Worst-Case and Average-Case Approximations by Simple Randomized Search Heuristics

Carsten Witt*

FB Informatik, LS 2
Univ. Dortmund, 44221 Dortmund, Germany
carsten.witt@cs.uni-dortmund.de

Abstract. In recent years, probabilistic analyses of algorithms have received increasing attention. Despite results on the average-case complexity and smoothed complexity of exact deterministic algorithms, little is known about the average-case behavior of randomized search heuristics (RSHs). In this paper, two simple RSHs are studied on a simple scheduling problem. While it turns out that in the worst case, both RSHs need exponential time to create solutions being significantly better than 4/3-approximate, an average-case analysis for two input distributions reveals that one RSH is convergent to optimality in polynomial time. Moreover, it is shown that for both RSHs, parallel runs yield a PRAS.

1 Introduction

It is widely acknowledged that worst-case analyses may provide too pessimistic estimations for the runtime of practically relevant algorithms and heuristics. Therefore, in recent years, there has been a growing interest in the probabilistic analysis of algorithms. Famous examples include results on the average-case time complexity of a classical algorithm for the knapsack problem [1] and of the simplex algorithm [2]. Both papers show a polynomial runtime in the even stronger model of so-called smoothed complexity.

Approximation is another way out of this worst-case way of thinking. It is well known that many NP-hard problems allow polynomial-time approximation algorithms or even approximation schemes [3]. However, if even no approximation algorithms are available, one often resorts to heuristic approaches, which are said to provide good solutions within a tolerable span of time. Such approaches may be the only choice if there are not enough resources (time, money, experts, ...) available to design problem-specific (approximation) algorithms.

Many general-purpose heuristics such as the Metropolis algorithm or Simulated Annealing [4] rely on the powerful concept of randomization. Another popular class of randomized search heuristics (RSHs) is formed by the so-called Evolutionary Algorithms (EAs), see [5]. Despite having been applied successfully for more than 30 years, a theoretical foundation of the computational time complexity of EAs has started only recently, see, e. g., [6–10].

* The author was supported by the Deutsche Forschungsgemeinschaft (DFG) as a part of the Collaborative Research Center "Computational Intelligence" (SFB 531).

However, almost all results on the time complexity of RSHs are concerned with exact optimization. Moreover, these results mostly refer to worst-case instances from the class of problems considered. In contrast, the real aims of heuristics are approximation and efficient average-case behavior. Therefore, we should consider these aspects when studying general-purpose heuristics such as EAs. Positive and negative results will help to understand under what circumstances such heuristics can be efficient (approximation) algorithms and to provide guidelines for the practitioner when and how to apply them. Our approach starts by investigating RSHs on well-studied combinatorial problems. Such analyses have already been carried out in the context of exact optimization, e. g., [9, 10]. Of course, our goal is not to compare RSHs with clever problem-specific algorithms.

In this paper, we consider two simple RSHs for a well-known optimization problem, namely the optimization variant of the NP-complete PARTITION problem: Given n positive integers $w_1, \dots, w_n$, find some subset $I \subseteq \{1, \dots, n\}$ such that $m(I) := \max\left\{\sum_{i \in I} w_i, \sum_{i \notin I} w_i\right\}$ becomes minimal. This is one of the easiest-to-state and easiest-to-solve NP-hard problems since it even allows an FPAS [3]; from a practical point of view, it may be regarded as a simple scheduling problem. In fact, there already are some average-case analyses of classical greedy heuristics designed for this problem [11, 12]. We will relate these results to those for the general-purpose RSHs considered by us.

Since the RSHs to be defined have been designed for pseudo-Boolean optimization, we encode a solution to a PARTITION instance by the characteristic vector of I and arrive at the pseudo-Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$, whose value $f(x)$ equals $m(I)$ if x encodes the set I. The following two simple RSHs are sometimes called hillclimbers. They store only one current search point and do not accept worsenings, which are, in this case, search points with some larger f-value. Both can be described by an initialization step and an infinite loop.

(1+1) EA

Initialization: Choose $a \in \{0,1\}^n$ randomly.
Loop: The loop consists of a mutation and a selection step.
Mutation: For each position i, decide independently whether a_i should be flipped (replaced by $1 - a_i$), the flipping probability equals $1/n$.
Selection: Replace a by a' iff $f(a') \leq f(a)$.

The (1+1) EA has a positive probability to create any search point from any search point and eventually optimizes each pseudo-Boolean function. This does not hold for Randomized Local Search (RLS), which flips only one bit per step.

RLS

This works like the (1+1) EA with a different mutation operator.
Mutation: Choose $i \in \{1, \dots, n\}$ randomly and flip a_i.

We ignore the stopping criterion needed by practical implementations of these RSHs and are interested in the f-value of the current search point by some time t, i. e., after t iterations of the infinite loop. Mainly, we try to estimate in how far this f-value approximates the optimum if t is bounded by some polynomial.

