Proceedings of the 2007 IEEE Symposium on
Foundations of Computational Intelligence (FOCI 2007)

When the Plus Strategy Outperforms the Comma Strategy—and When Not

Jens Jägersküpper and Tobias Storch
Department of Computer Science 2
University of Dortmund
44221 Dortmund, Germany
{jens.jaegerskuepper|tobias.storch}@cs.uni-dortmund.de

***Abstract*— Occasionally there have been long debates on whether to use elitist selection or not. In the present paper the simple (1,λ) EA and (1+λ) EA operating on $\{0,1\}^n$ are compared by means of a rigorous runtime analysis. It turns out that only values for λ that are logarithmic in n are interesting. An illustrative function is presented for which newly developed proof methods show that the (1,λ) EA—where λ is logarithmic in n—outperforms the (1+λ) EA for any λ. For smaller offspring populations the (1,λ) EA is inefficient on every function with a unique optimum, whereas for larger λ the two randomized search heuristics behave almost equivalently.**

I. Introduction

Evolutionary algorithms (EAs) belong to the broad class of general randomized search heuristics. Their area of application is as huge as their variety and they have been applied in numerous situations successfully. Among the best-known and simplest EAs are the (μ+λ) EA and (μ,λ) EA [1]. The "μ" indicates that a parent population of size μ is used, whereas "λ" denotes the application of an offspring population of size λ. Whether the elements of the descendant population are selected either from the parent and offspring population or from the offspring population only is indicated by "+" and "," respectively. Thus, for the comma strategy necessarily $\lambda \geq \mu$ (for $\lambda = \mu$, there is actually no selection).

Runtime analysis started with very simple EAs such as the (1+1) EA on example functions [2], [3]. Nowadays, one is able to analyze its runtime on practically relevant problems such as the maximum matching problem [4]. However, for more complex EAs and (typical) example functions, the effects of applying either a larger offspring or a large parent population size were investigated theoretically [5], [6]. In this paper, we aim at a systematic comparison of the plus and the comma strategy with respect to the offspring population size. These investigations improve our ability to choose an appropriate selection method, which has been debated a long time. Furthermore, they contribute to the discussion on the effects of selection pressure in evolutionary computation. In order to concentrate on these effects we consider simple EAs that allow for a rigorous analysis, but avoid unnecessary complications due to the effects of other EA components. Here we consider the maximization of pseudo-Boolean objective (fitness) functions $f : \{0,1\}^n \to \mathbb{R}$, $n \in \mathbb{N}$. We investigate the following optimization heuristics, known as (1+λ) EA and (1,λ) EA, using a parent population of size one and standard bit-mutation "mutate$_{1/n}(x)$", where each bit of $x \in \{0,1\}^n$ is flipped independently with probability $1/n$, cf. [1], [5].

(1+λ) EA and **(1,λ) EA**

1) Set $t := 1$ and choose $x_t \in \{0,1\}^n$ uniformly at random.
2) Set $t := t+1$ and let
   $y_{t,1} := \text{mutate}_{1/n}(x_{t-1}), \ldots, y_{t,\lambda} := \text{mutate}_{1/n}(x_{t-1})$.
3) Choose $y_t \in \{y_{t,1}, \ldots, y_{t,\lambda}\}$ arbitrarily among all elements with largest f-value.
   (1+λ) EA: If $f(y_t) \geq f(x_{t-1})$, then set $x_t := y_t$, else set $x_t := x_{t-1}$.
   (1,λ) EA: Set $x_t := y_t$.
4) Goto 2.

The number of f-evaluations which are performed until the $t(n)^{\text{th}}$ step is completed, equals $1+\lambda\cdot(t(n)-1)$ for $t(n) \geq 1$. In contrast to the (1+λ) EA, the (1,λ) EA occasionally accepts an element that is worse than the previous one (unless the function to be optimized is constant). This can avoid stagnation in local optima. However, it may also cause a slow(er) movement towards global optima. It was often argued that *the difference between an elitist* (1+λ) *EA and a non-elitist* (1,λ) *EA is less important in* $\{0,1\}^n$, e.g. [5]. Here we will point out in detail when this is correct—but we also demonstrate when this is definitely not the case. More precisely, in Section III we show that a comparison of plus and comma strategy is interesting in particular for offspring populations of size λ with $\ln(n)/14 < \lambda = \mathcal{O}(\ln n)$. Investigating $\lambda = 1$ for the (1+λ) EA does make sense, but for the (1,λ) EA it does not: For any $\lambda \leq \ln(n)/14$, the comma strategy indeed fails to optimize any function with a unique global optimum. Furthermore, for $\lambda = \omega(\ln n)$ (i.e., $\ln n/\lambda \to 0$ as $n \to \infty$) it is rather unlikely to observe any difference in the populations of the (1+λ) EA and the (1,λ) EA in a polynomial number of steps. These observations are applied and extended for a simple (unimodal) example function and (asymptotically) tight bounds on the heuristics' runtimes are obtained. In Section IV we extend the well-known proof technique of f-based partitions such that it can be applied to (1+λ) EA and (1,λ) EA. With the help of this method we demonstrate the algorithms' different strengths and weaknesses. Namely, for a simple (multimodal) example function we apply the method and demonstrate the possible major disadvantage of the plus

strategy compared to the comma strategy. The runtime bounds to be presented are again tight. We finish with a summary and some conclusions in Section V and continue with some preliminaries in the following Section II.

II. Preliminaries

The efficiency of a randomized algorithm is (usually) measured in the following way. For $\star \in \{\text{“}+\text{”};\text{“},\text{”}\}$ let $T^{\star}_{f_n,\lambda}(x)$ denote the random variable which corresponds to the number of function evaluations—the runtime—of the $(1\star\lambda)$ EA to create an optimum of $f_n : \{0,1\}^n \to \mathbb{R}$, $n \in \mathbb{N}$, for the first time, where the initial element is $x \in \{0,1\}^n$. (We can ignore a stopping criterion and analyze an infinite stochastic process.) If for a sequence of functions $f = (f_1, \ldots, f_n, \ldots)$ the expected runtime of the $(1\star\lambda)$ EA to optimize f_n, namely $\sum_{x\in\{0,1\}^n} \mathrm{E}[T^{\star}_{f_n,\lambda}(x)]/2^n$ (since the initial element x is chosen uniformly at random), is bounded by a polynomial in n, then we call the $(1\star\lambda)$ EA *efficient* on f, whereas we call it *totally inefficient* if the probability that an optimum is created remains exponentially small even after an exponential number of steps. In this case, a polynomially bounded number of (parallel) independent multistarts of the algorithm is still totally inefficient. For the notations on asymptotics see [7].

III. Small and Large Offspring Populations

A. Small Offspring Populations

We take a closer look at the smallest possible offspring population size. On the one hand, the (1+1) EA (can reasonably be applied) optimizes any function in an expected runtime $\mathcal{O}(n^n)$, and functions are known where it needs an expected runtime $\Theta(n^n)$. On the other hand, the (1+1) EA optimizes any linear function in an expected runtime $\mathcal{O}(n \ln n)$, and it needs an expected runtime $\Theta(n \ln n)$ if the linear function has a unique global optimum [2]. In contrast, the (1,1) EA (cannot reasonably be applied) optimizes any function in an expected runtime $\mathcal{O}(2^n)$, but it also needs an expected runtime $\Theta(2^n)$ if the function has a unique global optimum [3]. This is because the closer the search point x_t of the (1,1) EA is to the unique optimum, the larger the probability of x_{t+1} to be located farther away from the optimum. Let us consider the (1,λ) EA with larger offspring populations, yet $\lambda \le \ln(n)/14$. We demonstrate that a strong drift away from the optimum still exists. Namely, if x_t is reasonable close to the optimum, then with a large probability all elements in the offspring population are even farther away from the optimum. Since comma selection is applied, one of these elements becomes x_{t+1}. Thus, it is time-consuming to create the optimum.

Theorem 1 *Given a function $f : \{0,1\}^n \to \mathbb{R}$ with a unique global optimum $x' \in \{0,1\}^n$ and $\lambda \le \varepsilon(n)\ln(n)/7$ with $\varepsilon(n) \in [7/\ln n, 1/2]$, with probability $1 - 2^{-\Omega(n^{1-\varepsilon(n)})}$ the* (1,λ) EA *needs a runtime larger than $2^{n^{1-\varepsilon(n)}}$ to optimize f.*

With $\varepsilon(n) := 7/\ln n$ we obtain for the (1,1) EA a lower bound on the runtime to optimize f of $2^{\Omega(n)}$ which holds not only in expectation (cf. the result in [3]) yet also with an overwhelming probability. Even the $(1,\lfloor \ln(n)/14 \rfloor)$ EA (i.e., we choose $\varepsilon(n) := 1/2$) is still totally inefficient.

