector $\boldsymbol{x} \in \mathbb{B}^\ell$. The absorbing state is state ℓ. It can be reached from state $i \in \{0, 1, \ldots, \ell - 1\}$ within one step with probability

$$p_{i\ell} = p^{\ell - i}\,(1 - p)^i.$$

Let the Markov chain be in some state $i \in \{0, \ldots, \ell - 1\}$. Only transitions to some state $j < i$ or to state ℓ are possible. If the Markov chain transitions to state $j < i$, then the probability to transition to state ℓ has become smaller. Thus, it would be better to stay at state i than to move to state $j < i$ although the objective function value of state j is better than the objective function value of state i. This leads to the simplified Markov chain that has better performance than the original one. Then $p_{ii} = 1 - p_{i\ell}$ and the simplified Markov chain is described completely. Thus, the expected time to transition to state ℓ from state $i < \ell$ is

$$\mathbf{E}[\,T_{i,\ell}\,] = \frac{1}{p_{i\ell}} = \frac{1}{p^{\ell - i}\,(1 - p)^i} = \ell^\ell \left(\frac{1}{\ell - 1}\right)^i \geq \ell^{\ell - i}.$$

Assuming that the initial point is drawn from a uniform distribution over $\mathbb{B}^\ell$ the average time to absorption is larger than

$$2^{-\ell}\sum_{i=0}^{\ell-1}\binom{\ell}{i}\mathbf{E}[\,T_{i,\ell}\,] \geq 2^{-\ell}\sum_{i=0}^{\ell-1}\binom{\ell}{i}\ell^{\ell-i} = \left(\frac{\ell}{2}\right)^{\ell}\sum_{i=0}^{\ell-1}\binom{\ell}{i}\left(\frac{1}{\ell}\right)^{i} = \left(\frac{\ell+1}{2}\right)^{\ell} - 2^{-\ell}.$$

Since the lower bound on the absorption time is exponential in ℓ for $\ell \geq 2$ the proof is completed. □

Of course, this result does not imply that a GA must fail to solve this problem in a polynomially bounded number of generations. It may be that some crossover operator can help. But note that the objective function (B2.4.82) is *fully deceptive* as can be easily verified owing to the sufficient conditions presented by Deb and Goldberg (1994). Fully deceptive functions are the standard examples to show (empirically) that a GA fails. B2.7.1

B2.4.2.6 Almost-positive functions

Theorem B2.4.3 (Hansen and Simeone 1986, p 270). The maximum of an almost-positive pseudo-Boolean function (i.e. the coefficients of all nonlinear terms are nonnegative) can be determined in strongly polynomial time. □

Theorem B2.4.4. There exist supermodular functions that cannot be maximized by a $(1+1)$-EA with a number of mutations that is upper bounded by a polynomial in ℓ.

Proof. Theorem B2.4.2 has shown that the objective function in equation (B2.4.82) cannot be maximized by a number of mutations that is upper bounded by a polynomial in ℓ. Note that the function in (B2.4.82) has the alternative representation

$$f(\boldsymbol{x}) = \ell - \sum_{i=1}^{\ell} x_i + (\ell+1)\prod_{i=1}^{\ell} x_i$$

revealing that this function is also almost positive. This completes the proof. □

References

Abramowitz M and Stegun I A 1984 *Pocketbook of Mathematical Functions* (Thun: Harri Deutsch)

Bäck T 1992 The interaction of mutation rate, selection, and self-adaptation within a genetic algorithm *Parallel Problem Solving from Nature, 2 (Proc. 2nd Int. Conf. on Parallel Problem Solving from Nature, Brussels, 1992)* ed R Männer and B Manderick (Amsterdam: North-Holland) pp 85–94

——1993 Optimal mutation rates in genetic search *Proc. 5th Int. Conf. on Genetic Algorithms (Urbana-Champaign, IL, July 1993)* ed S Forrest (San Mateo, CA: Morgan Kaufmann) pp 2–8

Beyer H-G 1989 *Ein Evolutionsverfahren zur mathematischen Modellierung stationärer Zustände in dynamischen Systemen* Doctoral Dissertation, HAB-Dissertation 16, Hochschule für Architektur und Bauwesen, Weimar