The paper is structured as follows. In Sect. 2, we provide some basic definitions and proof techniques needed to estimate the progress of the RSHs. In Sect. 3, we prove worst-case results on the approximation ratios obtainable by the RSHs within polynomial time. Moreover, we show that these results can be extended to parallel runs of the RSHs so as to design a randomized approximation scheme. In Sect. 3, we extend our techniques toward a probabilistic average-case analysis for two well-known input distributions. We finish with some conclusions.

2 Definitions and Proof Methods

Throughout the paper, we adopt the following conventions. Given an instance $w_1, \dots, w_n$ for the optimization problem PARTITION, we assume w. l. o. g. that $w_1 \geq \cdots \geq w_n$. Moreover, we set $w := w_1 + \cdots + w_n$. We call the indices $1, \dots, n$ *objects* and call w_i the *volume* of the i-th object. Sometimes, the objects themselves are also called $w_1, \dots, w_n$. The optimization problem can be thought of as putting the objects in one of two bins, and a search point $x \in \{0,1\}^n$ is the characteristic vector of the set of objects put into the first bin. Then the goal function f corresponds to the total volume in the fuller bin w. r. t. x.

We will essentially exploit two proof methods in order to show bounds on the approximative quality of a solution output by the considered RSH. Both techniques only study progresses by so-called local steps, i. e., steps to search points at Hamming distance 1. Therefore, the analyses apply only until points of time where the heuristic RLS is able to get stuck in a local optimum. We are interested in sufficient conditions such that the considered RSH is able to improve the f-value by local steps. A first idea is to bound the volume of the largest object in the fuller bin from above. We do something similar but neglect the objects making the bin the fuller bin.

Definition 1 (Critical volume). *Let $W = (w_1, \dots, w_n)$ be an instance for the partition problem and let $\ell \geq w/2$ be a lower bound on the optimum f-value w. r. t. W. Moreover, let $x \in \{0,1\}^n$ be a characteristic vector s. t. $f(x) > \ell$. Let $w_{i_1} \geq w_{i_2} \geq \cdots \geq w_{i_k}$ be the objects in the fuller bin w. r. t. x, ranked in non-increasing order. Let $r := i_j$ for the smallest j such that $w_{i_1} + \cdots + w_{i_j} > \ell$. The volume w_r is called the* critical volume *(w. r. t. W, ℓ and x).*

The idea behind the critical volume is as follows. Suppose that x is the current search point of the RSH, leading to $f(x) > \ell$, and we know an upper bound v on the critical volume w. r. t. the instance, ℓ and x. Let r be the minimum i s. t. $w_i \leq v$. Due to $w_1 \geq \cdots \geq w_n$, we know that w_r is also an upper bound on the critical volume. By the definition of critical volume, there is some object $w_{r'}$, where $r' \geq r$, in the fuller bin. If we additionally know that $f(x) \geq \ell + v$ holds, together with $\ell \geq w/2$ this implies that $w_{r'}$ can be moved from the fuller bin into the emptier one, decreasing the f-value. Thus, a local step improves x.

The described sufficient condition for locally improvable search points can even be strengthened. Suppose that we have the same setting as before with the exception that now $\ell + v/2 < f(x) < \ell + v$ holds. If $w_{r'} \leq f(x) - w/2$, $w_{r'}$ can still be moved to the emptier bin. Otherwise, this step makes the fuller bin the emptier bin. Since $w_{r'} \leq w_r \leq v$, the total volume in this bin will be greater than $\ell + v/2 - w_r \geq w/2 - w_r/2$ and, therefore, the f-value less than $\ell + v/2$. Hence, the step is accepted by the RSH, too.

For RLS, a local step to a specific point at Hamming distance 1 has probability $1/n$, and for the (1+1) EA, the probability is at least $(1/n)(1-1/n)^{n-1} \geq 1/(en)$. If we know that the critical volume is always bounded by some small value, we can exploit this to show upper bounds on the f-values obtainable in expected polynomial time. The special case $w_1 \geq w/2$ can be solved exactly.

Lemma 1. *Let W and ℓ be as in Definition 1. Suppose that from some time t^* on, the critical volume w. r. t. W, ℓ and the current search point of the (1+1) EA or of RLS is at most v. Then the RSH reaches an f-value at most $\ell+v/2$ if $w_1 < w/2$ and at most w_1 otherwise in an expected number of another $O(n^2)$ steps.*

Proof. Let r be the smallest i where $w_i \leq v$. We consider the run of the RSH only from time t^* on. The proof uses a fitness-level argument [6]. Let $s := w_r + \cdots + w_n$, i. e., the sum of all volumes at most as large as w_r. Note that the conditions of the lemma and the definition of critical volume imply that $f(x) \leq \ell + s$ for all current search points x. According to $w_r, \ldots, w_n$, we partition the set of possible current search points by so-called fitness levels as follows. Let

$$L_i := \left\{ x \;\middle|\; \ell + s - \sum_{j=r}^{r+i-1} w_j \;\geq\; f(x) \;>\; \ell + s - \sum_{j=r}^{r+i} w_j \right\}$$