To prove Theorem 1, we recall a result on Markov processes (and drift). A Markov process $\mathcal{M}$ on $m < \infty$ states, namely $0, \ldots, m-1$, is described by a stochastic $m\times m$-matrix P of transition probabilities ($P_{i,j}$, $0 \le i,j \le m-1$: probability to transit from state i to state j) and a stochastic row vector $p \in [0,1]^m$ of initialization probabilities (p_i, $0 \le i \le m-1$: probability to initialize in state i). The ith entry of the stochastic row vector pP^{t-1} corresponds to the probability of $\mathcal{M}$ being in state i after the tth step for $0 \le i \le m-1$ and $t \ge 1$. For more detailed investigations of Markov processes see [7]. The following result was proven in [4] and goes back to a result in [8].

Lemma 2 *Given a Markov process $\mathcal{M}$ on m states, a state $\ell \in \{0, \ldots, m-1\}$, and $\alpha(\ell), \beta(\ell), \gamma(\ell) > 0$, if*
1) $\sum_{j=0}^{m-1} p_{i,j} \cdot e^{-\alpha(\ell)\cdot(j-i)} \le 1 - 1/\beta(\ell) \quad \forall i \in \{0, \ldots, \ell\}$,
2) $\sum_{j=0}^{m-1} p_{i,j} \cdot e^{-\alpha(\ell)\cdot(j-\ell)} \le 1 + \gamma(\ell) \quad \forall i \in \{\ell, \ldots, m-1\}$,

then the 0th entry of the m-vector pP^{t-1} is bounded from above by $t \cdot e^{-\alpha(\ell)\cdot\ell} \cdot \beta(\ell) \cdot (1+\gamma(\ell)) + \sum_{j=0}^{\ell-1} p_j$.

In the following we prove Theorem 1:

Proof: The runtime is larger than $1 + \lambda \cdot (t(n) - 1) \ge t(n)$, if the unique optimum x' is not created in the first $t(n)$ steps. We assume that, once the (1,λ) EA has created x', afterwards x' would be kept forever. Thus, we are interested in the event $x_{\lfloor 2^{n^{1-\varepsilon(n)}} \rfloor} = x'$. If its probability is $2^{-\Omega(n^{1-\varepsilon(n)})}$, then we obtain the claimed result.

We describe a Markov process $\mathcal{M}$ on $n+1$ states with the following property. At least with the same probability, $\mathcal{M}$ is in a state $i, \ldots, n$ after $t(n)$ steps as the (1,λ) EA generates an element $x_{t(n)}$ with Hamming distance $\mathrm{H}[x_{t(n)}, x'] \ge i$ from the optimum x'. If this holds for all $i \in \{0, \ldots, n\}$, the (1,λ) EA generates the optimum at most with the same probability as $\mathcal{M}$ reaches state 0 (in a given number of steps). With $p_i := \binom{n}{i}/2^n$, $\mathcal{M}$ has the desired property for $t(n) = 1$ (even equality holds). If $\mathcal{M}$ is in state i after $t(n)$ steps, with at least the same probability holds $\mathrm{H}[x_{t(n)}, x'] \ge i$ for the (1,λ) EA with $x_{t(n)}$. In this situation, assume that the probability of creating $x_{t(n)+1}$ is bounded above by $p_{i,\le j}$, where $\mathrm{H}[x_{t(n)+1}, x'] \le j$. Ensuring $p_{i,0} + \cdots + p_{i,j} \ge p_{i,\le j}$ is sufficient, so that $\mathcal{M}$ has the desired property also in the following step. For $j < i$ we set $p_{i,j}$ to (at least) the maximum of the probabilities that an $x_{t(n)}$ generates $x_{t(n)+1}$ with $\mathrm{H}[x_{t(n)+1}, x'] = j$, so that the inequality holds for $j < i$. We set $p_{0,0} := 1$ and the inequality holds for $i = 0$. We set $p_{i,i+1}$ for $i \ge 1$ to (at most) the minimum of the probabilities that $x_{t(n)}$ generates $x_{t(n)+1}$ with $\mathrm{H}[x_{t(n)+1}, x'] \ge i+1$. Moreover, $p_{i,j} := 0$ for $j \ge i+2$ a well as $p_{i,i} := 1 - p_{i,i+1} - \sum_{j=0}^{i-1} p_{i,j}$, so that the inequality holds for $j \ge i$ and $i \ge 1$.

In order to apply Lemma 2 for $\mathcal{M}$ with $\ell := \lfloor n^{1-\varepsilon(n)} \rfloor$, $\alpha(\ell) := 6/5$, $\beta(\ell) := 32\ell$, and $\gamma(\ell) := 1$, we have to prove that the following two conditions are fulfilled.
1) $\sum_{j=0}^{n} p_{i,j} \cdot e^{-(6/5)\cdot(j-i)} \le 1 - 1/(32\lfloor n^{1-\varepsilon(n)} \rfloor)$ for all $i \in \{1, \ldots, \lfloor n^{1-\varepsilon(n)} \rfloor - 1\}$.

Firstly, we consider $j < i$ and an element x with $\mathrm{H}[x, x'] = i + k$, $0 \le k \le n - i$. In order to decrease the Hamming distance from the optimum to j, for at least one of its λ offspring, $i + k - j$ out of $i + k$ specific bits must flip. Hence,

$$\begin{aligned} p_{i,j} &\le \max\{\lambda \cdot \tbinom{i+k}{i+k-j}/n^{i+k-j} \mid k \in \{0, \ldots, n-i\}\} \\ &\le \lambda \cdot \tbinom{i}{i-j}/n^{i-j} \ , \end{aligned}$$

since $\binom{i+k}{i+k-j} = \binom{i}{i-j} \cdot \frac{(i+k)\cdots(i+1)}{(j+k)\cdots(j+1)} \le \binom{i}{i-j} \cdot n^k$. Furthermore, with $\binom{i}{i-j} \le i^{i-j}$ it holds

$$\begin{aligned} &\lambda \cdot \tbinom{i}{i-j}/n^{i-j} \le \lambda \cdot i^{i-j}/n^{i-j} \\ \le\ &\lambda \cdot n^{(1-\varepsilon(n))(i-j)}/n^{i-j} = \lambda \cdot n^{-\varepsilon(n)(i-j)} \ . \end{aligned}$$

Secondly, we consider $j = i + 1$ for $i > 0$ and an element x with $\mathrm{H}[x, x'] = i + k$, $0 \le k \le n - i$.

- For $k = 0$ it is sufficient that each of its λ offspring equals x except for one bit which is flipped such that the Hamming distance to the optimum is increased.
- For $k \ge 1$ it is sufficient, that each of its λ offspring is a duplicate of x.

In these cases, the Hamming distance to the optimum is at least $i + 1$. Hence,

$$\begin{aligned} p_{i,i+1} &\ge \min\{\tbinom{n-i}{1} \cdot \tfrac{1}{n}\left(1 - \tfrac{1}{n}\right)^{n-1}, \left(1 - \tfrac{1}{n}\right)^n\}^\lambda \\ &\ge (6/17)^\lambda \end{aligned}$$

because $(1-1/n)^n \ge 6/17$ and $(n-i)/n \cdot (1-1/n)^{n-1} \ge (n - n^{1-7/\ln n})/(en) \ge 6/17$. Furthermore, using the fact that $\ln(6/17) \cdot 1/7 \ge -1/6$, we have

$$(6/17)^\lambda \ge (6/17)^{\varepsilon(n)\ln(n)/7} \ge n^{-\varepsilon(n)/6} \ .$$

It remains to prove that $\sum_{j=0}^{i-1} \lambda \cdot n^{-\varepsilon(n)(i-j)} e^{(6/5)\cdot(i-j)} + (1 - n^{-\varepsilon(n)/6} - \sum_{j=0}^{i-1} \lambda \cdot n^{-\varepsilon(n)(i-j)}) \cdot e^{(6/5)\cdot 0} + n^{-\varepsilon(n)/6} \cdot e^{(6/5)\cdot(-1)} \le 1 - n^{-(1-\varepsilon(n))}/32$ for $0 < i < \lfloor n^{1-\varepsilon(n)} \rfloor$. By an index transformation and due to the convergence property of infinite geometric series

$$\sum_{j=1}^{i} \lambda \cdot n^{-\varepsilon(n)j} e^{\frac{6j}{5}} \le \lambda \sum_{j=1}^{\infty} (n^{-\varepsilon(n)} e^{\frac{6}{5}})^j = \frac{\lambda}{n^{\varepsilon(n)} e^{-\frac{6}{5}} - 1}$$

follows. Furthermore, with $\lambda \le \varepsilon(n)\ln(n)/7$ it holds

$$\frac{\lambda}{n^{\varepsilon(n)} e^{-6/5} - 1} \le \frac{\varepsilon(n)\ln(n)}{7 n^{\varepsilon(n)} e^{-6/5} - 7} \le \frac{2n^{-\varepsilon(n)/6}}{3}$$

since $\varepsilon(n) \ge 7/\ln n$ and furthermore, with $e^{5\varepsilon(n)\ln(n)/6} \ge 1 + 5\varepsilon(n)\ln(n)/6$ it is

$$\begin{aligned} 0 &\le \varepsilon(n)\ln(n) \cdot (\frac{14}{3} \cdot \frac{5}{6} \cdot e^{-\frac{6}{5}} - 1) + \frac{14}{3}(e^{-\frac{6}{5}} - e^{-\frac{7}{6}}) \\ &\le \frac{14}{3} \cdot e^{-\frac{6}{5}} \cdot n^{\frac{5\varepsilon(n)}{6}} - \frac{14}{3} \cdot n^{-\frac{\varepsilon(n)}{6}} - \varepsilon(n)\ln(n) \\ &= \frac{2n^{-\frac{\varepsilon(n)}{6}}}{3} \cdot (7 n^{\varepsilon(n)} e^{-\frac{6}{5}} - 7) - \varepsilon(n)\ln(n) \ . \end{aligned}$$

Since

$$1 - n^{-\varepsilon(n)/6} - \sum_{j=1}^{i-1} \lambda \cdot n^{-\varepsilon(n)j} \le 1 - n^{-\varepsilon(n)/6} \ ,$$

the inequality mentioned above is fulfilled with $2n^{-\varepsilon(n)/6}/3 + (1 - n^{-\varepsilon(n)/6}) + n^{-\varepsilon(n)/6} e^{-6/5} \le 1 - n^{-\varepsilon(n)/6}/32 \le 1 - n^{-(1-\varepsilon(n))/6}/32$ since $\varepsilon(n) \le 1/2$.

