Runtime Analysis of the (1+1) Evolutionary Algorithm on Strings over Finite Alphabets*

Benjamin Doerr(1), Daniel Johannsen(1), and Martin Schmidt(2)

(1) Max Planck Institute for Informatics
Saarbrücken, Germany

(2) Saarland University
Saarbrücken, Germany

Abstract

In this work, we investigate a (1+1) Evolutionary Algorithm for optimizing functions over the space $\{0, \dots, r\}^n$, where r is a positive integer. We show that for linear functions over $\{0, 1, 2\}^n$, the expected runtime time of this algorithm is $O(n \log n)$. This result generalizes an existing result on pseudo-Boolean functions and is derived using drift analysis. We also show that for large values of r, no upper bound for the runtime of the (1+1) Evolutionary Algorithm for linear function on $\{0, \dots, r\}^n$ can be obtained with this approach nor with any other approach based on drift analysis with weight-independent linear potential functions.

1 Introduction

In their seminal work [8], Droste, Jansen, and Wegener analyzed the runtime of the (1+1) Evolutionary Algorithm ((1+1) EA) for several classes of pseudo-Boolean functions, including the class of linear pseudo-Boolean functions. Their results motivated a number of further studies of randomized search heuristics on basic pseudo-Boolean functions (e.g., [5, 11, 14]) and later on combinatorial optimization problems (e.g., [15, 17, 19]); see [16, 1] for an overview.

In the last decade, particular interest was given to the study of the (1+1) EA optimizing linear pseudo-Boolean functions of the form

$$f\colon \{0,1\}^n \to \mathbb{R}, \quad x \mapsto w_0 + \sum_{i=1}^{n} w_i x_i, \quad w_0, w_1, \dots, w_n \in \mathbb{R}.$$

For this class of functions, Droste, Jansen, and Wegener showed that the (expected) runtime of the (1+1) EA is in $\Theta(n \log n)$ for non-zero weights, and Doerr, Johannsen, and Winzen [5] showed that the leading constants in this bound are between 1 and 1.39.

*This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in GECCO-2010: Proceedings of Genetic and Evolutionary Computation Conference, 2010, ACM, http://doi.acm.org/10.1145/1873601.1873717.

One reason for our interest in the class of linear functions is the analytic challenge of its runtime analysis. For example, all known proofs that obtain asymptotically sharp runtime bounds of $O(n \log n)$ do not study the progress of the (1+1) EA with respect to the actual function value, but instead with respect to a structure-related potential function.

For example, in [8] and [6] the progress of the (1+1) EA minimizing linear functions is measured with respect to the potential function

$$g_0\colon \{0,1\}^n \to \mathbb{R}, \quad x \mapsto \sum_{i=1}^{\lfloor n/2 \rfloor} 2x_i + \sum_{i=\lfloor n/2 \rfloor+1}^{n} x_i,$$

in [11, 12] with respect to the potential function

$$g_1\colon \{0,1\}^n \to \mathbb{R}, \quad x \mapsto \log\Big(1 + \sum_{i=1}^{n} x_i\Big),$$

and in [14] and [5] with respect to the potential function

$$g_2\colon \{0,1\}^n \to \mathbb{R}, \quad x \mapsto |x|_1 := \sum_{i=1}^{n} x_i.$$

The study of the progress of an algorithm with respect to a potential function is called *drift analysis* (the *drift* is the expected change in potential) and is briefly introduced in Section 3, see He and Yao [11, 13], Doerr, Johannsen, Winzen [5, 6] and Hajek [10] for details on this technique.

An important progress in the understanding of the optimization behavior of the (1+1) EA on linear functions was made by Jägersküpper [14]. By studying the underlying Markov chain, he showed a basic but important invariance property on its distribution: For any search point $x \in \{0,1\}^n$ generated by the (1+1) EA minimizing a linear function, we have that

$$\Pr[x_i = 0 \wedge x_j = 1] \le \Pr[x_i = 1 \wedge x_j = 0]$$

whenever $w_i \le w_j$.

In fact, it is this invariance property that enabled him to use the simple (and intuitive) potential function g_2 to analyze the optimization behavior of the (1+1) EA on linear pseudo-Boolean functions.

In contrast to linear pseudo-Boolean functions, the optimization behavior of the (1+1) EA on combinatorial problems like the minimum spanning tree problems [15, 17, 18] or the shortest path tree problem [2, 3, 4, 19] are still poorly understood with respect to structural insights.

For example, consider the minimum spanning tree problem. As an optimal subgraph problem, search points can be represented by bitstrings indicating which edges appear in the respective subgraph. For this problem, we have a linear fitness function given by the sum of the edge weights in the subgraph (which we assume to be positive integers). However, in contrast to linear pseudo-Boolean functions, not all search points are feasible. More precisely, the (1+1) EA starts with an arbitrary spanning subgraph of the input graph and rejects all search points that represent disconnect subgraphs.

For the minimum spanning tree problem [15, 18], the best known upper bounds on the runtime are a general bound of $O(m^3 \log m)$ and a bound of $O(m^2(\log m + \log w_{\max}))$ depending on the maximum edge weight $w_{\max}$. Note that the latter bound is worse than the former if $w_{\max}$ is much larger than m^m. However, [15] also gives the lower bound $O(m^2 \log m)$. We conjecture that this is the right order of magnitude, i.e., that the logarithmic factor in the maximal weight is unnecessary.

This conjecture is motivated by the observation that the equally difficult problem of finding a maximum spanning forest generalizes the problem of maximizing a linear pseudo-Boolean function (represented by the problem of finding a maximum spanning forest on a path). Now, in the (equivalent) analysis of the minimum spanning tree problem, the progress of the (1+1) EA is simply measured by the current fitness distance to the optimum. If we apply the same analysis to linear pseudo-Boolean functions, we only obtain a weight-dependent runtime bound of $O(m(\log m + \log w_{\max}))$.

With this in mind, one might hope that the technique of measuring the progress of the (1+1) EA for the minimum spanning tree problem by a weight-independent potential function like in the case of linear pseudo-Boolean functions will remove the logarithmic factor in the maximal weight from the upper bound of the runtime.

In this work, we approach this question by analyzing the optimization behavior of the (1+1) EA for linear functions over the space $\{0, \ldots, r\}^n$, i.e., function of the form

$$f\colon \{0, \ldots, r\}^n \to \mathbb{R}, \quad x \mapsto w_0 + \sum_{i=1}^{n} w_i x_i$$

where n and r are positive integers and $w_0, \ldots, w_n$ are real-valued weights. One might expect that the known methods by Droste, Jansen, and Wegener [8], He and Yao [13], and Jägersküpper [14] generalize to this setting. Indeed, in Section 3 we introduce a potential function that allows us to bound the runtime of the (1+1) EA for the case $r = 2$, i.e., for the search space $\{0, 1, 2\}^n$.

Theorem 1. *The expected runtime of the (1+1) Evolutionary Algorithm (Algorithm 1) minimizing a linear function over* $\{0, 1, 2\}^n$ *is in* $O(n \log n)$.

Surprisingly, it turns out that the proof of the invariance property in the approach of Jägersküpper does not immediately generalize to this setting. In fact, in line with the results in [5], we show in Section 4 that this approach and also more general approaches using arbitrary weight-independent linear potential functions fail for values of r larger than 43 (the actual value of this constant is of minor importance and we expect it to be much smaller than 43).