for $0 \leq i \leq n-r$ and $L_{n-r+1} := \{x \mid \ell = f(x)\}$. Now consider some x such that $f(x) > \ell + w_r/2$. By the definition of critical volume, there must be an object from $w_r, \ldots, w_n$ in the fuller bin whose move to the emptier bin decreases the f-value by its volume or leads to an f-value of at most $\ell + w_r/2 \leq \ell + v/2$. If $x \in L_i$, due to $w_r \geq \cdots \geq w_n$, there is even an object from $w_r, \ldots, w_{r+i}$ with this property. By the above considerations, moving this object to the emptier bin by a local step of the RSH has probability at least $1/(en)$ and, due to $w_r \geq \cdots \geq w_n$, leads to some $x' \in L_j$ such that $j > i$. The expected waiting time for such a step is at most en. After at most $n-r+1$ sets have been left, the f-value has dropped to at most $\ell + w_r/2$. Hence, the total expected time after time t^* is $O(n^2)$.

If $w_1 \geq w/2$, we can apply the previous arguments with the special values $\ell := w_1$ and $r := 2$. The only difference is that in case that $f(x) > \ell$, there must be an object of volume at most $f(x) - \ell$ in the fuller bin. Hence, the RSH cannot be in a local optimum and is able to reach L_{n-r+1} by local steps. □

If we are satisfied with slightly larger f-values than guaranteed by Lemma 1, significantly smaller upper bounds on the expected time can be shown.

Lemma 2. *Let W and ℓ be as in Definition 1. Suppose that from some time t^* on, the critical volume w. r. t. W, ℓ and the current search point of the (1+1) EA or of RLS is at most v. Then for any $\gamma > 1$ and $0 < \delta < 1$, the (1+1) EA (RLS) reaches an f-value at most $\ell + v/2 + \delta w/2$ if $w_1 < w/2$ and at most $w_1 + \delta w/2$*

otherwise in at most $\lceil en \ln(\gamma/\delta)\rceil$ *(*$\lceil n \ln(\gamma/\delta)\rceil$*) another steps with probability at least* $1 - \gamma^{-1}$*. Moreover, the expected number of another steps is at most* $2\lceil en \ln(2/\delta)\rceil$ *(*$2\lceil n \ln(2/\delta)\rceil$*).*

Proof. Let r be the smallest i where $w_i \leq v$. First, we consider the run of the (1+1) EA from time t^* on. Let x be a current search point s. t. $f(x) > \ell + w_r/2$. We are interested in the contribution of the so-called small objects $w_r, \dots, w_n$ to the f-value and want to estimate the average decrease of the f-value by a similar method as presented in [10]. Let $p(x) := \max\{f(x) - \ell - w_r/2, 0\}$ and note that due to the definition of critical volume and the conditions of the lemma, $p(x)$ is a lower bound on the contribution of small objects to $f(x)$. Moreover, as long as $p(x) > 0$, all steps moving only a small object to the emptier bin are accepted and decrease the p-value by its volume or lead to an f-value of at most $\ell + v/2$. Let p_0 be some current p-value. Since a local step of the (1+1) EA has probability at least $1/(en)$, the expected p-decrease is at least $p_0/(en)$ and the expected p-value after the step, therefore, at most $(1 - 1/(en))p_0$. Since the steps of the (1+1) EA are independent, this argumentation remains valid if p_0 is only an expected value and can be iterated until the p-value equals 0. Hence, the expected p-value p_t after t steps is at most $(1 - 1/(en))^t p_0$. For $t' := en \ln(\gamma/\delta)$, we have $p_{t'} \leq \delta p_0/\gamma \leq \delta w/(2\gamma)$. Since the p-value is non-negative, we can apply Markov's inequality, implying $p_{t'} \leq \delta w/2$ with probability at least $1 - 1/\gamma$. Since the previous arguments make no assumptions on p_0, we can repeat independent phases of length $\lceil en \ln(2/\delta)\rceil$. The expected number of phases until the p-value is at most $\delta w/2$ is at most 2, implying the lemma for the case $w_1 < w/2$.

If $w_1 \geq w/2$, we can apply the previous arguments with the special values $\ell := w_1$ and $r := 2$. The only difference is that in case that $f(x) > \ell$, there must be an object of volume at most $f(x) - \ell$ in the fuller bin. Hence, the (1+1) EA cannot be in a local optimum. Redefining $p(x) := f(x) - \ell$, the lemma follows for the (1+1) EA. The statements on RLS follow in the same way, taking into account that a local step has probability $1/n$. □

3 Worst-Case Analyses

In this section, we will study bounds on the approximation ratios obtainable by the RSHs within polynomial time regardless of the problem instance.

Theorem 1. *Let* $\varepsilon > 0$ *be a constant. On any instance for the partition problem, the (1+1) EA and RLS reach an* f*-value that is at least* $(4/3+\varepsilon)$*-approximate in an expected number of* $O(n)$ *steps and an* f*-value that is at least* $4/3$*-approximate in an expected number of* $O(n^2)$ *steps.*

Proof. We start by studying trivial instances with $w_1 \geq w/2$. Then even both statements follow for $\delta := 1/3$ by means of Lemma 2.