2) $\sum_{j=0}^{n} p_{i,j} \cdot e^{-(6/5)\cdot(j - \lfloor n^{1-\varepsilon(n)} \rfloor)} \le 2$ for all $i \in \{\lfloor n^{1-\varepsilon(n)} \rfloor, \ldots, n\}$.
Similar to the proof that the first condition is met, we have $p_{i,j} \le \lambda \cdot n^{-\varepsilon(n)(\lfloor n^{1-\varepsilon(n)} \rfloor - j)}$ for $j < i$. Thus, $\sum_{j=0}^{\lfloor n^{1-\varepsilon(n)} \rfloor - 1} \lambda \cdot n^{-\varepsilon(n)(\lfloor n^{1-\varepsilon(n)} \rfloor - j)} e^{(6/5)\cdot(\lfloor n^{1-\varepsilon(n)} \rfloor - j)} + (1 - \sum_{j=0}^{\lfloor n^{1-\varepsilon(n)} \rfloor - 1} \lambda \cdot n^{-\varepsilon(n)(\lfloor n^{1-\varepsilon(n)} \rfloor - j)}) \cdot e^{(6/5)\cdot 0} \le 2n^{-\varepsilon(n)/6}/3 + 1 \le 2$ for $\lfloor n^{1-\varepsilon(n)} \rfloor \le i \le n$.

To apply Lemma 2 we must finally estimate $\sum_{j=0}^{\ell-1} p_j$. Since $\ell \le \lfloor n^{1-\varepsilon(n)} \rfloor \le \lfloor n/e^7 \rfloor$ with $\varepsilon(n) \ge 7/\ln n$,

$$\begin{aligned} &\sum_{j=0}^{\ell-1} \tbinom{n}{j}/2^n \le \sum_{j=0}^{\lfloor n/e^7 \rfloor - 1} \tbinom{n}{j}/2^n \le n \cdot \tbinom{n}{\lfloor n/e^7 \rfloor}/2^n \\ &\le n \cdot \left(\frac{en}{n/e^7}\right)^{n/e^7}/2^n = e^{\ln n + 8n/e^7 - n\ln 2} \le e^{-n/3} \ . \end{aligned}$$

Now, applying Lemma 2 with $t = \lfloor 2^{n^{1-\varepsilon(n)}} \rfloor$ leads to a probability of at most $\lfloor 2^{n^{1-\varepsilon(n)}} \rfloor \cdot e^{-(6/5)\cdot\lfloor n^{1-\varepsilon(n)} \rfloor} \cdot 32\lfloor n^{1-\varepsilon(n)} \rfloor \cdot (1+1) + e^{-n/3} = 2^{-\Omega(n^{1-\varepsilon(n)})}$ that $\mathcal{M}$ reaches state 0 in the first $\lfloor 2^{n^{1-\varepsilon(n)}} \rfloor$ steps. ■

B. Large Offspring Populations

With an offspring population size λ of any appreciable size, the (1+λ) EA and the (1,λ) EA will not differ significantly in the way they search $\{0,1\}^n$. This was claimed in [5] since in this situation *... the offspring population will almost surely contain at least one exact copy of the parent.* We extend this statement and make it more precise in the following. Therefore, let f be a function and for $t(n) \ge 1$ let

$$\begin{aligned} s_{t(n)} \ := \ & x_1, y_{2,1}, \ldots, y_{2,\lambda}, y_2, \quad x_2, y_{3,1}, \ldots, y_{3,\lambda}, y_3, \\ & \ldots, \quad x_{t(n)}, y_{t(n)+1,1}, \ldots, y_{t(n)+1,\lambda}, y_{t(n)+1} \end{aligned}$$

be a sequence of $(\lambda + 2) \cdot t(n)$ elements from $\{0,1\}^n$. The $(1\star\lambda)$ EA *observes* $s_{t(n)}$ (while optimizing f) if with positive probability the following holds: The elements $x_1, \ldots, x_{t(n)}$ appear as the first $t(n)$ parents and, for $t \in \{2, \ldots, t(n)+1\}$, y_t can appear as the selected offspring out of the λ offspring $y_{t,1}, \ldots, y_{t,\lambda}$ of x_{t-1}. We consider a sequence $s_{t(n)}$ observed by the $(1\star\lambda)$ EA. Recall that the $(1\star\lambda)$ EA and the $(1\bar{\star}\lambda)$ EA differ only in step 3, where $\bar{\star}$ denotes the other of the two symbols $\{\text{“}+\text{”}; \text{“},\text{”}\}$.

If $\forall\, t \in \{2, \ldots, t(n)\} : f(y_t) \ge f(x_{t-1})$, then the condition in step 3 is always fulfilled. The $(1\star\lambda)$ EA and $(1\bar{\star}\lambda)$ EA observe with equal probability: the same x_1 *and* with the same parent x_{t-1} the same sequence of offspring $y_{t,1}, \ldots, y_{t,\lambda}$ *and* even the selected offspring y_t is determined identically. Thus, the $(1\star\lambda)$ EA and $(1\bar{\star}\lambda)$ EA observe the sequence $s_{t(n)}$

with equal probability while optimizing f. The set of these sequences is denoted by $S_{\dot{+},f,t(n)}$.

If $\exists\, t \in \{2,\ldots,t(n)\} : f(y_t) < f(x_{t-1})$, then the condition in step 3 is in step t not fulfilled. In case $f(y_t) < f(x_{t-1})$ also $y_t \neq x_{t-1}$ and the (1⋆λ) EA and (1$\bar{\star}\lambda$) EA surely select different elements to be x_t. Thus, the sequence $s_{t(n)}$ is not observed by the (1$\bar{\star}\lambda$) EA while optimizing f. The set of these sequences is denoted by $S_{\star,f,t(n)}$.

We bound the probability to observe a sequence of $S_{\dot{+},f,t(n)}$ by the (1+λ) EA and (1,λ) EA. If at least one of the offspring $y_{t,1},\ldots,y_{t,\lambda}$ is a duplicate of its parent x_{t-1}, then necessarily $f(y_t) \geq f(x_{t-1})$. Its probability is bounded from below by $1-\big(1-(1-1/n)^n\big)^\lambda \geq 1-\big(1-6/17\big)^\lambda = 1-(11/17)^\lambda$ for n large enough. With probability at most $(t(n)-1)\cdot(11/17)^\lambda$, this does not happen for at least one $t \in \{2,\ldots,t(n)\}$.

Lemma 3 *Given $f : \{0,1\}^n \to \mathbb{R}$ and n large enough, with probability at least $1-(t(n)-1)\cdot(11/17)^\lambda$ the* (1+λ) EA *as well as the* (1,λ) EA (*with an arbitrary x_1*) *observe a sequence from $S_{\dot{+},f,t(n)}$ for $t(n) \geq 1$.*

This lemma helps to transfer success probabilities and even expectation values for optimization from the (1+λ) EA to the (1,λ) EA and vice versa. In particular, when the offspring population is large – with respect to the period considered.

We consider a runtime of $\ell(n)$ with $1 \leq \ell(n) \leq 1+\lambda$, i.e. at most two steps. For any λ, the (1⋆λ) EA and (1$\bar{\star}\lambda$) EA optimize a function f within the first $\ell(n)$ function evaluations with equal probability in this case.