Similar restrictions on the current techniques of analysis have been encounter in [5, 6] for the study of the (1+1) EA with mutation rates larger than $1/n$.

2 A (1+1) Evolutionary Algorithm for Strings Over Finite Alphabets

In this section, we present the (1+1) Evolutionary Algorithm ((1+1) EA) for minimizing functions over the search space $\{0, 1, \ldots, r\}^n$ with positive integers n and r. This algorithm is a generalization of the classical (1+1) EA for pseudo-Boolean functions (see, e.g., [8]).

In the classical (1+1) EA, the mutation operator flips each bit of the current search point with probability $1/n$. Accordingly, we define a mutation operator on $\{0, 1, \ldots, r\}^n$ that still "flips" each position with probability $1/n$, where "flipping" means that the respective entry is changed to a different value chosen uniformly at random.

Definition 2 ($\mathbf{mut}(x)$)**.** *Let* n *and* r *be positive integers. Let* $x \in \{0, 1, \ldots, r\}^n$*. The mutation operator* $\mathrm{mut}(x)$ *generates a random search point* y *in* $\{0, 1, \ldots, r\}^n$*. Independently for each* $i \in \{1, \ldots, n\}$*, the probability that* $y_i \neq x_i$ *is* $1/n$*. In this case,* y_i *is chosen uniformly at random from* $\{0, 1, \ldots, r\} \setminus \{x_i\}$*.*

Using this mutation operator, we generalize the search space of the classical (1+1) EA optimizing pseudo-Boolean functions to $\{0, 1, \ldots, r\}^n$. This (1+1) Evolutionary Algorithm on $\{0, 1, \ldots, r\}^n$ is given by Algorithm 1. It was studied before by Gunia [9] to analyze the minimum make-span scheduling problem with $r + 1$ machines.

Algorithm 1: The (1+1) EA on $\{0, 1, \ldots, r\}^n$.

input function $f\colon \{0, 1, \ldots, r\}^n \to \mathbb{R}$;
initialization Choose $x \in \{0, 1, \ldots, r\}^n$ uniformly at random;
repeat
  $y := \mathrm{mut}(x)$;
  if $f(y) \leq f(x)$ **then** $x := y$;
until *forever*;

Note that, as common in theoretical runtime analysis, we do not specify a stopping criterion (see [8] for a discussion of this). Our focus is determining the *runtime* of the algorithm, which we define to be the random variable T that denotes the number of fitness evaluations (Step 4) until a global optimum (i.e., minimum) of f is found.

In the remainder of this work we focus on linear functions over $\{0,1,\dots,r\}^n$. A *linear* function f over $\{0,1,\dots,r\}^n$ is a function

$$f\colon \{0,1,\dots,r\}^n \to \mathbb{R}, \quad x \mapsto w_0 + \sum_{i=1}^{n} w_i x_i$$

with $w_i \in \mathbb{R}$ for $0 \le i \le n$. Note that for $r = 1$ this definition specializes to linear pseudo-Boolean functions. Like for linear pseudo-Boolean functions, we may assume that $w_0 = 0$ and all other weights are positive and non-increasing, i.e.,

$$0 = w_0 < w_1 \le w_2 \le \dots \le w_n$$

This is not a real restriction. First, the size of the weight w_0 does not influence the behavior of the (1+1) EA. Second, since Algorithm 1 ensures that every position flips independently, reordering the weights is not a restriction and in case a weight w_i is negative we may replace w_i and x_i by $-w_i$ and $1-x_i$ respectively, without changing the optimization behavior.

3 The Runtime of the (1+1) Evolutionary Algorithm Optimizing Linear Functions Over $\{0,1,2\}^n$

The aim of this section is to prove Theorem 1, i.e., that the expected runtime of the (1+1) EA for linear functions on $\{0,1,2\}^n$ is $O(n \log n)$. For this, we apply drift analysis, more precisely, the multiplicative drift theorem from [7].

Theorem 3. *Let $\{x^{[t]}\}_{t\in\mathbb{N}}$ be a sequence of random variables in a finite space $\mathcal{S}$. Furthermore, let $g\colon \mathcal{S} \to \mathbb{R}$ be a potential function on $\mathcal{S}$ such that $\min_{x\in\mathcal{S}\colon g(x)>0}\{g(x)\} \ge 1$. Let T be the random variable that denotes the first point in time $t \in \mathbb{N}$ such that $g(x^{[t]}) = 0$.*

Suppose that

$$\mathrm{E}\left[g(x^{[t]}) - g(x^{[t+1]}) \,\middle|\, g(x^{[t]})\right] \ge \delta g(x^{[t]}). \tag{1}$$

holds for all $t \in \mathbb{N}$. Then,

$$\mathrm{E}\left[T \,\middle|\, g(x^{[0]})\right] \le \frac{1 + \ln g(x^{[0]})}{\delta}.$$

We apply this theorem to the search points $\{x^{[t]}\}_{t\in\mathbb{N}}$ generated by the (1+1) EA (Algorithm 1) minimizing a linear function

$$f\colon \{0,1,2\}^n \to \mathbb{R}, \quad x \mapsto \sum_{i=1}^{n} w_i x_i$$

with

$$0 < w_1 \le \dots \le w_n.$$

Let $x^{[0]}$ be the random search point represented by x after Step 2 in a run of the (1+1) EA. For $t \ge 1$, let $x^{[t]}$ be the search points represented by x after the t-th iteration of Step 5 in a run of the (1+1) EA.

In order to apply Theorem 3 and prove Theorem 1, we define the potential function

$$g\colon \{0,1,2\}^n \to \mathbb{R}, \quad x \mapsto \sum_{i=1}^{n} g_i x_i := \sum_{i=1}^{n} \left(1+\tfrac{1}{n}\right)^{i-1} x_i.$$

For this potential function, we have for $g_{n+1} := (1+1/n)^n$ that

$$1 = g_1 \le \dots \le g_n \le g_{n+1} \le \mathrm{e}. \tag{2}$$

Furthermore, by the formula for geometric sums, it holds that

$$\sum_{i=1}^{k} g_i = \sum_{i=0}^{k-1} \left(1+\tfrac{1}{n}\right)^i = \frac{\left(1+\frac{1}{n}\right)^k - 1}{\left(1+\frac{1}{n}\right) - 1} = n(g_{k+1} - 1) \tag{3}$$

for all $k \in \{1,\dots,n\}$ and therefore, for all $x \in \{0,1,2\}^n$,

$$g(x) \le \sum_{i=1}^{n} 2g_i = 2n(g_{n+1} - 1) = O(n).$$

Let $x \in \{0,1,2\}^n$ be a fixed point and let $y \in \{0,1,2\}^n$ be the random search point generated by $\mathrm{mut}(x)$. Furthermore, let $z \in \{-2,-1,0,1,2\}^n$ be the vector given by

$$z := x - y,$$

and let its positive and negative parts $z^+, z^- \in \{0,1,2\}^n$ be given by

$$z_i^+ := \max\{x_i - y_i, 0\} \quad \text{and} \quad z_i^- := \max\{y_i - x_i, 0\}$$