Now let $w_1 < w/2$ and $\ell := w/2$. We still have to distinguish two cases. The first case holds if $w_1 + w_2 > 2w/3$. This implies $w_1 > w/3$ and, therefore, $w - w_1 < 2w/3$. Hence, if we start with w_1 and w_2 in the same bin, a step separating

w_1 and w_2 by putting w_2 into the emptier bin is accepted, and these objects will remain separated afterwards. The expected time until such a separating step occurs is $O(n)$. Afterwards, the critical volume according to Definition 1 is always bounded above by w_3. Since $w_3 + \cdots + w_n < w/3$, we know that $w_i < w/3$ for $i \geq 3$. Hence, the first statement of the theorem follows for $\delta := \varepsilon$ by Lemma 2 and the second one by Lemma 1. If $w_1 + w_2 \leq 2w/3$, we have $w_i \leq w/3$ for $i \geq 2$. Since $w_1 < w/2$, this implies that the critical volume is always at most $w_2 \leq w/3$. Therefore, the theorem holds also in this case. □

The approximation ratio 4/3 that the RSHs are able to obtain within expected polynomial time is at least almost tight. Let n be even and $\varepsilon > 0$ be some arbitrarily small constant. Then the instance W^*_ε, an almost worst-case instance, contains two objects w_1 and w_2 of volume $1/3-\varepsilon/4$ each and $n-2$ objects of volume $(1/3+\varepsilon/2)/(n-2)$. Note that the total volume has been normalized to 1 and that the instance has an exponential number of perfect partitions.

Theorem 2. *Let ε be any constant s. t. $0 < \varepsilon < 1/3$. With probability $\Omega(1)$, both the (1+1) EA and RLS take $n^{\Omega(n)}$ steps to create a solution better than $(4/3-\varepsilon)$-approximate for the instance W^*_ε.*

Proof. The proof idea is to show that the RSH reaches a situation where w_1 and w_2 are in one bin and at least $k := n-2-(n-2)\varepsilon/2$ of the remaining so-called small objects are in the other one. Since $\varepsilon < 1/3$, at least k objects yield a total volume of more than $1/3+\varepsilon/4$. To leave the situation by separating w_1 and w_2, the RSH has to transfer small objects of a total volume of at least $\varepsilon/4$ from one bin to the other one in a single step. For this, $(n-2)\varepsilon/2$ small objects are not enough. Flipping $\Omega(n)$ bits in one step of the (1+1) EA has probability $n^{-\Omega(n)}$, and flipping $\Omega(n)$ bits at least once within n^{cn} steps is, therefore, still exponentially unlikely if the constant c is small enough. For RLS, the probability is even 0. Since the total volume in the fuller bin is at least $2/3-\varepsilon/2$ unless w_1 and w_2 are separated, this will imply the theorem.

To show the claim that the described situation is reached with probability $\Omega(1)$, we consider the initial search point of the RSH. With probability 1/2, it puts w_1 and w_2 into the same bin. Therefore, we estimate the probability that enough small objects are transferred from this bin to the other one in order to reach the situation, before a bit at the first two positions (denoting the large objects) flips. In a phase of length cn for any constant c, with probability $(1-2/n)^{cn} = \Omega(1)$, the latter never happens. Under this assumption, each step moving a small object into the emptier bin is accepted. By the same idea as in the proof of Lemma 2, we estimate the expected decrease of the contribution of small objects to the f-value. Reducing it to at most an $\varepsilon/2$-fraction of its initial contribution suffices to obtain at least k objects in the emptier bin. Each step leads to an expected decrease by at least a $1/(en)$-fraction. Since ε is a positive constant, $O(n)$ steps are sufficient to decrease the contribution to at most an expected $\varepsilon/4$-fraction. By Markov's inequality, we obtain the desired fraction within $O(n)$ steps with probability at least 1/2. Since c may be chosen appropriately, this proves the theorem. □

The worst-case example studied in Theorem 2 suggests that the RSH is likely to arrive at a bad approximation if it misplaces objects of high volume. On the other hand, it can easily be shown for the example that the RSH is able to find an optimal solution with probability $\Omega(1)$ in polynomial time if it separates the two largest objects in the beginning. We try to generalize this to arbitrary instances. In order to obtain a $(1+\varepsilon)$-approximation in polynomial time according to Lemma 1, the critical volume should be bounded above by εw. Due to the ordering $w_1 \geq \cdots \geq w_n$, all objects of index at least $s := \lceil 1/\varepsilon \rceil$ are bounded by this volume. Therefore, the crucial idea is to bound the probability that the RSH distributes the first $s-1$ objects in such a nice way that the critical volume is at most w_s. Interestingly, this is essentially the same idea as for the classical PTAS for the partition problem presented by Graham [13]. Even if the RSH does not know of this algorithmic idea, it is able to behave accordingly by chance.