——1993 Toward a theory of evolution strategies: some asymptotical results from the $(1 \dagger \lambda)$-theory *Evolutionary Comput.* **1** 165–88

——1994a Towards a theory of 'evolution strategies': progress rates and quality gain for $(1 \dagger \lambda)$-strategies on (nearly) arbitrary fitness functions *Parallel Problem Solving from Nature—PPSN III (Proc. Int. Conf. on Evolutionary Computation and 3rd Conf. on Parallel Problem Solving from Nature, Jerusalem, October 1994) (Lecture Notes in Computer Science 866)* ed Yu Davidor, H-P Schwefel and R Männer (Berlin: Springer) pp 58–67

——1994b *Towards a Theory of 'Evolution Strategies': Results from the N-dependent* (μ, λ) *and the Multi-Recombinant* $(\mu/\mu, \lambda)$ *Theory* Department of Computer Science Technical Report SYS-5/94, University of Dortmund

——1995a *How GAs do* NOT *Work—Understanding GAs without Schemata and Building Blocks* Department of Computer Science Technical Report SYS-2/95, University of Dortmund

——1995b Toward a theory of evolution strategies: the (μ, λ)-theory *Evolutionary Comput.* **2** 381–407

——1995c Toward a theory of evolution strategies: on the benefit of sex—the $(\mu/\mu, \lambda)$-theory *Evolutionary Comput.* **3** 81–111

——1996a Towards a theory of evolution strategies: self-adaptation *Evolutionary Comput.* **3** 311–47

——1996b An alternative explanation for the manner in which genetic algorithms operate *Biosystems* at press

Bronstein I N and Semendjajew K A 1981 *Taschenbuch der Mathematik* (Leipzig: Teubner)
David H A 1970 *Order Statistics* (New York: Wiley)
Deb K and Goldberg D E 1994 Sufficient conditions for deceptive and easy binary functions *Ann. Math. Artificial Intell.* **10** 385–408
Fogel D B 1992 *Evolving Artificial Intelligence* PhD Thesis, University of California, San Diego
——1994 Asymptotic convergence properties of genetic algorithms and evolutionary programming: analysis and experiments *Cybernet. Syst.* **25** 389–408
——1995 *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence* (Piscataway, NJ: IEEE)
Goldberg D 1989 *Genetic Algorithms in Search, Optimization, and Machine Learning* (Reading, MA: Addison-Wesley)
Grötschel M, Lovász L and Schrijver A 1993 *Geometric Algorithms and Combinatorial Optimization* 2nd edn (Berlin: Springer)
Hansen P and Simeone B 1986 Unimodular functions *Discrete Appl. Math.* **14** 269–81
Horn J, Goldberg D E and Deb K 1994 Long path problems *Parallel Problem Solving from Nature—PPSN III (Proc. Int. Conf. on Evolutionary Computation and 3rd Conf. on Parallel Problem Solving from Nature, Jerusalem, October 1994) (Lecture Notes in Computer Science 866)* ed Yu Davidor, H-P Schwefel and R Männer (Berlin: Springer) pp 149–58
Iosifescu M 1980 *Finite Markov Processes and Their Applications* (Chichester: Wiley)
Johnson D S, Papadimitriou C H and Yannakakis M 1988 How easy is local search? *J. Comput. Syst. Sci.* **37** 79–100
Mühlenbein H 1992 How genetic algorithms really work: mutation and hillclimbing *Parallel Problem Solving from Nature, 2 (Proc. 2nd Int. Conf. on Parallel Problem Solving from Nature, Brussels, 1992)* ed R Männer and B Manderick (Amsterdam: North-Holland) pp 15–25
Mühlenbein H and Schlierkamp-Voosen D 1993 Predictive models for the breeder genetic algorithm *Evolutionary Comput.* **1** 25–49
Rechenberg I 1994 *Evolutionsstrategie '94* (Stuttgart: Frommann-Holzboog)
Rudolph G 1996 How mutation and selection solve long path problems in polynomial expected time *Evolutionary Comput.* **4** at press
Schwefel H-P 1995 *Evolution and Optimum Seeking* (New York: Wiley)

Further reading

1. Arnold B C, Balakrishnan N and Nagaraja H N 1992 *A First Course in Order Statistics* (New York: Wiley)

   As does the book of David (1970), this course gives a good introduction into order statistics, which builds the mathematical basis for truncation selection.