for all $i \in \{1,\dots,n\}$, such that

$$z = z^+ - z^-.$$

Let $t \in \mathbb{N}$. Suppose that $x^{[t]} = x$ and y is the search point generated by mutation of x in Step 4 of the (1+1) EA. By the selection performed in Step 5, we have $x^{[t+1]} = y$ if $f(z) \ge 0$ and $x^{[t+1]} = x$ otherwise. Thus,

$$g(x^{[t]}) - g(x^{[t+1]}) = g(z)\chi(f(z) \ge 0)$$

where g and f are extended to $\{-2,-1,0,1,2\}^n$ in the obvious way and $\chi(\mathcal{A}) \in \{0,1\}$ is the random variable that denotes the indicator function of the event $\mathcal{A}$ such that

$$\Pr[\chi(\mathcal{A}) = 1] = \Pr[\mathcal{A}].$$

Therefore,

$$\mathrm{E}\left[g(x^{[t]}) - g(x^{[t+1]}) \,\middle|\, x^{[t]} = x\right] = \mathrm{E}\left[g(z)\chi(f(z) \ge 0)\right].$$

If we show the following statement then Theorem 1 directly follows from the previous equation and from Theorem 3.

Lemma 4. *Independently of the choice of $x \in \{0,1,2\}^n$, we have that*

$$\mathrm{E}\left[g(z)\chi(f(z) \ge 0)\right] = \Omega\big(g(x)/n\big).$$

The remainder of this section is devoted to the proof of this lemma (which immediately concludes the proof of Theorem 1). We distinguish four main cases, namely whether the value of $|z^+| := \sum_{i+1}^n z_i^+$ is zero, one, two, or at least three.

$$\begin{aligned} \mathrm{E}\left[g(z)\chi(f(z) \geq 0)\right] &= \mathrm{E}\left[g(z)\chi(|z^+| = 0 \wedge f(z) \geq 0)\right] \\ &+ \mathrm{E}\left[g(z)\chi(|z^+| = 1 \wedge f(z) \geq 0)\right] \quad (4) \\ &+ \mathrm{E}\left[g(z)\chi(|z^+| = 2 \wedge f(z) \geq 0)\right] \\ &+ \mathrm{E}\left[g(z)\chi(|z^+| \geq 3 \wedge f(z) \geq 0)\right]. \end{aligned}$$

The first term in the sum of (4) is easily seen to be zero.

Proposition 5. *It holds that*

$$\mathrm{E}\left[g(z)\chi(|z^+| = 0 \wedge f(z) \geq 0)\right] = 0.$$

Proof. Together, the conditions $|z^+| = 0$ and $f(z) \geq 0$ imply $x = y$ and therefore $g(z) = 0$. □

Since $\mathrm{E}[g(z^+)]$ is reasonably bounded, we see that the fourth (and last) term in the sum of (4) is at least zero.

Proposition 6. *It holds that*

$$\mathrm{E}\left[g(z)\chi(|z^+| \geq 3 \wedge f(z) \geq 0)\right] \geq 0.$$

Proof. Let $\mathcal{A}$ be the event that we have $|z^+| \geq 3$ and $f(z) \geq 0$. We consider the (conditional) expected values of $g(z^+)$ and $g(z^-)$ independently.

On the one hand,

$$\mathrm{E}\left[g(z^+)\chi(\mathcal{A})\right] = \mathrm{E}\left[g(z^+) \mid \mathcal{A}\right] \Pr\left[\mathcal{A}\right] \geq 3 \Pr\left[\mathcal{A}\right],$$

since $|z^+| \geq 3$ implies $|g(z^+)| \geq 3$.

On the other hand,

$$\mathrm{E}\left[g(z^-)\chi(\mathcal{A})\right] = \sum_{j=1}^{n} g_j \,\mathrm{E}\left[z_j^- \mid \mathcal{A}\right] \Pr\left[\mathcal{A}\right].$$

For all $j \in \{1, \ldots, n\}$, the condition $\mathcal{A}$ does not increase the expected value of z_j^-. Therefore,

$$\mathrm{E}\left[z_j^- \mid \mathcal{A}\right] \leq \mathrm{E}\left[z_j^-\right] \leq \tfrac{1}{2n} \cdot 1 + \tfrac{1}{2n} \cdot 2 = \tfrac{3}{2n},$$

where for the last inequality we pessimistically assume that we have $x_j = 0$. Together, this gives us

$$\mathrm{E}\left[g(z^-)\chi(\mathcal{A})\right] \leq \sum_{j=1}^{n} \tfrac{3}{2n} g_j \Pr\left[\mathcal{A}\right]$$

and by applying (3) we get

$$\mathrm{E}\left[g(z^-)\chi(\mathcal{A})\right] \leq \tfrac{3(\mathrm{e}-1)}{2} \Pr\left[\mathcal{A}\right] \leq 3 \Pr\left[\mathcal{A}\right].$$

The proposition follows from

$$\mathrm{E}\left[g(z)\chi(\mathcal{A}_i)\right] = \mathrm{E}\left[g(z^+)\chi(\mathcal{A}_i)\right] - \mathrm{E}\left[g(z^-)\chi(\mathcal{A}_i)\right].$$

In the proof of Proposition 8 we will review this argument. For this, note that we did not actually needed the condition $|z^+| \geq 3$ and that the weaker condition $g(z^+) \geq 3(\mathrm{e}-1)/2$ is sufficient to prove non-negative drift. □

The substantial contribution to the sum in (4) comes from the second term which is of order $\Omega(g(x)/n)$.

Proposition 7. *It holds that*

$$\mathrm{E}\left[g(z)\chi(|z^+| = 1 \wedge f(z) \geq 0)\right] \geq \tfrac{3-\mathrm{e}}{8\mathrm{e}^2} \cdot \tfrac{g(x)}{n}.$$

Proof. For all $i \in \{1, \ldots, n\}$ let $\mathcal{A}_i$ be the event that z_i^+ is the i-th vector of unity and that $f(z) \geq 0$.

Then the conditions $|z^+| = 1$ and $f(z) \geq 0$ imply that $\mathcal{A}_i$ holds for exactly one position $i \in \{1, \ldots, n\}$ and

$$\mathrm{E}\left[g(z)\chi(|z^+| = 1 \wedge f(z) \geq 0)\right] = \sum_{i=1}^{n} \mathrm{E}\left[g(z)\chi(\mathcal{A}_i)\right].$$

Like in the proof of Proposition 7, we consider the (conditional) expected values of $g(z^+)$ and $g(z^-)$ independently.