Theorem 3. *Let $\varepsilon \geq 4/n$. With probability at least $2^{-(\mathrm{e}\log \mathrm{e}+\mathrm{e})\lceil 2/\varepsilon\rceil \ln(4/\varepsilon)-\lceil 2/\varepsilon\rceil}$, the (1+1) EA on any instance for the partition problem creates a $(1+\varepsilon)$-approximate solution in $\lceil \mathrm{e}n \ln(4/\varepsilon))\rceil$ steps. The same holds for RLS with $\lceil n\ln(4/\varepsilon))\rceil$ steps and a probability of even at least $2^{-(\log \mathrm{e}+1)\lceil 2/\varepsilon\rceil \ln(4/\varepsilon)-\lceil 2/\varepsilon\rceil}$.*

Proof. Let $s := \lceil 2/\varepsilon\rceil \leq n/2+1$. Since $w_1 \geq \cdots \geq w_n$, it holds that $w_i \leq \varepsilon w/2$ for $i \geq s$. If $w_1 + \cdots + w_{s-1} \leq w/2$, the critical volume w. r. t. $\ell := w/2$ is always bounded above by w_s and, therefore, by $\varepsilon w/2$. Therefore, in this case, the theorem follows for $\delta := \varepsilon/2$ and $\gamma := 2$ by Lemma 2.

In the following, we assume $w_1 + \cdots + w_{s-1} > w/2$. Consider all partitions of only the first $s-1$ objects. Let ℓ^* be the minimum volume of the fuller bin over all these partitions and $\ell := \max\{w/2, \ell^*\}$. Then with a probability at least 2^{-s+2}, in the beginning, neither bin receives a contribution of more than ℓ by these objects. As long as the property remains valid, we can be sure that the critical volume w. r. t. ℓ is at most $w_s \leq \varepsilon w/2$, and we can apply the arguments from the first paragraph. The probability that in a phase of $t := \lceil \mathrm{e}n \ln(4/\varepsilon)\rceil$ steps, it never happens that at least one of the first $s-1$ bits flips is bounded below by

$$\left(1-\frac{s-1}{n}\right)^{\mathrm{e}n(\ln(4/\varepsilon))+1} \geq \mathrm{e}^{-\mathrm{e}(\ln(4/\varepsilon))(s-1)}\left(1-\frac{s-1}{n}\right)^{s\ln(4/\varepsilon)},$$

which is at least $2^{-(\mathrm{e}\log \mathrm{e}+\mathrm{e})\lceil 2/\varepsilon\rceil \ln(4/\varepsilon)}$ since $s-1 \leq n/2$. Under the mentioned conditions, by Lemma 2 for $\delta := \varepsilon/2$ and $\gamma := 2$, the (1+1) EA reaches a $(1+\varepsilon)$-approximation within t steps with probability at least $1/2$. Altogether, the desired approximation is reached within t steps with probability at least

$$\frac{1}{2}\cdot 2^{-\lceil 2/\varepsilon\rceil+2}\cdot 2^{-(\mathrm{e}\log \mathrm{e}+\mathrm{e})\lceil 2/\varepsilon\rceil \ln(4/\varepsilon)} \geq 2^{-(\mathrm{e}\log \mathrm{e}+\mathrm{e})\lceil 2/\varepsilon\rceil \ln(4/\varepsilon)-\lceil 2/\varepsilon\rceil}.$$

The statement for RLS follows by redefining $t := \lceil n\ln(4/\varepsilon)\rceil$. □

Theorem 3 allows us to design a PRAS (polynomial-time randomized approximation scheme, see [14]) for the partition problem using multistart variants of the considered RSH. If $\ell(n)$ is a lower bound on the probability that a single run of the RSH achieves the desired approximation in $O(n \ln(1/\varepsilon))$ steps then this holds for at least one out of $\lceil 2/\ell(n) \rceil$ parallel runs with a probability of at least $1 - \mathrm{e}^{-2} > 3/4$. According to the lower bounds $\ell(n)$ given in Theorem 3, the computational effort $c(n)$ incurred by the parallel runs is bounded above by $O(n \ln(1/\varepsilon)) \cdot 2^{(\mathrm{e} \log \mathrm{e} + \mathrm{e}) \lceil 2/\varepsilon \rceil \ln(4/\varepsilon) + O(1/\varepsilon)}$. For $\varepsilon > 0$ a constant, $c(n) = O(n)$ holds, and $c(n)$ is still a polynomial for any $\varepsilon = \Omega(\log \log n / \log n)$. This is the first example where it could be shown that an RSH serves as a PRAS for an NP-hard optimization problem. Before, a characterization of an EA as a PRAS was only known for the maximum matching problem [9].

4 Average-Case Analyses

A probabilistic analysis of RSHs on random inputs must take into account two sources of randomness. Since this constitutes one of the first attempts in this respect, we concentrate on two fairly simple and well-known distributions. First, we assume the volumes w_i to be independent random variables drawn uniformly from the interval $[0, 1]$. This is called the *uniform-distribution model.* Second, we rather consider exponentially distributed random variables with parameter 1, which is called the *exponential-distribution model.*

In the last two decades, some average-case analyses of deterministic heuristics for the partition problem have been performed. The first such analyses studied the LPT rule, a greedy algorithm sorting the volumes decreasingly and putting each object from the resulting sequence into the currently emptier bin. Extending a result that stated convergence in expectation, Frenk and Rinnooy Kan [11] were able to prove that the LPT rule converges to optimality at a speed of $O(\log n/n)$ almost surely in several input models, including the uniform-distribution and exponential-distribution model. Further results on average-case analyses of more elaborate deterministic heuristics are contained in [12].