2. Beyer H-G 1992 *Towards a Theory of 'Evolution Strategies'. Some Asymptotical Results from the* $(1 \dot{+} \lambda)$*-Theory* Department of Computer Science Technical Report SYS-5/92, University of Dortmund

   In this report the derivations for the $(1 \dot{+} \lambda)$ theory on noisy fitness data can be found.

3. Beyer H-G 1994 *Towards a Theory of 'Evolution Strategies': Results from the* N*-dependent* (μ, λ) *and the Multi-Recombinant* $(\mu/\mu, \lambda)$ *Theory* Department of Computer Science Technical Report SYS-5/94, University of Dortmund

   This report contains the 'hairy details' of the progress rate theory for (μ, λ) and $(\mu/\mu, \lambda)$ ESs as well as the derivations for the differential geometry approach.

4. Beyer H-G 1995 *Towards a Theory of 'Evolution Strategies': the* $(1, \lambda)$*-Self-Adaptation* Department of Computer Science Technical Report SYS-1/95, University of Dortmund

   This report is devoted to the theory of $(1, \lambda)$ σ selfadaptation and contains the derivations of the results presented in the article by Beyer (1996).

5. Michod R E and Levin B R (eds) 1988 *The Evolution of Sex: an Examination of Current Ideas* (Sunderland, MA: Sinauer)

   Concerning the current ideas on the benefits of recombination in biology, this book reflects the different hypotheses on the evolution of sex. Biological arguments and theories should receive more attention within the EA theory.

Handbook of Evolutionary Computation

Editors in Chief

Thomas Bäck, David B Fogel
and Zbigniew Michalewicz

INSTITUTE OF PHYSICS PUBLISHING
Bristol Philadelphia

and

OXFORD UNIVERSITY PRESS
New York Oxford

1997

INSTITUTE OF PHYSICS PUBLISHING
Bristol Philadelphia

and

OXFORD UNIVERSITY PRESS
Oxford New York
Athens Auckland Bangkok Bogotá
Bombay Buenos Aires Calcutta Cape Town
Dar es Salaam Delhi Florence Hong Kong Istanbul
Karachi Kuala Lumpur Madras Madrid Melbourne
Mexico City Nairobi Paris Singapore
Taipei Tokyo Toronto

and associated companies in

Berlin Ibadan

Published by Institute of Physics Publishing,
Techno House, Redcliffe Way, Bristol BS1 6NX, United Kingdom
(US Editorial Office: The Public Ledger Building, Suite 1035, 150 South Independence Mall West,
Philadelphia, PA 19106, USA)
and Oxford University Press, Inc., 198 Madison Avenue, New York, NY 10016, USA

Oxford is a registered trademark of Oxford University Press

British Library Cataloguing-in-Publication Data and
Library of Congress Cataloging-in-Publication Data are available

ISBN 0 7503 0392 1

This handbook is a joint publication of Institute of Physics Publishing
and Oxford University Press

PROJECT STAFF

INSTITUTE OF PHYSICS PUBLISHING

Publisher: Robin Rees
Project Editor: Sarah Hood
Production Editor: Martin Beavis
Production Manager: Sharon Toop
Assistant Production Manager: Jenny Troyano
Production Assistant: Sarah Plenty
Electronic Production Manager: Tony Cox

OXFORD UNIVERSITY PRESS

Senior Editor: Sean Pidgeon
Project Editor: Matthew Giarratano
Editorial Assistant: Merilee Johnson
Cover Design: Joan Greenfield

Printing (last digit): 9 8 7 6 5 4 3 2 1
Printed in the United Kingdom

♾™ The paper used in this publication meets the minimum requirements
of American National Standard for Information Sciences – Permanence of Paper
for Printed Library Materials, ANSI Z39.48-1984

Contents