We consider a runtime $\ell(n)$ with $2+\lambda\cdot(t(n)-1) \leq \ell(n) \leq 1+\lambda\cdot t(n)$ for $t(n) \geq 2$, i.e. at most $t(n)+1$ steps. Let $\mathcal{E}^\star$ be the event that the (1⋆λ) EA has *not* optimized the function f in the first $\ell(n)$ function evaluations. This event occurs iff the (1⋆λ) EA observes a sequence $s_{t(n)}$ where all $\ell(n)$ elements $x_1, y_{2,1},\ldots,y_{2,\lambda}, y_{3,1},\ldots,y_{3,\lambda},\ldots,y_{t(n)+1,1},\ldots,$ $y_{t(n)+1,\ell(n)-\lambda\cdot(t(n)-1)-1}$ are non-optimal. We decompose $\mathcal{E}^\star$ into two disjoint events $\mathcal{E}_1^\star$, that $s_{t(n)} \in S_{\dot{+},f,t(n)}$, and $\mathcal{E}_2^\star$, that $s_{t(n)} \notin S_{\dot{+},f,t(n)}$, i.e., $s_{t(n)} \in S_{\star,f,t(n)}$. As we have seen, each sequence from $S_{\dot{+},f,t(n)}$ occurs with the same probability for the (1⋆λ) EA and (1$\bar{\star}\lambda$) EA. Thus, $\Pr[\mathcal{E}_1^\star] = \Pr[\mathcal{E}_1^{\bar{\star}}]$ and hence, $\Pr[\mathcal{E}^{\bar{\star}}] = \Pr[\mathcal{E}_2^{\bar{\star}}] + \Pr[\mathcal{E}_1^{\bar{\star}}] = \Pr[\mathcal{E}_2^{\bar{\star}}] + \Pr[\mathcal{E}_1^\star] = \Pr[\mathcal{E}_2^{\bar{\star}}] + \Pr[\mathcal{E}^\star] - \Pr[\mathcal{E}_2^\star]$.

Consider $\lambda \geq (5/2)\cdot(1+c(n))\cdot\ln t(n)$, where $c(n) \geq 0$. By Lemma 3 the (1⋆λ) EA and (1$\bar{\star}\lambda$) EA observe a sequence from $S_{\dot{+},f,t(n)}$ with probability at least $1-(t(n)-1)\cdot(11/17)^\lambda \geq 1-t(n)\cdot(11/17)^{(5/2)\cdot(1+c(n))\cdot\ln t(n)} \geq 1-t(n)\cdot 1/t(n)^{1+c(n)} = 1-1/t(n)^{c(n)}$ since $\ln(11/17)\cdot(5/2) \leq -1$. So, a sequence of $S_{\star,f,t(n)}$ is observed with probability at most $1/t(n)^{c(n)}$ by the (1⋆λ) EA. Since $\mathcal{E}_2^\star$ implies that the (1⋆λ) EA observes a sequence from $S_{\star,f,t(n)}$, $0 \leq \Pr[\mathcal{E}_2^\star]$ and $\Pr[\mathcal{E}_2^{\bar{\star}}] \leq 1/t(n)^{c(n)}$, and hence, $-1/t(n)^{c(n)} \leq \Pr[\mathcal{E}_2^{\bar{\star}}] - \Pr[\mathcal{E}_2^\star] \leq 1/t(n)^{c(n)}$.

Theorem 4 *Let $f : \{0,1\}^n \to \mathbb{R}$, $x \in \{0,1\}^n$ and n large enough be given.*

1) *For $0 \leq \ell(n) \leq 1+\lambda$ holds:*

$$\Pr[T_{f,\lambda}^{\bar{\star}}(x) > \ell(n)] = \Pr[T_{f,\lambda}^{\star}(x) > \ell(n)]$$

2) *For $2+\lambda\cdot(t(n)-1) \leq \ell(n) \leq 1+\lambda\cdot t(n)$, where $t(n) \geq 2$ and $\lambda \geq (5/2)\cdot(1+c(n))\cdot\ln t(n)$, $c(n) \geq 0$, holds:*

$$\begin{aligned}\Pr[T_{f,\lambda}^{\bar{\star}}(x) > \ell(n)] &\leq \Pr[T_{f,\lambda}^{\star}(x) > \ell(n)] + 1/t(n)^{c(n)}\\ \Pr[T_{f,\lambda}^{\bar{\star}}(x) > \ell(n)] &\geq \Pr[T_{f,\lambda}^{\star}(x) > \ell(n)] - 1/t(n)^{c(n)}\end{aligned}$$

The next section shows an exemplary application of this result.

C. Application to ONEMAX

Let us investigate one of the best-known functions, namely ONEMAX : $\{0,1\}^n \to \mathbb{R}$, where

$$\textsc{OneMax}(x) := |x| \ .$$

Even its analogue in continuous search spaces is well-studied, e.g. in [9]. Part 1 of the following theorem was proven in [5] (it even holds for the (1+λ) EA with an arbitrarily fixed x_1). Let us consider a phase of

$$\lceil 3\max\{\mathrm{E}[T^+_{\textsc{OneMax},\lambda}(x)] \,|\, x \in \{0,1\}^n\}/\lambda + 6^7\rceil =: E_\lambda$$

steps, each creating λ offspring. By Markov's inequality [7] the (1+λ) EA does not create the optimum in such a phase with probability at most $\mathrm{E}[T^+_{\textsc{OneMax},\lambda}(x)]/(\lambda\cdot E_\lambda) \leq 1/3$ for every x. We observe, for $\lambda \geq 3\ln n$ holds $(5/2)\cdot(1+1/7)\cdot\ln E_\lambda \leq \lambda$ since $E_\lambda \leq cn$ for an appropriate large constant c. Hence, by Theorem 4.1, the (1,λ) EA does not create the optimum in such a phase with probability at most $1/3+1/E_\lambda^{1/7} \leq 1/2$, i.e., with probability at least $1/2$ it does. In the case of a failure we can repeat the argumentation. The expected number of repetitions is upper bounded by 2 and we obtain part 3 of the following theorem since $1+2\cdot\lambda\cdot E_\lambda = \mathcal{O}(\max\{\mathrm{E}[T^+_{\textsc{OneMax},\lambda}(x)] \mid x \in \{0,1\}^n\})$. Finally, part 2 of the following theorem results by Theorem 1 since ONEMAX has the unique optimum 1^n.

Theorem 5

1) *The expected runtime of the* (1+λ) EA *on* ONEMAX *is*
   - $\mathcal{O}(n\ln n)$ *if* $\lambda = \mathcal{O}((\ln n)(\ln\ln n)/\ln\ln\ln n)$, *and*
   - $\mathcal{O}(\lambda n)$ *if* $\lambda = \Omega(\ln n)$.
2) *If $\lambda \leq \varepsilon(n)\ln(n)/7$ for $\varepsilon(n) \in [7/\ln n, 1/2]$, then with probability $1-2^{-\Omega(n^{1-\varepsilon(n)})}$ the* (1,λ) EA *needs a runtime larger than $2^{n^{1-\varepsilon(n)}}$ to optimize* ONEMAX.
3) *If $\lambda \geq 3\ln n$, then the expected runtime of the* (1,λ) EA *on* ONEMAX *is*
   - $\mathcal{O}(n\ln n)$ *if* $\lambda = \mathcal{O}((\ln n)(\ln\ln n)/\ln\ln\ln n)$, *and*
   - $\mathcal{O}(\lambda n)$.

IV. OFFSPRING POPULATIONS WHICH ARE NEITHER LARGE NOR SMALL

We present two proof techniques – one for the (1+λ) EA and one for the (1,λ) EA – which are inspired by the method of f-based partitions from [10]. They demonstrate the different strengths and weaknesses of the two selection strategies. The original method of f-based partitions helps to upper bound the expected runtime of the (1+1) EA to optimize a particular function and is widely applied. Recently, this method was successfully extended for a (μ+1) EA in [11].

Given $f : \{0,1\}^n \to \mathbb{R}$ and $A, B \subseteq \{0,1\}^n$, $A, B \neq \emptyset$, the relation $A <_f B$ holds, iff $f(a) < f(b)$ for all $a \in A$, $b \in B$. We call $A_0, \ldots, A_m$ an *f-based partition*, iff $A_0, \ldots, A_m$ is a partition of $\{0,1\}^n$, $A_0 <_f \cdots <_f A_m$, and A_m contains optima only, i.e., $f(a) = \max\{f(b) \mid b \in \{0,1\}^n\}$ for each $a \in A_m$. Moreover, for $i \in \{0, \ldots, m-1\}$ let $p(a)$, $a \in A_i$, denote the probability that a mutation of a is in $A_{i+1} \cup \cdots \cup A_m$ and $p(A_i) := \min\{p(a) \mid a \in A_i\}$, i.e., $p(A_i)$ is a lower bound on the probability to leave A_i with a mutation.

A. (1+λ) EA

For the (1+λ) EA to leave A_i, $i < m$, once and for all, at least one of the λ offspring must be in $A_{i+1} \cup \cdots \cup A_m$.