Let $i \in \{1, \ldots, n\}$ with $\Pr[\mathcal{A}_i] > 0$. Then on the one hand,

$$\mathrm{E}\left[g(z^+)\chi(\mathcal{A}_i)\right] = \mathrm{E}\left[g(z^+) \mid \mathcal{A}_i\right] \Pr\left[\mathcal{A}_i\right] \geq g_i \Pr\left[\mathcal{A}_i\right].$$

On the other hand,

$$\mathrm{E}\left[g(z^-)\chi(\mathcal{A}_i)\right] = \sum_{j=1}^{n} g_j \,\mathrm{E}\left[z_j^- \mid \mathcal{A}_i\right] \Pr\left[\mathcal{A}_i\right].$$

For all $j \in \{1, \ldots, i-1\}$, the condition $\mathcal{A}_i$ does not increase the expected value of z_j^-. Therefore,

$$\mathrm{E}\left[z_j^- \mid \mathcal{A}_i\right] \leq \mathrm{E}\left[z_j^-\right] \leq \tfrac{1}{2n} \cdot 1 + \tfrac{1}{2n} \cdot 2 = \tfrac{3}{2n},$$

where we pessimistically assume that $x_i = 0$ in the last inequality. However, for all $j \in \{i+1, \ldots, n\}$, condition $\mathcal{A}_i$ implies $f(z^-) \leq w_i$ and therefore $z_j^- \leq 1$. Hence, in this case

$$\mathrm{E}\left[z_j^- \mid \mathcal{A}_i\right] \leq \mathrm{E}\left[z_j^-\right] \leq \tfrac{1}{2n} \cdot 1 = \tfrac{1}{2n},$$

where again we pessimistically assume that $x_i \neq 2$ in the last inequality. Together, this gives us

$$\mathrm{E}\left[g(z^-)\chi(\mathcal{A}_i)\right] \leq \sum_{j=1}^{i-1} \tfrac{3}{2n} g_j \Pr\left[\mathcal{A}_i\right] + \sum_{j=i+1}^{n} \tfrac{1}{2n} g_j \Pr\left[\mathcal{A}_i\right]$$

and by using (3) we get

$$\sum_{j=1}^{i-1} \tfrac{3}{2n} g_j = \tfrac{3}{2}(g_i - 1)$$

and

$$\sum_{j=i+1}^{n} \tfrac{1}{2n} g_j = \tfrac{1}{2n}\Big(\sum_{j=1}^{n} g_j - \sum_{j=1}^{i} g_j\Big) = \tfrac{1}{2}(g_{n+1} - g_{i+1}).$$

Therefore, with $g_{n+1} \leq \mathrm{e}$ and $g_{i+1} \geq g_i$, we have

$$\mathrm{E}\left[g(z^-)\chi(\mathcal{A}_i)\right] \leq \left(g_i - \tfrac{3-\mathrm{e}}{2}\right) \Pr\left[\mathcal{A}_i\right].$$

Since

$$\mathrm{E}\left[g(z)\chi(\mathcal{A}_i)\right] = \mathrm{E}\left[g(z^+)\chi(\mathcal{A}_i)\right] - \mathrm{E}\left[g(z^-)\chi(\mathcal{A}_i)\right]$$

and

$$\Pr\left[\mathcal{A}_i\right] \geq \tfrac{1}{2n}\left(1-\tfrac{1}{1n}\right)^{n-1} \geq \tfrac{1}{2\mathrm{e}n},$$

and $g_i x_i \leq 2\mathrm{e}$ we have

$$\mathrm{E}\left[g(z)\chi(\mathcal{A}_i)\right] \geq \tfrac{3-\mathrm{e}}{8\mathrm{e}^2 n} g_i x_i$$

for all $i \in \{1,\ldots,n\}$ with $x_i > 0$. Hence,

$$\mathrm{E}\left[g(z)\chi(\mathcal{A})\right] \geq \sum_{i\colon x_i>0} \tfrac{3-\mathrm{e}}{8\mathrm{e}^2 n} g_i x_i = \tfrac{3-\mathrm{e}}{8\mathrm{e}^2} \cdot \tfrac{g(x)}{n}$$

and the proposition follows. □

In order to complete the proof of Lemma 4 and thus of Theorem 1, we still need to show that the third term in the sum of (4) is non-negative.

Proposition 8. *Let $n \geq 200$. Then it holds that*

$$\mathrm{E}\left[g(z)\chi(|z^+| = 2 \wedge f(z) \geq 0)\right] \geq 0.$$

Proof. For all $i \in \{1,\ldots,n\}$ and $j \in \{i,\ldots,n\}$ let $\mathcal{A}_{i,j}$ be the event that $z_j^+ = e_i + e_j$ (where e_i and e_j are the i-th and, respectively, j-th vector of unity), and that $f(z) \geq 0$. Furthermore, let

$$p_{i,j} := \begin{cases} \left(1-\frac{1}{n}\right)^{n-1}\frac{1}{2n} & \text{if } i=j,\\ \left(1-\frac{1}{n}\right)^{n-2}\frac{1}{4n^2} & \text{otherwise.}\end{cases}$$

Then either $\Pr[\mathcal{A}_{i,j}] = 0$ (which happens if $x_i = 0$, or $x_j = 0$, or $x_j = 1$ and $i = j$), or $\Pr[\mathcal{A}_{i,j}] \geq p_{i,j}$.

Then the conditions $|z^+| = 2$ and $f(z) \geq 0$ imply that $\mathcal{A}_{i,j}$ holds for exactly one such pair (i,j) and therefore

$$\mathrm{E}\left[g(z)\chi(|z^+| = 2 \wedge f(z) \geq 0)\right] = \sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathrm{E}\left[g(z)\chi(\mathcal{A}_{i,j})\right].$$

Let $i \in \{1,\ldots,n\}$ and $j \in \{i+1,\ldots,n\}$ be two indices such that $\Pr[\mathcal{A}_{i,j}] \geq p_{i,j}$.

We may assume that $g_i + g_j \leq 3(\mathrm{e}-1)/2$. The reason for this is given at the end of the proof of Proposition 6. Let

$\mathcal{B}$ be the event that $z_\ell^- = 0$ for all $\ell \in \{j+1,\ldots,n\}$.

For all $k \in \{j+1,\ldots,n\}$ let

$\mathcal{C}_k$ be the event that $z_k^- = 1$ and that $z_\ell^- = 0$ for all $\ell \in \{j+1,\ldots,n\} \setminus \{k\}$, and let

$\mathcal{D}_k$ be the event that $z^- = 2e_k$.

For all $k \in \{j+1,\ldots,n\}$ and $\ell \in \{k+1,\ldots,n\}$ let

$\mathcal{E}_{k,\ell}$ be the event that $z^- = e_k + e_\ell$.

Then all of these events are pairwise disjoint.

Conditioned on $\mathcal{A}_{i,j}$ we have $f(z^-) \leq w_i + w_j$. Since $w_k \geq w_j$ for all $k > j$, this implies

$$\sum_{m=j+1}^{n} z_m^- \leq 2.$$

That is, there is either no position $k > j$ such that $z_k^- > 0$ (event $\mathcal{B}$), or exactly one position $k > j$ such that $z_k^- = 1$ (event $\mathcal{C}_k$), or exactly one position $k > j$ such that $z_k^- = 2$ (event $\mathcal{D}_k$), or exactly two distinct positions $\ell > k > j$ such that $z_k^- = 1$ and $z_\ell^- = 1$ (event $\mathcal{E}_{k,\ell}$). Note that for the events $\mathcal{D}_k$ and $\mathcal{E}_{k,\ell}$ we have

$$\sum_{m=j+1}^{n} z_m^- = 2.$$

In this case, we necessarily have

$$w_i = w_j = w_k\ (= w_\ell)$$

and hence also

$$\sum_{m=1}^{j-1} z_m^- = 0,$$

that is, $z_m^- = 0$ unless $m = k$ or $m = \ell$.