In our models, the optimum f-value is random. Therefore, for a current search point, we now consider the so-called discrepancy measure rather than an approximation ratio. The discrepancy denotes the absolute difference of the total volumes in the bins. It is easy to see that the initial discrepancy in both models is $\Omega(\sqrt{n})$ with constant probability. We start with a simple upper bound on the discrepancy after polynomially many steps in the uniform-distribution model.

Lemma 3. *The discrepancy of the (1+1) EA (RLS) in the uniform-distribution model is bounded above by* 1 *after an expected number of* $O(n^2)$ $(O(n \log n))$ *steps. Moreover, for any constant* $c \geq 1$*, it is bounded above by* 1 *with probability at least* $1 - O(1/n^c)$ *after* $O(n^2 \log n)$ $(O(n \log n))$ *steps.*

Proof. Recall the argumentation given after Definition 1. Hence, if the discrepancy is greater than 1, local steps can improve the f-value by the volume moved or lead to a discrepancy of less than 1. By a fitness-level argument like in the proof of Lemma 1, we obtain the $O(n^2)$ bound for the (1+1) EA. This holds for any random instance. Hence, by Markov's inequality and repeating phases, the discrepancy is at most 1 with probability $1 - O(1/n^c)$ after $O(n^2 \log n)$ steps. The statements for RLS follow immediately by the Coupon Collector's Theorem. □

The foregoing upper bound on the discrepancy was easy to obtain; however, for the (1+1) EA, we can show that with a high probability, the discrepancy provably becomes much lower than 1 in a polynomial number of steps. The reason is as follows. All preceding proofs considered only local steps; however, the (1+1) EA is able to leave local optima by flipping several bits in a step.

The following two theorems will use the following simple properties of order statistics (e. g., [15]). Let $X_{(1)} \geq \cdots \geq X_{(n)}$ be the order statistics of the volumes in the uniform-distribution model. Then for $1 \leq i \leq n-1$ and $0 < t < 1$, $\mathrm{Prob}(X_{(i)} - X_{(i+1)} \geq t) = \mathrm{Prob}(X_{(n)} \geq t) = (1-t)^n$. In the exponential-distribution model, there is a sequence of independent, parameter-1 exponentially distributed random variables $Y_1, \ldots, Y_n$ s. t. $X_{(i)} = \sum_{j=i}^{n} \frac{Y_j}{j}$ for $1 \leq i \leq n$.

Theorem 4. *Let $c \geq 1$ be an arbitrary constant. After $O(n^{c+4} \log n)$ steps, the discrepancy of the (1+1) EA in the uniform-distribution model is bounded above by $O(\log n/n)$ with probability at least $1 - O(1/n^c)$. Moreover, the expected discrepancy after $O(n^5 \log n)$ steps is also bounded by $O(\log n/n)$.*

Proof. By Lemma 3, the discrepancy is at most 1 after $O(n^2 \log n)$ steps with probability at least $1 - O(1/n^2)$. Since the discrepancy is always bounded by n, the failure probability contributes only an $O(1/n)$-term to the expected discrepancy after $O(n^5 \log n)$ steps. From now on, we consider the time after the first step where the discrepancy is at most 1 and concentrate on steps flipping two bits. If an accepted step moves an object of volume w' from the fuller to the emptier bin and one of volume $w'' < w'$ the other way round, the discrepancy may be decreased by $2(w' - w'')$. We look for combinations s. t. $w' - w''$ is small.

Let $X_{(1)} \geq \cdots \geq X_{(n)}$ be the order statistics of the random volumes. If for the current search point, there is some i s. t. $X_{(i)}$ is the order statistic of an object in the fuller and $X_{(i+1)}$ is in the emptier bin then a step exchanging $X_{(i)}$ and $X_{(i+1)}$ may decrease the discrepancy by $2(X_{(i)} - X_{(i+1)})$. If no such i exists, all objects in the emptier bin are larger than any object in the fuller bin. In this case, $X_{(n)}$ can be moved into the emptier bin, possibly decreasing the discrepancy by $2X_{(n)}$. Hence, we need upper bounds on $X_{(i)} - X_{(i+1)}$ and $X_{(n)}$.

Let $t^* := (c+1)(\ln n)/n$, i. e., $t^* = O(\log n/n)$ since c is a constant. We obtain $(1-t^*)^n \leq n^{-c-1}$. By the above-mentioned statement, this implies that with probability $1 - O(1/n^c)$, $X_{(i)} - X_{(i+1)} \leq t^*$ holds for all i and $\mathrm{Prob}(X_{(n)} \geq t^*) = O(1/n^{c+1})$. Now assume $X_{(i)} - X_{(i+1)} \leq t^*$ for all i and $X_{(n)} \leq t^*$. If this does not hold, we bound the expected discrepancy after $O(n^{c+4} \log n)$ steps by 1, yielding a term of $O(1/n^c) = O(1/n)$ in the total expected discrepancy. By the argumentation given after Definition 1, there is always a step flipping at most 2 bits that decreases the discrepancy as long as the discrepancy is greater than t^*.