Lemma 6 *Given $f : \{0,1\}^n \to \mathbb{R}$ and an f-based partition $A_0, \ldots, A_m$, let*

$$p_i^+ := 1 - (1 - p(A_i))^\lambda \text{ for } i \in \{0, \ldots, m-1\} \ .$$

The (1+λ) EA (*with an arbitrarily initialized x_1*) *optimizes f in an expected runtime of at most*

$$1 + \lambda \cdot \Big(\frac{1}{p_0^+} + \cdots + \frac{1}{p_{m-1}^+}\Big) \ .$$

Proof: We describe a Markov process $\mathcal{M}$ on $m+1$ states with the following property. The probability that $\mathcal{M}$ is in a state $i, \ldots, m$ after $t(n)$ steps is at most the probability that the (1+λ) EA generates an element $x_{t(n)}$ with $x_{t(n)} \in A_i \cup \cdots \cup A_m$. If this holds for all $i \in \{0, \ldots, m\}$, the (1+λ) EA generates at least with the same probability an optimum as $\mathcal{M}$ reaches state m (in a given number of steps). We set $p_0 := 1$ and $p_i := 0$ for $i \geq 1$, so that $\mathcal{M}$ has the claimed property for $t(n) = 1$ and arbitrary x_1. If $\mathcal{M}$ is in state i after $t(n)$ steps, with at least the same probability $x_{t(n)} \in A_i \cup \cdots \cup A_m$ for the (1+λ) EA with $x_{t(n)}$. In this situation, it is impossible to create an $x_{t(n)+1} \in A_0 \cup \cdots \cup A_{i-1}$. Moreover, p_i^+ is a lower bound on the probability to create $x_{t(n)+1} \in A_{i+1} \cup \cdots \cup A_m$ since it suffices that at least one of λ offspring is therein. Thus, we set $p_{i,j} := 0$ for $0 \leq j < i \leq m$ and $i+2 \leq j \leq m$, $p_{i,i+1} := p_i^+$, and $p_{i,i} := 1 - p_i^+$ for $0 \leq i < m$, and $p_{m,m} := 1$. This ensures that $\mathcal{M}$ has the desired property also in the following step.

The expected number of steps to move from state i to state m equals $E_i := 1 + p_i^+ \cdot E_{i+1} + (1 - p_i^+) \cdot E_i = 1/p_i^+ + E_{i+1}$ for $i \in \{0, \ldots, m-1\}$, and $E_m = 0$. Thus, E_0 equals $1/p_0^+ + \cdots + 1/p_{m-1}^+$. With the initialization and the λ function evaluations in each further step, the (1+λ) EA optimizes f in an expected runtime of at most $1 + \lambda \cdot E_0$. ■

For $\lambda = 1$ we obtain the original result for the (1+1) EA presented in [10]. We apply this method exemplarily to ONEMAX. We consider the partition $A_0, \ldots, A_n$ with $A_i := \{x \mid |x| = i\}$. Then $p_i^+ \geq 1 - \big(1 - \frac{n-i}{en}\big)^\lambda \geq 1 - e^{-\frac{\lambda(n-i)}{en}} \geq 1 - \frac{1}{1 + \frac{\lambda(n-i)}{en}} = \frac{en + \lambda(n-i)}{\lambda(n-i)}$ (cf. [5]). Hence, by applying Lemma 6, the (1+λ) EA optimizes ONEMAX in an expected runtime of at most $1 + \lambda \sum_{i=0}^{n-1} \frac{en + \lambda(n-i)}{\lambda(n-i)} = \mathcal{O}(n \ln n + \lambda n)$. This already proves a major part of Theorem 5.1.

B. (1,λ) EA

Theorem 4 enables us to easily transfer Lemma 6 to the (1,λ) EA. For steps which do not create a duplicate, we may pessimistically assume that they lead to a disadvantage, or we may optimistically assume that they lead to an advantage, depending on whether we aim at an upper or at a lower bound on the (expected) runtime.

Given $f : \{0,1\}^n \to \mathbb{R}$, let $A_0, \ldots, A_m$ be a (not necessarily f-based) partition of $\{0,1\}^n$ such that A_m consists of optima only. Let the probability that a mutation of $a \in A_i$ generates some b

- in $A_0 \cup \cdots \cup A_{i-1}$ such that $f(b) \geq f(a)$ be denoted by $p^-(a)$,
- in $A_{i+1} \cup \cdots \cup A_m$ such that $f(b) > f(a)$ be denoted by $p^+(a)$.

Thus, $p^-(A_i) := \max\{p^-(a) \mid a \in A_i\}$ is an upper bound on the probability that an offspring is generated such that A_i is (possibly) left, but in the wrong direction (namely $A_0 \cup \cdots \cup A_{i-1}$ is hit) – even if a duplicate of the parent is generated. Moreover, $p^+(A_i) := \min\{p^+(a) \mid a \in A_i\}$ is a lower bound on the probability that an offspring is generated such that A_i is left in the right direction, namely $A_{i+1} \cup \cdots \cup A_m$ is hit.

Lemma 7 *Given $f : \{0,1\}^n \to \mathbb{R}$ and a partition $A_0, \ldots, A_m$ of $\{0,1\}^n$ such that A_m consists of optima only, let for $i \in \{0, \ldots, m\}$*

$$\begin{aligned} p_i^+ &:= \max\{\big(1 - p^-(A_i)\big)^\lambda - \big(1 - p^-(A_i) - p^+(A_i)\big)^\lambda, \\ &\qquad p(A_i)^\lambda\} \text{ and} \\ p_i^- &:= 1 - \big(1 - p^-(A_i)\big)^\lambda + \big(\tfrac{11}{17} - p^-(A_i) - p^+(A_i)\big)^\lambda \ . \end{aligned}$$

The (1,λ) EA (*with an arbitrarily initialized x_1*) *optimizes f in an expected runtime of at most*

$$\begin{aligned} 1 + \lambda \ &\cdot \ \Big(\frac{1}{p_0^+} + \frac{1}{p_1^+ + p_1^-} + \cdots + \frac{1}{p_{m-1}^+ + p_{m-1}^-}\Big) \\ &\cdot \ \frac{p_1^+ + p_1^-}{p_1^+} \cdots \frac{p_{m-1}^+ + p_{m-1}^-}{p_{m-1}^+} \ . \end{aligned}$$

Proof: Note that p_i^+ is a lower bound on the probability that the (1,λ) EA with $x_{t(n)} \in A_i$ generates an $x_{t(n)+1} \in A_{i+1} \cup \cdots \cup A_m$, since – for this to happen – it is sufficient that either

- at least one offspring in $A_{i+1} \cup \cdots \cup A_m$ with a larger function value than each element of A_i and no offspring in $A_0 \cup \cdots \cup A_{i-1}$ with a function value at least as large as the one of each element in A_i are created (the probability of this event is bounded from below by $\sum_{j=1}^{\lambda} \binom{\lambda}{j} \cdot p^+(A_i)^j \cdot \big(1 - p^-(A_i) - p^+(A_i)\big)^{\lambda - j} = \big(1 - p^-(A_i)\big)^\lambda - \big(1 - p^-(A_i) - p^+(A_i)\big)^\lambda$), or
- all offspring are in $A_{i+1} \cup \cdots \cup A_m$ (the probability is bounded from below by $p(A_i)^\lambda$).

Moreover, p_i^- is an upper bound on the probability that the (1,λ) EA with $x_{t(n)} \in A_i$ generates $x_{t(n)+1} \in A_0 \cup \cdots \cup A_{i-1}$, since for this *not* to happen it is sufficient that

- at least one offspring in $A_i \cup A_{i+1} \cup \cdots \cup A_m$ with a function value at least as large as the one of each element in A_i is generated but no offspring in $A_0 \cup \cdots \cup A_{i-1}$ with a function value at least as large as the function value of an element in A_i. The probability for this is bounded from below by $\sum_{j=1}^{\lambda} \binom{\lambda}{j} \big(6/17 + p^+(A_i)\big)^j \cdot \big(1 - p^-(A_i) - (6/17 + p^+(A_i))\big)^{\lambda-j} = \big(1 - p^-(A_i)\big)^\lambda - \big(11/17 - p^-(A_i) - p^+(A_i)\big)^\lambda$ since the probability of generating a duplicate equals $(1 - 1/n)^n \geq 6/17$.