Therefore, conditioned on $\mathcal{A}_{i,j}$, one of the events $\mathcal{B}$, $\mathcal{C}_k$, $\mathcal{D}_k$, or $\mathcal{E}_{k,\ell}$ holds with $k \in \{j+1,\ldots,n\}$ and $\ell \in \{k+1,\ldots,n\}$. Thus,

$$\begin{aligned}\mathrm{E}\left[g(z)\chi(\mathcal{A}_{i,j})\right] &= \mathrm{E}\left[g(z)\chi(\mathcal{B} \wedge \mathcal{A}_{i,j})\right]\\ &\quad+ \sum_{k=j+1}^{n} \mathrm{E}\left[g(z)\chi(\mathcal{C}_k \wedge \mathcal{A}_{i,j})\right]\\ &\quad+ \sum_{k=j+1}^{n} \mathrm{E}\left[g(z)\chi(\mathcal{D}_k \wedge \mathcal{A}_{i,j})\right]\\ &\quad+ \sum_{k=j+1}^{n}\sum_{\ell=k+1}^{n} \mathrm{E}\left[g(z)\chi(\mathcal{E}_{k,\ell} \wedge \mathcal{A}_{i,j})\right].\end{aligned}$$

Since $\Pr[\mathcal{A}_{i,j}] \geq p_{i,j}$, we know that $x_i \geq 1$ and $x_j \geq 1$ if $i \neq j$ and $x_i = x_j = 2$ if $i = j$ (otherwise this probability would be zero). In this case we have $\Pr[z = e_i + e_j] = p_{i,j}$. Now, the event "$z = e_i + e_j$" is a sub-event of the event $\mathcal{B} \wedge \mathcal{A}_{i,j}$. Therefore, we also have $\Pr[\mathcal{B} \wedge \mathcal{A}_{i,j}] \geq p_{i,j}$. Hence

$$\mathrm{E}\left[g(z)\chi(\mathcal{B} \wedge \mathcal{A}_{i,j})\right] = \mathrm{E}\left[g(z) \,\middle|\, \mathcal{B} \wedge \mathcal{A}_{i,j}\right]\Pr\left[\mathcal{B} \wedge \mathcal{A}_{i,j}\right].$$

To condition on $\mathcal{B}$ and $\mathcal{A}_{i,j}$ implies that $z_\ell^- = 0$ for all $\ell > j$ and does not increase the expected value of z_ℓ^- for $\ell < j$. Therefore, similar to Proposition 7 and Proposition 6

$$\mathrm{E}\left[g(z) \,\middle|\, \mathcal{B} \wedge \mathcal{A}_{i,j}\right] \geq g_i + g_j - \tfrac{3}{2n}\sum_{\ell=1}^{j-1} g_i \geq \tfrac{5}{2} - \tfrac{1}{2}g_j$$

by (3) and $g_i \geq 1$. Thus,

$$\mathrm{E}\left[g(z)\chi(\mathcal{B} \wedge \mathcal{A}_{i,j})\right] \geq \left(\tfrac{5}{2} - \tfrac{1}{2}g_j\right)\Pr[\mathcal{B} \wedge \mathcal{A}_{i,j}] \tag{5}$$

Next, let $k \in \{j+1,\ldots,n\}$ such that $\Pr[\mathcal{C}_k \wedge \mathcal{A}_{i,j}] > 0$. Then

$$\mathrm{E}\left[g(z)\chi(\mathcal{C}_k \wedge \mathcal{A}_{i,j})\right] = \mathrm{E}\left[g(z) \,\middle|\, \mathcal{C}_k \wedge \mathcal{A}_{i,j}\right]\Pr\left[\mathcal{C}_k \wedge \mathcal{A}_{i,j}\right].$$

To condition on $\mathcal{C}_k$ and $\mathcal{A}_{i,j}$ does imply $z_k = 1$, $z_\ell^- = 0$ for all $\ell > j$ with $\ell \neq k$ and does not increase the expected value of z_ℓ^- for $\ell < j$. Therefore,

$$\mathrm{E}\left[g(z) \,\middle|\, \mathcal{C}_k \wedge \mathcal{A}_{i,j}\right] \geq g_i + g_j - g_k - \tfrac{3}{2n} \sum_{\ell=1}^{j-1} g_i \geq g_j + g_k$$

where we use (3) and the facts that $3/2 \geq (g_i + g_j)/(\mathrm{e} - 1)$ and $g_i \geq g_j/(3(\mathrm{e} - 1)/2 - 1)$.

Furthermore, if z satisfies $\mathcal{C}_k$ and $\mathcal{A}_{i,j}$ then $z' = z + e_k$ satisfies $\mathcal{B}$ and $\mathcal{A}_{i,j}$. Since the k-th entry flips in z but not in z' these two probabilities differ by a factor of $(1/2n)/(1+1/n)$ and we have

$$\Pr\left[\mathcal{C}_k \wedge \mathcal{A}_{i,j}\right] \leq \tfrac{1}{2(n-1)} \Pr\left[\mathcal{B} \wedge \mathcal{A}_{i,j}\right]$$

Thus,

$$\mathrm{E}\left[g(z)\chi(\mathcal{C}_k \wedge \mathcal{A}_{i,j})\right] \geq \left(g_j - g_k\right) \tfrac{1}{2(n-1)} \Pr\left[\mathcal{B} \wedge \mathcal{A}_{i,j}\right]$$

and, since for $n \geq 200$ we have $n\mathrm{e}/(n-1) \leq 11/4$, we get

$$\sum_{k=j+1}^{n} \mathrm{E}\left[g(z)\chi(\mathcal{C}_k \wedge \mathcal{A}_{i,j})\right] \geq \left((1 - \tfrac{j}{2n})g_j - \tfrac{11}{8}\right) \Pr\left[\mathcal{B} \wedge \mathcal{A}_{i,j}\right]$$

Hence, together with (5), we have

$$\begin{aligned}
&\mathrm{E}\left[g(z)\chi(\mathcal{B} \wedge \mathcal{A}_{i,j})\right] + \sum_{k=j+1}^{n} \mathrm{E}\left[g(z)\chi(\mathcal{C}_k \wedge \mathcal{A}_{i,j})\right] \\
&\geq \left(\tfrac{9}{8} + (\tfrac{1}{2} - \tfrac{j}{2n})g_j\right) \Pr\left[\mathcal{B} \wedge \mathcal{A}_{i,j}\right] \\
&\geq \left((4 - 4\tfrac{j}{n})g_j + 9\right) \tfrac{p_{i,j}}{8}
\end{aligned}$$