It remains to estimate the time to decrease the discrepancy. Therefore, we need lower bounds on $X_{(i)} - X_{(i+1)}$ and X_n. Let $\ell^* := 1/n^{c+2}$. We obtain $\mathrm{Prob}(X_{(i)} - X_{(i+1)} \geq \ell^*) \geq \mathrm{e}^{-2/n^{c+1}} \geq 1 - 2/n^{c+1}$. Hence, with probability $1 - O(1/n^c)$, $X_{(i)} - X_{(i+1)} \geq \ell^*$ for all i. Moreover, $X_{(n)} \geq \ell^*$ with probability $1 - O(1/n^{c+1})$. We assume these lower bounds to hold, introducing a failure probability of only $O(1/n^c)$, whose contribution to the expected discrepancy is negligible as above. A step flipping 1 resp. 2 specific bits has probability at least $n^{-2}(1-1/n)^{n-2} \geq 1/(\mathrm{e}n^2)$. Hence, the discrepancy is decreased by at least ℓ^* or drops below t^* with probability $\Omega(1/n^2)$ in each step. The expected time until the discrepancy becomes at most t^* is, therefore, bounded above by $O(\ell^* n^2) = O(n^{c+4})$, and, by repeating phases, the time is at most $O(n^{c+4}\log n)$ with probability $1-O(1/n^c)$. The sum of all failure probabilities is $O(1/n^c)$. □

Theorem 5. *Let $c \geq 1$ be an arbitrary constant. With probability $1-O(1/n^c)$, the discrepancy of the (1+1) EA in the exponential-distribution model is bounded above by $O(\log n)$ after $O(n^2 \log n)$ steps and by $O(\log n/n)$ after $O(n^{c+4}\log^2 n)$ steps. Moreover, the expected discrepancy is $O(\log n)$ after $O(n^2\log n)$ steps and it is $O(\log n/n)$ after $O(n^6 \log^2 n)$ steps.*

Proof. The expected value of the initial discrepancy is bounded above by n since each object has an expected volume of 1. In the following, all failure probabilities can be bounded by $O(1/n^2)$. In case of a failure, we will tacitly bound the failure's contribution to the expected discrepancy after $O(n^2\log n)$ resp. $O(n^6\log^2 n)$ steps by $O(1/n)$. Next, we will show that with probability $1-O(1/n^c)$, the critical volume w. r. t. $\ell := w/2$ is always $O(\log n)$. Together with Lemma 1, this claim implies the theorem for the situation after $O(n^2 \log n)$ steps.

To show the claim, like in the proof of Theorem 4, we consider the order statistics $X_{(1)} \geq \cdots \geq X_{(n)}$ of the random volumes. Our goal is to show that with high probability, $X_{(1)} + \cdots + X_{(k)} \leq w/2$ holds for $k := \lceil \delta n \rceil$ and some constant $\delta > 0$. Afterwards, we will prove that $X_{(k)} = O(\log n)$ with high probability.

Each object has a volume of at least 1 with probability $\mathrm{e}^{-1} > 1/3$. By Chernoff bounds, $w \geq n/3$ with probability $1-2^{-\Omega(n)}$. To bound $X_{(1)}+\cdots+X_{(k)}$, we use the above-mentioned identity $X_{(i)} = \sum_{j=i}^{n} Y_j/j$. Hence,

$$
\begin{aligned}
X_{(1)} + \cdots + X_{(k)} &= Y_1 + 2\cdot\frac{Y_2}{2} + \cdots + k\cdot\frac{Y_k}{k} + k\sum_{i=k+1}^{n}\frac{Y_i}{i} \\
&\leq \sum_{j=1}^{k} Y_j + \sum_{i=1}^{\lceil n/k\rceil}\frac{1}{i}\sum_{j=ik+1}^{(i+1)k} Y_j,
\end{aligned}
$$

where $Y_j := 0$ for $j > n$. Essentially, we are confronted with $\lceil n/k \rceil$ sums of k exponentially distributed random variables each. A simple calculation (deferred to the last paragraph of this proof) yields that a single sum is bounded above by $2k$ with probability $1-2^{-\Omega(k)}$, which is at least $1-2^{-\Omega(n)}$ for the values of k considered. Since we consider at most n sums, this statement also holds for all sums together. Hence, with probability $1-2^{-\Omega(n)}$, the considered expression is bounded above by

$$
2\lceil \delta n\rceil + \sum_{i=1}^{1/\delta}\frac{2\lceil \delta n\rceil}{i} \leq 2(\delta n + 1)\ln(1/\delta + 2),
$$

which is strictly less than $n/6$ for $\delta \leq 1/50$ and n large enough. Together with the above lower bound on w, this implies that with probability $1-2^{-\Omega(n)}$, the critical volume is always bounded above by the $\lceil n/50\rceil$-th largest volume.