Let $p_{x,x'}$ denote the probability that the (1,λ) EA generates x' as next parent when x is mutated, and let T_x denote the expected number of steps until an element in A_m is generated (for the first time) when starting with x. Obviously, $T_x = 0$ if $x \in A_m$. For $x \in A_i$ with $i < m$,

$$T_x = 1 + \sum_{x' \in A_0 \cup \cdots \cup A_{i-1}} p_{x,x'} T_{x'} + \sum_{x' \in A_i} p_{x,x'} T_{x'} + \sum_{x' \in A_{i+1} \cup \cdots \cup A_m} p_{x,x'} T_{x'} \ .$$

Since $T_x \leq \max\{T_x \,|\, x \in A_i \cup \cdots \cup A_m\} =: T_i$ (so that $T_i \geq T_{i+1}$),

$$T_x \leq 1 + \sum_{x' \in A_0 \cup \cdots \cup A_{i-1}} p_{x,x'} T_0 + \sum_{x' \in A_i} p_{x,x'} T_i + \sum_{x' \in A_{i+1} \cup \cdots \cup A_m} p_{x,x'} T_{i+1} \ .$$

As we have seen above, $\sum_{x' \in A_{i+1} \cup \cdots \cup A_m} p_{x,x'} \geq p_i^+$ and $\sum_{x' \in A_0 \cup \cdots \cup A_{i-1}} p_{x,x'} \leq p_i^-$. Hence, for each $x \in A_i$

$$T_x \leq 1 + p_i^- T_0 + (1 - p_i^- - p_i^+) T_i + p_i^+ T_{i+1} \ .$$

Thus, $\max\{T_x \,|\, x \in A_i\} \leq 1 + p_i^- T_0 + (1 - p_i^- - p_i^+) T_i + p_i^+ T_{i+1}$ and with $T_{i+1} \leq 1 + p_i^- T_0 + (1 - p_i^- - p_i^+) T_i + p_i^+ T_{i+1}$

$$\begin{aligned} T_i &= \max\{\max\{T_x \,|\, x \in A_i\}, T_{i+1}\} \\ &\leq 1 + p_i^- T_0 + (1 - p_i^- - p_i^+) T_i + p_i^+ T_{i+1} \ . \end{aligned}$$

Since $T_x \leq T_0$ for all $x \in \{0,1\}^n$, we are interested in an upper bound on T_0.

We consider the following Markov process $\mathcal{M}$ on $m+1$ states. For $i \in \{1, \ldots, m-1\}$ let $p_{i,0} := p_i^-$, $p_{i,i+1} := p_i^+$, $p_{i,i} := 1 - p_i^- - p_i^+$, and $p_{0,0} := 1 - p_0^+$, $p_{0,1} := p_0^+$, $p_{m,0} := p_m^-$, $p_{m,m} := 1 - p_m^-$. Moreover, let $p_{i,j} := 0$ for $i \in \{1, \ldots, m-1\}$ and $j \in \{1, \ldots, i-1, i+2, \ldots, m\}$.

For the expected number of steps E_i to move in $\mathcal{M}$ from state i to state m in fact $E_i \geq T_i$. We prove

$$E_0 \leq \Big(\frac{1}{p_0^+} + \frac{1}{p_1^+ + p_1^-} + \cdots + \frac{1}{p_{i-1}^+ + p_{i-1}^-}\Big) \cdot \frac{p_1^+ + p_1^-}{p_1^+} \cdots \frac{p_{i-1}^+ + p_{i-1}^-}{p_{i-1}^+} + E_i$$

for all $i \in \{0, \ldots, m\}$ by induction over i. With the first step and the λ function evaluations in each further step, and with $i = m$, the (1,λ) EA optimizes f in an expected runtime at most $1 + \lambda \cdot E_0$ since $E_m = 0$. Obviously, $E_0 = E_0$, and for $i = 1$, it is readily seen that $E_0 = E_1 / p_0^+$. Similarly, for the estimation of E_{i+1} we utilize that

$$\begin{aligned} E_i &= 1 + p_i^- E_0 + (1 - p_i^- - p_i^+) E_i + p_i^+ E_{i+1} \\ &= \frac{1 + p_i^- E_0 + p_i^+ E_{i+1}}{p_i^+ + p_i^-} \ . \end{aligned}$$

Now, since $\frac{1}{p_i^+ + p_i^-} \leq \frac{1}{p_i^+ + p_i^-} \cdot \frac{p_1^+ + p_1^-}{p_1^+} \cdots \frac{p_{i-1}^+ + p_{i-1}^-}{p_{i-1}^+}$, using the estimate for E_i (induction) yields

$$\begin{aligned} & E_0 \cdot \Big(1 - \frac{p_i^-}{p_i^+ + p_i^-}\Big) \\ \leq\ & \Big(\frac{1}{p_0^+} + \frac{1}{p_1^+ + p_1^-} + \cdots + \frac{1}{p_{i-1}^+ + p_{i-1}^-} + \frac{1}{p_i^+ + p_i^-}\Big) \\ & \cdot \frac{p_1^+ + p_1^-}{p_1^+} \cdots \frac{p_{i-1}^+ + p_{i-1}^-}{p_{i-1}^+} + \frac{p_i^+ E_{i+1}}{p_i^+ + p_i^-} \ . \end{aligned}$$

Finally, $\big(1 - \frac{p_i^-}{p_i^+ + p_i^-}\big) \cdot \frac{p_i^+ + p_i^-}{p_i^+} = 1$, so that the claimed inequality holds also for $i + 1$. ■

We apply this lemma exemplary in the following section.

C. Application to CLIFF

A comparison of the (1+λ) EA and the (1,λ) EA for the optimization of a function f with a unique optimum is interesting especially for $\lambda = \Theta(\ln n)$: On the one hand, for $\lambda \leq \ln(n)/14$, the (1,λ) EA cannot optimize f efficiently at all. On the other hand, for $\lambda = \omega(\ln n)$, it is impossible that the (1$\star\lambda$) EA is efficient for f when the (1$\bar{\star}\lambda$) EA is totally inefficient for f.

Let us investigate the function CLIFF : $\{0,1\}^n \to \mathbb{R}$ with

$$\text{CLIFF}(x) := \begin{cases} \text{ONEMAX}(x) - \lfloor n/3 \rfloor & \text{if } |x| \geq n - \lfloor n/3 \rfloor, \\ \text{ONEMAX}(x) & \text{if } |x| < n - \lfloor n/3 \rfloor. \end{cases}$$

Its analogue in continuous search spaces has been studied in [12]. Typically, the (1+λ) EA waits for a long time at the cliff, which consists of all elements x with $|x| < n - \lfloor n/3 \rfloor$, whereas the (1,$\lambda$) EA approaches the border of the cliff and, after a short while, jumps over the cliff and hardly ever drops back. The following theorem proves that the (1,λ) EA is efficient with an offspring population size that is logarithmic in n, whereas the (1+λ) EA is totally inefficient for *any* offspring population sizes. (The opposite effect could also be illustrated.)

Theorem 8

1) *With probability* $1 - 2^{-\Omega(n)}$ *the* (1+λ) EA *needs a runtime larger than* $n^{n/4}$ *to optimize* CLIFF.
2) *If* $\lambda \leq \varepsilon(n) \ln(n)/7$ *for* $\varepsilon(n) \in [7/\ln n, 1/2]$, *with probability* $1 - 2^{-\Omega(n^{1-\varepsilon(n)})}$ *the* (1,λ) EA *needs a runtime larger than* $2^{n^{1-\varepsilon(n)}}$ *to optimize* CLIFF.
3) *If* $\lambda \geq 5 \ln n$, *the expected runtime of the* (1,λ) EA *on* CLIFF *is* $\mathcal{O}(e^{5\lambda})$.
4) *The expected runtime of the* (1,λ) EA *on* CLIFF *is larger than* $\min\{n^{n/4}, e^{\lambda/4}\}/3$.

Proof:

1) The (1+λ) EA with $x_{t(n)}$, where $|x_{t(n)}| < n - \lfloor n/3 \rfloor$, generates $x_{t(n)+1}$, where $|x_{t(n)+1}| < n - \lfloor n/3 \rfloor$ (case 1) or $|x_{t(n)+1}| \geq |x_{t(n)}| + \lfloor n/3 \rfloor$ (case 2) only. For case 2 to occur, at least any $\lfloor n/3 \rfloor$ bits have to flip. The probability that this happens at least once in $\lfloor n^{n/4} \rfloor$ mutations is bounded from above by $\lfloor n^{n/4} \rfloor \cdot 1/\lfloor n/3 \rfloor! \leq n^{n/4} \cdot (e/\lfloor n/3 \rfloor)^{\lfloor n/3 \rfloor} = 2^{-\Omega(n \ln n)}$. As long as only case 1 occurs, the optimum is not generated. The probability that for x_1 holds $|x_1| \geq n - \lfloor n/3 \rfloor$ is bounded from above by $\sum_{i=n-\lfloor n/3 \rfloor}^{n} \binom{n}{i}/2^n \leq n \cdot \binom{n}{\lfloor n/3 \rfloor}/2^n$. Furthermore, $n \cdot \binom{n}{\lfloor n/3 \rfloor} \leq n \cdot \frac{n \cdots (n-\lceil n/6 \rceil+1) \cdot (n-\lceil n/6 \rceil) \cdots (n-\lfloor n/3 \rfloor+1)}{(1/n) \cdot \lfloor n/3+1 \rfloor!} \leq n^2 \cdot \frac{n^{\lceil n/6 \rceil} \cdot (5n/6)^{\lfloor n/3 \rfloor - \lceil n/6 \rceil}}{2\sqrt{n/3} \cdot (n/(3e))^{\lfloor n/3 \rfloor}}$, where Stirling's formula is applied. Moreover, the former expression is bounded from above by $n^2 \cdot 2^{\lfloor n/3 \rfloor \cdot (\log_2(5/6)/2 + \log_2(3e))} \leq 2^{28n/29}$. Hence, $\sum_{i=n-\lfloor n/3 \rfloor}^{n} \binom{n}{i}/2^n \leq 2^{28n/29}/2^n = 2^{-\Omega(n)}$ and with probability $1 - 2^{-\Omega(n \ln n)} - 2^{-\Omega(n)}$ the (1+λ) EA has not optimized CLIFF in runtime $n^{n/4}$.
2) The result follows by Theorem 1 since CLIFF has a unique optimum, namely 1^n.
3) For $\lambda > n \ln n$ the result follows since the optimum is generated with probability at least $1/n^n$ in each mutation as at most n bits have to flip. So, let $\lambda \leq n \ln n$ and $\ell := \lfloor \ln \lambda / \ln \ln \lambda \rfloor$. In order to apply Lemma 7, we distinguish three classes of partitions A_i and determine p_i^+ and p_i^- in each case. Note that $(11/17)^\lambda \leq (11/17)^{5 \ln n} \leq 1/n^2$.
   a) $A_i := \{x \mid |x| = i\}$, $0 \leq i \leq n - \lfloor n/3 \rfloor - 1$. Since $p^-(A_i) = 0$ it holds
   $$p_i^- \leq 1/n^2 \text{ for } 0 \leq i \leq n - \lfloor n/3 \rfloor - 1 \ .$$
   Moreover, $(1-p^-(A_i))^\lambda - (1-p^-(A_i)-p^+(A_i))^\lambda \geq 1-(1-p^+(A_i)) \geq \lfloor n/3 \rfloor \cdot (1/n)(1-1/n)^{n-1} \geq 1/9$, and hence, (movement towards the cliff)
   $$p_i^+ \geq 1/9 \text{ for } 0 \leq i < n - \lfloor n/3 \rfloor - 1$$
   and (jump over the cliff)
   $$p_{n-\lfloor n/3 \rfloor - 1}^+ \geq 9^{-\lambda}$$
   since we obtain (similarly to above) $p(A_{n-\lfloor n/3 \rfloor - 1})^\lambda \geq (\lfloor n/3 \rfloor \cdot (1/n)(1-1/n)^{n-1})^\lambda \geq (1/9)^\lambda$.
   b) $A_{n-\lfloor n/3 \rfloor + i} := \{x \mid n - \lfloor n/3 \rfloor + i\ell \leq |x| < n - \lfloor n/3 \rfloor + (i+1)\ell\}$, $0 \leq i < \lfloor n/(12\ell) \rfloor$. It holds
   $$p^+(A_{n-\lfloor n/3 \rfloor + i}) \geq \frac{\binom{\lfloor n/3 \rfloor - (i+1)\ell}{\ell}}{en^\ell}$$
   since it is sufficient to flip exactly ℓ out of $n - |x| \geq \lfloor n/3 \rfloor - (i+1)\ell \geq n/5$ specific bits of $x \in A_{n-\lfloor n/3 \rfloor + i}$. Hence,
   $$\begin{aligned} \frac{\binom{\lfloor n/3 \rfloor - (i+1)\ell}{\ell}}{en^\ell} &\geq \frac{\left(\frac{n/5}{\ell}\right)^\ell}{en^\ell} = e^{-\ell \ln(\ell \cdot 5) - 1} \\ &\geq e^{-\frac{\ln \lambda}{\ln \ln \lambda} \cdot (\ln \ln \lambda - \ln \ln \ln \lambda + \ln 5) - 1} \geq \frac{1}{\lambda} \end{aligned}$$
   since $(\ln \lambda / \ln \ln \lambda) \cdot (\ln \ln \lambda - \ln 5) - 1 \geq 0$. Furthermore, we obtain (probable return to the cliff)
   $$p_{n-\lfloor n/3 \rfloor + i}^- \leq 1 - p_{n-\lfloor n/3 \rfloor + i}^+ \text{ for } 0 \leq i \leq 1$$
   and (improbable movement off the cliff)
   $$\begin{aligned} & p_{n-\lfloor n/3 \rfloor + i}^+ \\ \geq\ & \left(1 - p^-(A_{n-\lfloor n/3 \rfloor + 1})\right)^\lambda \\ & -\left(1 - p^-(A_{n-\lfloor n/3 \rfloor + 1}) - p^+(A_{n-\lfloor n/3 \rfloor + 1})\right)^\lambda \\ =\ & \sum_{j=1}^{\lambda} \binom{\lambda}{j} \cdot p^+(A_{n-\lfloor n/3 \rfloor + i})^j \\ & \cdot \left(1 - p^-(A_{n-\lfloor n/3 \rfloor + i}) - p^+(A_{n-\lfloor n/3 \rfloor + i})\right)^{\lambda - j} \\ \geq\ & p^+(A_{n-\lfloor n/3 \rfloor + i}) \\ & \cdot \left(1 - p^-(A_{n-\lfloor n/3 \rfloor + i}) - p^+(A_{n-\lfloor n/3 \rfloor + i})\right)^{\lambda - 1} \\ \geq\ & (6/17)^\lambda / \lambda \text{ for } 0 \leq i \leq 1 \end{aligned}$$
   since $p^+(A_{n-\lfloor n/3 \rfloor + i}) \geq 1/\lambda$ and $p^-(A_{n-\lfloor n/3 \rfloor + i}) + p^+(A_{n-\lfloor n/3 \rfloor + i}) \leq 11/17$. Hence, we obtain
   $$\begin{aligned} & p^-(A_{n-\lfloor n/3 \rfloor + i}) \leq 1/(\ell i)! \leq (e/(\ell i))^{\ell i} \\ \leq\ & e^{-(1 + \frac{3 \ln \ln \lambda}{4}) \cdot (\frac{\ln \lambda}{\ln \ln \lambda} - 1) \cdot i} \leq 1/\lambda^{3i/4} \ , \end{aligned}$$
   for $2 \leq i < \lfloor n/(12\ell) \rfloor$ since at least any ℓi bits have to flip, for the last but one inequality holds $e/(\ell i) \leq 1/(e \ln^{3/4} \lambda)$, and for the last inequality
   $$-\frac{\ln \lambda}{\ln \ln \lambda} + 1 - \frac{3 \ln \lambda}{4} + \frac{3 \ln \ln \lambda}{4} \leq \frac{-3 \ln \lambda}{4} \ .$$
   Thus,
   $$\left(1 - p^-(A_{n-\lfloor n/3 \rfloor + i})\right)^\lambda \geq \left(1 - 1/\lambda^{3i/4}\right)^\lambda \geq 1 - \lambda^{1-3i/4}$$
   and we obtain (improbable return to the cliff)
   $$p_{n-\lfloor n/3 \rfloor + i}^- \leq \lambda^{1-3i/4} + 1/n^2 \text{ for } 2 \leq i < \lfloor n/(12\ell) \rfloor$$
   as well as (probable movement off the cliff)
   $$p_{n-\lfloor n/3 \rfloor + i}^+ \geq \left(1 - \lambda^{1-3i/4}\right) - \left(1 - 1/\lambda\right)^\lambda \geq 1/2 \quad \text{for } 2 \leq i < \lfloor n/(12\ell) \rfloor \ .$$
   c) $A_{n-\lfloor n/3 \rfloor + \lfloor n/(12\ell) \rfloor + i} := \{x \mid x = n - \lfloor n/3 \rfloor + \lfloor n/(12\ell) \rfloor \ell + i\}$, $0 \leq i \leq \lfloor n/3 \rfloor - \lfloor n/(12\ell) \rfloor \ell$. It holds
   $$\begin{aligned} & \left(1 - p^-(A_{n-\lfloor n/3 \rfloor + \lfloor n/(12\ell) \rfloor + i})\right)^\lambda \geq 1/2 \\ \geq\ & \left(1 - \frac{1}{\lceil n/13 \rceil!}\right)^\lambda \geq 1 - \lambda \cdot \left(\frac{e}{\lceil n/13 \rceil}\right)^{\lceil n/13 \rceil} \\ \geq\ & 1 - e^{-n} \end{aligned}$$
   since at least any $\lfloor n/(12\ell) \rfloor \ell \geq \lceil n/13 \rceil$ bits have to flip. We obtain (improbable return to the cliff)
   $$\begin{aligned} & p_{n-\lfloor n/3 \rfloor + \lfloor n/(12\ell) \rfloor + i}^- \\ \leq\ & e^{-n} + 1/n^2 \text{ for } 0 \leq i \leq \lfloor n/3 \rfloor - \lfloor n/(12\ell) \rfloor \ell \end{aligned}$$
   and furthermore, since $p^+(A_{n-\lfloor n/3 \rfloor + \lfloor n/(12\ell) \rfloor + i}) \geq (1/n)(1-1/n)^{n-1} \geq 1/(en)$ for $i < \lfloor n/3 \rfloor - \lfloor n/(12\ell) \rfloor \ell$, we have (probable movement off the cliff)
   $$\begin{aligned} & p_{n-\lfloor n/3 \rfloor + \lfloor n/(12\ell) \rfloor + i}^+ \\ \geq\ & \left(1 - p^-(A_{n-\lfloor n/3 \rfloor + \lfloor n/(12\ell) \rfloor + i})\right)^\lambda \\ & -\left(1 - p^+(A_{n-\lfloor n/3 \rfloor + \lfloor n/(12\ell) \rfloor + i})\right)^\lambda \end{aligned}$$

$$\begin{aligned} &\geq (1-e^{-n}) - \left(1-(1/n)(1-1/n)^{n-1}\right)^{\lambda} \\ &\geq 1/(3n) \text{ for } 0 \leq i < \lfloor n/3 \rfloor - \lfloor n/(12\ell) \rfloor \ell \ . \end{aligned}$$