Next, let $k \in \{j+1, \ldots, n\}$ such that $\Pr[\mathcal{D}_k \wedge \mathcal{A}_{i,j}] > 0$. This implies that $z_k = e_i + e_j - 2e_k$ and in particular that $w_i = w_j = w_k$. In this case,

$$\mathrm{E}[g(z) \,|\, \mathcal{D}_k \wedge \mathcal{A}_{i,j}] = g_i + g_j - 2g_k \geq 1 + g_j - 2g_k$$

and

$$\Pr[\mathcal{D}_k \wedge \mathcal{A}_{i,j}] \leq \tfrac{1}{2(n-1)} p_{i,j}.$$

Hence,

$$\mathrm{E}[g(z)\chi(\mathcal{D}_k \wedge \mathcal{A}_{i,j})] \geq \Big(1 + g_j - 2g_k\Big) \tfrac{1}{2(n-1)} p_{i,j}$$

and (again using $n\mathrm{e}/(n-1) \leq 11/4$) we have

$$\sum_{k=j+1}^{n} \mathrm{E}[g(z)\chi(\mathcal{D}_k \wedge \mathcal{A}_{i,j})] \geq \left((12 - 4\tfrac{j}{n})g_j - 18 - 4\tfrac{j}{n}\right) \tfrac{p_{i,j}}{8}$$

Finally, let $k \in \{j+1, \ldots, n\}$ and $\ell \in \{k+1, \ldots, n\}$ such that $\Pr[\mathcal{D}_k \wedge \mathcal{A}_{i,j}] > 0$. This implies that $z_k = e_i + e_j - e_k - e_\ell$ and in particular that $w_i = w_j = w_k = w_\ell$. In this case,

$$\mathrm{E}[g(z) \,|\, \mathcal{D}_k \wedge \mathcal{A}_{i,j}] \geq 1 + g_j - g_k - g_\ell$$

and

$$\Pr[\mathcal{D}_k \wedge \mathcal{A}_{i,j}] \leq \left(1 - \tfrac{1}{n}\right)^{-2} 4n^2 p_{i,j}.$$

Hence,

$$\mathrm{E}[g(z)\chi(\mathcal{D}_k \wedge \mathcal{A}_{i,j})] \geq \left(1 + g_j - g_k - g_\ell\right)\left(1 - \tfrac{1}{n}\right)^{-2} \tfrac{1}{4n^2} p_{i,j}$$

and with $e/(1-1/n)^2 \leq 11/4$ we get

$$\begin{aligned}
&\sum_{k=j+1}^{n} \sum_{\ell=k+1}^{n} \mathrm{E}[g(z)\chi(\mathcal{D}_k \wedge \mathcal{A}_{i,j})] \\
&\geq \left((3 - 4\tfrac{j}{n})g_j - 5 + 4\tfrac{j}{n}\right) \tfrac{p_{i,j}}{8}.
\end{aligned}$$

Summing up, we get

$$\mathrm{E}\left[g(z)\chi(\mathcal{A}_{i,j})\right] \geq \left((9 - 6\tfrac{j}{n})g_j - 7\right) \tfrac{p_{i,j}}{4}.$$

To do so, we apply

$$g_j = \left(1 + \tfrac{1}{n}\right)^{j-1} \geq 0.99\mathrm{e}^{j/n},$$

where the last inequality follows from $1 + a \geq \mathrm{e}^{a - a^2}$ for all $a \in [0,1]$ and from $n \geq 200$. Now, we substitute $\alpha := j/n$ and verify that the function

$$h(\alpha) := 0.99(9 - 6\alpha)\mathrm{e}^\alpha - 7$$

is positive on the interval $[0,1]$. This holds since we have

$$h'(\alpha) = 0.99(3 - 6\alpha)\mathrm{e}^\alpha$$

and

$$h''(\alpha) = -0.99(3 + 6\alpha)\mathrm{e}^\alpha,$$

that is on the interval $[0,1]$ the function h has a maximum at $\alpha = 1/2$ and two minima

$$h(0) = 1.91 \quad \text{and} \quad h(1) = 2.97\mathrm{e} - 7 > 0$$

at $\alpha = 0$ and $\alpha = 1$ which are both positive. □

4 Limitations of Fixed Potential Functions

In the previous section we have seen how to use a single linear potential function to analyze all linear functions over $\{0,1,2\}^n$. One may ask whether this is possible also for linear function over $\{0,\ldots,r\}^n$ with larger r, potentially even growing in n.

In this section, we show that this is not the case. Already for $r = 3$, the proof in Section 3 does not generalize immediately. More importantly, we show that for $r \geq 43$, there exists no single ("universal") linear potential function over $\{0,1,\ldots,r\}^n$ that allows us to bound the expected runtime of the (1+1) EA on all linear functions over the space $\{0,1,\ldots,r\}^n$. In this, we do not try to optimize the constant, that is, we do not know exactly up to which constant r the proof method of the previous section can be applied and from which point on the non-existence results for universal drift functions presented in this section holds.

In the previous section, we showed that our potential function g has a positive drift of $\Omega(g(x)/n)$ in every single point $x \in \{0,1,2\}^n$. In [14], Jägersküpper shows that this assumption is not necessary. Instead, he shows that for the search space $\{0,1\}^n$ it is sufficient to assume that x is chosen according to a semi-balanced distribution.

Definition 9 (semi-balanced)**.** *Let n and r be two positive integers. Let $\mathcal{D}$ be a distribution over $\{0, 1 \ldots, r\}^n$. Then, $\mathcal{D}$ is* semi-balanced *if for $x \in \{0, \ldots, r\}^n$ drawn according to $\mathcal{D}$ it holds that*

$$\Pr(x_i = a \wedge x_j = b) \leq \Pr(x_i = b \wedge x_j = a)$$

for all $i, j \in \{1 \ldots n\}$ and $a, b \in \{0, \ldots, r\}$ with $i < j$ and $a < b$.

It turns out that it is difficult to extend Jägersküpper's result and show that in each iteration of the (1+1) EA optimizing a linear function over $\{0, 1, \ldots, r\}^n$ the search point is semi-balanced. However, even under the assumption that this is possible, we see that ultimately this alone will not yield a proof of a $O(n \log n)$ runtime bound for $r \geq 43$.

Theorem 10. *Let $r \geq 43$ and let*

$$g\colon \{0, 1, \ldots, r\}^n \to \mathbb{R} \quad g\colon x \mapsto \sum_{i=1}^{n} g_i x_i$$

be a linear potential function with real-valued weights

$$0 < g_1 \leq \cdots \leq g_n.$$

Then there exists a linear fitness function

$$f\colon \{0, 1, \ldots, r\}^n \to \mathbb{R} \quad f\colon x \mapsto \sum_{i=1}^{n} w_i x_i$$

with real-valued weights

$$0 < w_1 \leq \cdots \leq w_n$$

and a semi-balanced distribution $\mathcal{D}$ on $\{0, 1, \ldots, r\}^n$ such that if x is drawn according to $\mathcal{D}$ and $y = \mathrm{mut}(x)$, then

$$\mathrm{E}_{x,y}\left[(g(x) - g(y))\chi(f(y) \leq f(x))\right] \leq 0\,.$$

where again $\chi(\mathcal{A}) \in \{0, 1\}$ is the random variable that denotes the indicator function of the event $\mathcal{A}$.

Note that this theorem does not imply that the runtime of the (1+1) EA on linear functions over $\{0, 1, \ldots, r\}^n$ is not in $O(n \log n)$ (which we actually conjecture to be true). Nor does it say that this result cannot be obtained using (multiplicative) drift analysis. Instead, it states that if such a proof is possible, the used potential function has to include more information on the weights $w_1, \ldots, w_n$ than just their order.

To prove the previous theorem, we first show two propositions. Without loss of generality, we may assume that $g_0 = 1$ since Theorem 3 is invariant under scaling. The first proposition tells us that the sum of weights in a suitable potential function may not be to large.

Proposition 11. *Let n and r be two positive integers. Let*

$$g\colon \{0, 1, \ldots, r\}^n \to \mathbb{R} \quad g\colon x \mapsto \sum_{i=1}^{n} g_i x_i$$

be a linear potential function with real-valued weights

$$1 = g_1 \leq \cdots \leq g_n$$

such that

$$\sum_{i=1}^{n} g_i > (r+1)n. \tag{6}$$

Let

$$f\colon \{0, 1, \ldots, r\}^n \to \mathbb{R} \quad f\colon x \mapsto \sum_{i=1}^{n} x_i$$

and let $\mathcal{D}$ be the semi-balanced distribution on $\{0, 1, \ldots, r\}^n$ that chooses e_1, i.e., the first unit vector, with certainty.

Suppose x is drawn according to $\mathcal{D}$ and $y = \mathrm{mut}(x)$, then

$$\mathrm{E}_{x,y}\left[(g(x) - g(y))\chi(f(y) \leq f(x))\right] < 0\,.$$

Proof. We know that $x = e_1$ with $f(x) = w_1 = 1$ and $g(x) = g_1 = 1$. Hence, the only random variable we have to consider is $y = \mathrm{mut}(e_1)$. Let

$$\Delta := (g(x) - g(y))\chi(f(y) \leq f(x)) = (1 - g(y))\chi(f(y) \leq 1).$$

There are three cases such that $f(y) \leq 1$.

The first case is that $y = x = e_1$. The probability of this event is $(1 - 1/n)^n$ and conditioned on this event we have $\Delta = 0$.

The second case is that $y = (0, \ldots, 0)$. The probability of this event is $(1/rn)(1 - 1/n)^{n-1}$ and conditioned on this event we have $\Delta = 1$.

The third case is that $y = e_i$ with $i \geq 2$. For all $i \geq 2$, the probability that $y = e_i$ is $(1/rn)^2(1 - 1/n)^{n-2}$ and in this case we have $\Delta = 1 - g_i$. Thus,

$$\begin{aligned}\mathrm{E}_{x,y}[\Delta] &= \tfrac{1}{rn}\left(1 - \tfrac{1}{n}\right)^{n-1} + \sum_{i=2}^{n}\left(\tfrac{1}{rn}\right)^2\left(1 - \tfrac{1}{n}\right)^{n-2}\left(1 - g_i\right)\\ &= \tfrac{1}{rn}\left(1 - \tfrac{1}{n}\right)^{n-2}\left(\tfrac{(r+1)n-1}{rn} - \tfrac{1}{rn}\sum_{i=1}^{n} g_i\right).\end{aligned}$$

The proposition follows from (6) □

The second proposition tells us that the sum of the first k weights in a suitable potential function may not grow too slowly.

Proposition 12. *Let n and r be two positive integers. Let*

$$g\colon \{0, 1, \ldots, r\}^n \to \mathbb{R} \quad g\colon x \mapsto \sum_{i=1}^{n} g_i x_i$$

be a linear potential function with real-valued weights

$$1 = g_1 \leq \cdots \leq g_n$$

such that

$$\sum_{i=1}^{k} g_i > \tfrac{r+3}{r+1}\sum_{i=1}^{k}\left(1 - \tfrac{1}{n}\right)^{k-i} g_i \tag{7}$$

Let $k \in \{1, \ldots, n\}$. Let

$$f\colon \{0, 1, \ldots, r\}^n \to \mathbb{R} \quad f\colon x \mapsto \sum_{i=1}^{n} (r+1)^i x_i$$

and let $\mathcal{D}_k$ be the (semi-balanced) uniform distribution over $\{e_1, \ldots, e_k\}$.

Suppose x is drawn according to $\mathcal{D}$ and $y = \mathrm{mut}(x)$, then

$$\mathrm{E}_{x,y}\left[(g(x) - g(y))\chi(f(y) \le f(x))\right] < 0\,.$$

Proof. Let $\Delta := (g(x) - g(y))\chi(f(y) \le f(x))$. Then

$$\mathrm{E}_{x,y}[\Delta] = \tfrac{1}{k}\sum_{i=1}^{k}\mathrm{E}_{x,y}[\Delta \mid x = e_i].$$

Suppose that $x = e_i$ with $i \in \{1, \ldots, k\}$. Then we have that $\Delta \ne 0$ if and only if $y_i = \cdots = y_n = 0$ and this event happens with probability $\frac{1}{rn}\left(1 - \frac{1}{n}\right)^{n-i}$. Furthermore, we have

$$\mathrm{E}_{x,y}[\Delta \mid x = e_i \wedge y_i = \cdots = y_n = 0] = g_i - \sum_{j=1}^{i-1} g_j\,\mathrm{E}[y_j].$$

Therefore, since $\mathrm{E}[y_j] = \frac{r+1}{2n}$, we get

$$\mathrm{E}_{x,y}[\Delta \mid x = e_i \wedge y_i = \cdots = y_n = 0] = g_i - \tfrac{r+1}{2n}\sum_{j=1}^{i-1} g_j.$$

It follows that

$$\mathrm{E}_{x,y}[\Delta \mid x = e_i] = \tfrac{1}{rn}\left(1 - \tfrac{1}{n}\right)^{n-i}\left(g_i - \tfrac{r+1}{2n}\sum_{j=1}^{i-1} g_j\right)$$

and thus

$$\mathrm{E}_{x,y}[\Delta] = \tfrac{1}{krn}\sum_{i=1}^{k}\left(1 - \tfrac{1}{n}\right)^{n-i}\left(g_i - \tfrac{r+1}{2n}\sum_{j=1}^{i-1} g_j\right)$$

Furthermore, we see that

$$\sum_{i=1}^{k}\sum_{j=1}^{i-1}\left(1 - \tfrac{1}{n}\right)^{k-i} g_j = n\sum_{i=1}^{k}\left(1 - \left(1 - \tfrac{1}{n}\right)^{k-i}\right) g_i$$

and therefore

$$\mathrm{E}_{x,y}[\Delta] = \frac{\left(1 - \frac{1}{n}\right)^{n-k}}{2krn}\sum_{i=1}^{k}\left((r+3)\left(1 - \tfrac{1}{n}\right)^{k-i} - (r+1)\right) g_i.$$

The proposition follows from (7). ☐

To prove Theorem 10, we now need to show that at least one of the two inequalities (6) or (7) holds for $r \ge 43$.

of Theorem 10. Let $\ell := \lceil \frac{2n}{r+1} \rceil$. Aiming at a contradiction, suppose that neither (6) nor (7) holds. Then we have by (7) that

$$\sum_{i=1}^{k} g_i \le \left(1 + \tfrac{2}{r+1}\right)\sum_{i=1}^{k}\left(1 - \tfrac{1}{n}\right)^{k-i} g_i$$

for all $k \in \{\ell + 1, \ldots, n\}$ simultaneously.