As argued, Lemma 7 implies for the (1,λ) EA an upper bound on the expected runtime to optimize CLIFF of

$$\Big(1+\lambda\cdot\big((\sum_{i=0}^{n-\lfloor n/3\rfloor-2} 9) + n^2 + 1 + 1 + (\sum_{i=2}^{\lfloor n/(12\ell)\rfloor-1} 2) + (\sum_{i=0}^{\lfloor n/3\rfloor-\lfloor n/(12\ell)\rfloor\ell} 3n)\big)\Big)$$

$$\text{(a)} \quad \cdot \prod_{i=0}^{n-\lfloor n/3\rfloor-2} \left(\frac{\frac{1}{9}+\frac{1}{n^2}}{\frac{1}{9}}\right) \cdot \frac{\frac{1}{9^{\lambda}}+\frac{1}{n^2}}{\frac{1}{9^{\lambda}}}$$

$$\text{(b)} \quad \cdot \frac{1}{\frac{(6/17)^{\lambda}}{\lambda}} \cdot \frac{1}{\frac{(6/17)^{\lambda}}{\lambda}} \cdot \prod_{i=2}^{\lfloor n/(12\ell)\rfloor-1} \left(\frac{\frac{1}{2}+(\lambda^{1-3i/4}+\frac{1}{n^2})}{\frac{1}{2}}\right)$$

$$\text{(c)} \quad \cdot \prod_{i=0}^{\lfloor n/3\rfloor-\lfloor n/(12\ell)\rfloor\ell} \left(\frac{\frac{1}{3n}+(\frac{1}{e^n}+\frac{1}{n^2})}{\frac{1}{3n}}\right) \ .$$

The letters on the left identify the class of partition investigated. The expression in the first line is bounded by $\mathcal{O}(n^2\lambda)$, and (a) is bounded by $\mathcal{O}(1)\cdot\mathcal{O}(9^{\lambda}/n^2) = \mathcal{O}(e^{9\lambda/4}/n^2)$ since $\ln 9 \leq 9/4$. The expression (b) is bounded by $(17/6)^{\lambda}\lambda \cdot (17/6)^{\lambda}\lambda \cdot \mathcal{O}(e^{\lambda/4}/\lambda^3) = \mathcal{O}(e^{11\lambda/4}/\lambda)$ since $\ln(17/6) \leq 5/4$ and

$$\begin{aligned} &\prod_{i=2}^{\lfloor n/(12\ell)\rfloor-1} 1+2\lambda^{1-3i/4}+2/n^2 \\ \leq\ &\prod_{i=2}^{\lfloor \ln n\rfloor} 2 \cdot \prod_{i=\lfloor \ln n\rfloor+1}^{\lfloor n/(12\ell)\rfloor-1} (1+3/n^2) \\ \leq\ &n\cdot 2 = \mathcal{O}(e^{\lambda/4}/\lambda^3) \ . \end{aligned}$$

Finally, the expression (c) is bounded by $\mathcal{O}(1)$. Therefore, the (1,λ) EA optimizes CLIFF in an expected runtime $\mathcal{O}(n^2\lambda)\cdot\mathcal{O}(e^{9\lambda/4}/n^2)\cdot\mathcal{O}(e^{11\lambda/4}/\lambda)\cdot\mathcal{O}(1) = \mathcal{O}(e^{5\lambda})$.

4) By part 1), with probability $1-2^{-\Omega(n)}$ the (1+λ) EA needs a runtime larger than $\min\{n^{n/4}, e^{\lambda/4}\}$ to optimize CLIFF. If $\min\{n^{n/4}, e^{\lambda/4}\} \leq 1+\lambda$, by Theorem 4.1, the (1,λ) EA needs a runtime larger than $\min\{n^{n/4}, e^{\lambda/4}\}$ to optimize CLIFF with probability $1-2^{-\Omega(n)} \geq 1/3$, too.
If $\min\{n^{n/4}, e^{\lambda/4}\} > 1+\lambda$, by Theorem 4.2, where $c(n) = 3/5$, the (1,λ) EA needs a runtime larger than $\min\{n^{n/4}, e^{\lambda/4}\}$ to optimize CLIFF with probability at least $(1-2^{-\Omega(n)}) - 1/2^{3/5} \geq 1/3$ since $2 \leq t(n) \leq e^{\lambda/4}$ and $\lambda \geq (5/2)\cdot(1+3/5)\ln t(n)$.
Consequently, the (1,λ) EA needs an expected runtime larger than $\min\{n^{n/4}, e^{\lambda/4}\}/3$ to optimize CLIFF.

■

It is worth to note that the proof of Theorem 8.3 implies the following: For $\lambda \geq 5\ln n$, the (1,λ) EA creates the optimum of CLIFF in runtime $\mathcal{O}(\lambda n^2)$ with probability $\Omega(e^{-5\lambda})$, at least.

Let us take a short look at the multistart variant/extension of EAs, where a particular EA A is (independently) restarted after runtime $\ell(n)$, denoted by $A_{\ell(n)}$. We observe the following for $\lambda \geq 3\ln \ell(n)$ and polynomially bounded values of $\ell(n)$. If the (1+λ) EA$_{\ell(n)}$ is efficient, then also the (1,λ) EA$_{\ell(n)}$ is efficient. And as we have seen, there exist functions where the (1,λ) EA$_{\ell(n)}$ is efficient, but the (1+λ) EA$_{\ell(n)}$ is not. So, in this situation and in case of doubt, one should prefer the (1,λ) EA$_{\ell(n)}$. In cases when $\lambda \leq \ln(n)/14$, however, one should definitely prefer the (1+λ) EA$_{\ell(n)}$. Thus, we have obtained a somewhat general rule when to apply the comma or the plus selection.

V. Summary and Conclusions

We have compared the (1,λ) EA and the (1+λ) EA operating on $\{0,1\}^n$, and it has been pointed out why only the consideration of offspring populations of logarithmic size in n are interesting. For smaller values of λ, the (1,λ) is totally inefficient on every function with a unique optimum, whereas for larger values, the (1,λ) EA and (1+λ) EA behave equivalently with a high probability. These investigations have been exemplified by ONEMAX. For the example function CLIFF, we have analyzed rigorously when and why the (1,λ) EA outperforms the (1+λ) EA. Therefore, a simple but powerful proof method has been developed. However, our results support – depending on the offspring population size – the importance of a correct choice of the selection operator.

Acknowledgments: This work was supported by the German Research Foundation (DFG) through the collaborative research center "Computational Intelligence" (SFB 531) and grant We 1066/11.

References

[1] H.-P. Schwefel, *Evolution and Optimum Seeking*. Wiley, 1995.
[2] S. Droste, T. Jansen, and I. Wegener, "On the analysis of the (1+1) evolutionary algorithm," *Theoretical Computer Science*, vol. 276, pp. 51–81, 2002.
[3] J. Garnier, L. Kallel, and M. Schoenauer, "Rigorous hitting times for binary mutations," *Evolutionary Computation*, vol. 7, pp. 173–203, 1999.
[4] O. Giel and I. Wegener, "Searching randomly for maximum matchings," *Electronic Colloquium on Computational Complexity*, vol. 76, 2004.
[5] T. Jansen, K. De Jong, and I. Wegener, "On the choice of the offspring population size in evolutionary algorithms," *Evolutionary Computation*, vol. 13, pp. 413–440, 2006.
[6] C. Witt, "Runtime analysis of the (μ+1) EA on simple pseudo-boolean functions," *Evolutionary Computation*, vol. 14, pp. 65–86, 2006.
[7] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge University Press, 1995.
[8] J. He and X. Yao, "Drift analysis and average time complexity of evolutionary algorithms," *Artificial Intelligence*, vol. 127, pp. 57–85, 2001.
[9] J. Jägersküpper, "Analysis of a simple evolutionary algorithm for minimization in Euclidean spaces," in *Proc. 30th International Colloquium on Automata, Languages, and Programming – ICALP 2003*, 2003, pp. 1068–1079.
[10] I. Wegener, "Methods for the analysis of evolutionary algorithms on pseudo-boolean functions," in *Evolutionary Optimization*, 2002, pp. 349–369.
[11] T. Storch, "On the choice of the population size," in *Proc. Genetic and Evolutionary Computation Conference – GECCO 2004, LNCS 3102*, 2004, pp. 748–760.
[12] J. Jägersküpper and T. Storch, "How comma selection helps with the escape from local optima," in *Proc. International Conference on Parallel Problem Solving From Nature IX – PPSN 2006*, 2006, pp. 52–61.