For all $i \in \{1, \ldots, k\}$ let $a_i := \left(1 + \frac{2}{r+1}\right)\left(1 - \frac{1}{n}\right)^{k-i}$. Then

$$\sum_{i=1}^{k} g_i = \sum_{i=1}^{k-\ell} a_i g_i + \sum_{i=k-\ell+1}^{k} a_i g_i.$$

Since $a_i \le 1 + \frac{2}{r+1}$ for all $i \le k$, we have

$$\sum_{i=1}^{k} g_i \le \sum_{i=1}^{k-\ell} a_i g_i + \sum_{i=k-\ell+1}^{k}\left(1 + \tfrac{2}{r+1}\right) g_i$$

and hence

$$\sum_{i=1}^{k-\ell}(1 - a_i) g_i \le \frac{2}{r+1}\sum_{i=k-\ell+1}^{k} g_i. \tag{8}$$

Since it holds for all $z \in \mathbb{R}$ that $1 + z \le \mathrm{e}^z$, we have

$$a_i \le \mathrm{e}^{\frac{2}{r+1} - \frac{k-i}{n}} \le 1$$

for all $1 \le i \le k - \ell$. Thus, the coefficients on the left hand-side of (8) are all positive. Since $1 = g_1 \le \cdots \le g_k$, we obtain

$$\sum_{i=1}^{k-\ell}(1 - a_i) \le \frac{2}{r+1}\sum_{i=k-\ell+1}^{k} g_i = \frac{2\ell}{r+1} g_k.$$

Now,

$$\begin{aligned}\sum_{i=1}^{k-\ell} a_i &= \left(1 + \tfrac{2}{r+1}\right)\sum_{i=1}^{k-\ell}\left(1 - \tfrac{1}{n}\right)^{k-i} \\ &= a_{k-\ell}\sum_{i=1}^{k-\ell}\left(1 - \tfrac{1}{n}\right)^{k-\ell-i} \\ &= a_{k-\ell}\sum_{i=0}^{k-\ell-1}\left(1 - \tfrac{1}{n}\right)^{i} \\ &= a_{k-\ell}\, n\left(1 - \left(1 - \tfrac{1}{n}\right)^{k-\ell}\right) \\ &\le n - \left(1 - \tfrac{1}{n}\right)^{k-\ell} n.\end{aligned}$$

Thus,

$$g_k \ge \tfrac{r+1}{2\ell}\left(k - \ell - n + \left(1 - \tfrac{1}{n}\right)^{k-\ell} n\right)$$

and

$$\begin{aligned}\sum_{k=\ell+1}^{n} g_k &\ge \tfrac{r+1}{2\ell}\sum_{k=\ell+1}^{n}\left(k - \ell - n + \left(1 - \tfrac{1}{n}\right)^{k-\ell} n\right) \\ &= \tfrac{r+1}{2\ell}\sum_{k=1}^{n-\ell}\left(k - n + \left(1 - \tfrac{1}{n}\right)^{k} n\right) \\ &= \tfrac{r+1}{2\ell}\left(\sum_{k=1}^{n-\ell} k - (n-\ell)n + n\sum_{k=1}^{n-\ell}\left(1 - \tfrac{1}{n}\right)^{k}\right) \\ &\ge \tfrac{r+1}{4\ell}\left(\left(1 - 2\mathrm{e}^{-(n-\ell)/n}\right)n(n-1) + \ell(\ell-1)\right) \\ &\ge \tfrac{\left(1 - 2\mathrm{e}^{-(r-2)/(r+1)}\right)(r+1)^2}{9} n \\ &> (r+1)n\end{aligned}$$

for $r \ge 43$. This is a contradiction to (6). ☐

5 Conclusion

We have seen that drift analysis can be applied to analyze the (1+1) EA on linear functions over $\{0,1,2\}^n$. However, for larger values of r, the approach to model the behavior of the (1+1) EA for all linear functions over $\{0,1,\ldots,r\}^n$ with a single linear potential function fails, even if we invoke a balance argument as in [14].

Observing such difficulties already for relatively simple optimization problems as the one regarded here, we feel that it is still a long way to go until drift analysis is sufficiently well-understood to solve problems like proving a runtime of $\Theta(m^2 \log m)$ for the (1+1) EA on the minimum spanning tree problem (which we conjecture to be true).

We still believe that (multiplicative) drift analysis is an adequate way to approach this problem. Most likely, one needs a more problem-specific potential function that reveals (or needs?) more insight into the structure of the minimum spanning tree problem.

References

[1] A. Auger and B. Doerr, editors. *Theory of Randomized Seach Heuristics*. World Scientific, 2011.

[2] S. Baswana, S. Biswas, B. Doerr, T. Friedrich, P. P. Kurur, and F. Neumann. Computing single source shortest paths using single-objective fitness. In *FOGA '09*, pages 59–66. ACM, 2009.

[3] B. Doerr, E. Happ, and C. Klein. A tight analysis of the (1+1)-EA for the single source shortest path problem. In *CEC '07*, pages 1890–1895. IEEE, 2007.

[4] B. Doerr and D. Johannsen. Edge-based representation beats vertex-based representation in shortest path problems. In *GECCO '10*, pages 759–766. ACM, 2010.

[5] B. Doerr, D. Johannsen, and C. Winzen. Drift analysis and linear functions revisited. In *CEC '10*, pages 1–8. ACM, 2010.

[6] B. Doerr, D. Johannsen, and C. Winzen. Multiplicative drift analysis. In *GECCO '10*, pages 1449–1456. ACM, 2010.

[7] B. Doerr, D. Johannsen, and C. Winzen. Multiplicative drift analysis. In *GECCO '10*, pages 1449–1456. ACM, 2010.

[8] S. Droste, T. Jansen, and I. Wegener. On the analysis of the (1+1) evolutionary algorithm. *Theoretical Computer Science*, 276(1–2):51–81, 2002.

[9] C. Gunia. On the analysis of the approximation capability of simple evolutionary algorithms for scheduling problems. In *GECCO '05*, pages 571–578. ACM, 2005.

[10] B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 14(3):502–525, 1982.

[11] J. He and X. Yao. Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence*, 127(1):57–85, 2001.

[12] J. He and X. Yao. Erratum to: drift analysis and average time complexity of evolutionary algorithms [Artificial Intelligence 127 (2001) 57–85]. *Artificial Intelligence*, 140(1–2):245–248, 2002.

[13] J. He and X. Yao. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing*, 3(1):21–35, 2004.

[14] J. Jägersküpper. A blend of Markov-chain and drift analysis. In *PPSN '08*, pages 41–51. Springer, 2008.

[15] F. Neumann and I. Wegener. Randomized local search, evolutionary algorithms, and the minimum spanning tree problem. *Theoretical Computer Science*, 378(1):32–40, 2007.

[16] F. Neumann and C. Witt. *Bioinspired Computation in Combinatorial Optimization*. Springer, 2010.

[17] J. Reichel and M. Skutella. Evolutionary algorithms and matroid optimization problems. In *GECCO '07*, pages 947–954. ACM, 2007.

[18] J. Reichel and M. Skutella. On the size of weights in randomized search heuristics. In *FOGA '09*, pages 21–28. ACM, 2009.

[19] J. Scharnow, K. Tinnefeld, and I. Wegener. The analysis of evolutionary algorithms on sorting and shortest paths problems. *Journal of Mathematical Modelling and Algorithms*, 3(4):349–366, 2004.