How large is $X_{(\lceil n/50\rceil)}$? Since with probability at least $1 - n\mathrm{e}^{-(c+1)\ln n} \geq 1 - n^{-c}$, all random variables Y_j are bounded above by $(c+1)\ln n$, it follows that with at least the same probability, we have

$$X_{\lceil n/50\rceil} = \sum_{j=\lceil n/50\rceil}^{n} \frac{Y_j}{j} \leq (c+1)(\ln n)((\ln n) + 1 - \ln(n/49))$$

(for n large enough), which equals $(c+1)(\ln(49)+1)(\ln n) = O(\log n)$. The sum of all failure probabilities is $O(1/n^c)$, bounding the critical volume as desired.

We still have to show the theorem for the case of $O(n^{c+4}\log^2 n)$ steps. Now we assume that the discrepancy has been decreased to $O(\log n)$ and use the same idea as in the proof of Theorem 4 by investigating steps swapping $X_{(i)}$ and $X_{(i+1)}$ or moving $X_{(n)}$. Above, we have shown that with probability $1 - O(1/n^c)$, the smallest object in the fuller bin is always at most $X_{(k)}$ for some $k \geq n/50$. Since $X_{(k)} - X_{(k+1)} = Y_k/k$, we obtain $X_{(k)} - X_{(k+1)} \leq 50Y_k/n$ with the mentioned probability. Moreover, it was shown that $Y_j \leq (c+1)\ln n$ for all j with at least the same probability. Altogether, $X_{(k)} - X_{(k+1)} \leq 50(c+1)(\ln n/n) =: t^*$ with probability $1 - O(1/n^c)$. Since $X_{(n)} = Y_n/n$, $\mathrm{Prob}(X_{(n)} \leq t^*)$ with probability $1 - O(1/n^c)$, too. In the following, we assume these upper bounds to hold. This implies that as long as the discrepancy is greater than t^*, there is a step flipping at most 2 bits and decreasing the discrepancy.

It remains to establish lower bounds on $X_{(k)} - X_{(k+1)}$ and $X_{(n)}$. We know that $X_{(k)} - X_{(k+1)} \geq Y_k/n$ and obtain $\mathrm{Prob}(X_{(k)} - X_{(k+1)} \geq 1/n^{c+2}) \geq \mathrm{e}^{-1/n^{c+1}} \geq 1 - 1/n^{c+1}$ for any fixed k and $\mathrm{Prob}(X_{(n)} \geq 1/n^{c+2}) \geq 1 - 1/n^{c+1}$. All events together occur with probability $1 - O(1/n^c)$. By the same arguments as in the proof of Theorem 4, the expected time until the discrepancy becomes at most t^* is $O(n^{c+4}\log n)$, and the time is bounded by $O(n^{c+4}\log^2 n)$ with probability $1 - O(1/n^c)$. The sum of all failure probabilities is $O(1/n^c)$. This will imply the theorem.

We still have to show the following claim. The sum S_k of k exponentially distributed random variables with parameter 1 is at most $2k$ with probability $1 - 2^{-\Omega(k)}$. Observe that S_k follows a gamma distribution, i. e.,

$$\mathrm{Prob}(S_k \geq 2k) = \mathrm{e}^{-2k}\left(1 + \frac{2k}{1!} + \cdots + \frac{(2k)^{k-1}}{(k-1)!}\right) \leq \frac{k\mathrm{e}^{-2k}(2k)^{k-1}}{(k-1)!}.$$

By Stirling's formula, the last expression is bounded above by

$$\frac{\mathrm{e}^{-2k+(k-1)} \cdot 2^{k-1} \cdot k \cdot k^{k-1}}{(k-1)^{k-1}} = \mathrm{e}^{-2k+(k-1)} \cdot 2^{k-1} \cdot k \cdot \left(1 - \frac{1}{k}\right)^{-(k-1)} = 2^{-\Omega(k)}.$$

This proves the claim and, therefore, the theorem. □

Theorem 4 and Theorem 5 imply that in both models, the solution of the (1+1) EA after a polynomial number of steps converges to optimality *in expectation.* Moreover, the asymptotic discrepancy after a polynomial number of steps is at most $O(\log n/n)$, i. e., convergent to 0, with probability $1 - O(1/n^c)$, i. e., convergent to 1 polynomially fast. This is almost as strong as the above-mentioned result for the LPT rule.

Conclusions

In this paper, we have presented a probabilistic analysis for randomized search heuristics on the optimization variant of the PARTITION problem. In the worst case, both the (1+1) EA and RLS with constant probability need exponential time to create solutions being better than $(4/3-\varepsilon)$-approximate; however, parallel runs of the heuristics lead to a PRAS. An average-case analysis with respect to two input distributions shows that the (1+1) EA, inspected after a polynomial number of steps, creates solutions that are in some sense convergent to optimality. By this average-case analysis, we have made a step towards a theoretical justification of the efficiency of randomized search heuristics for practically relevant problem instances.

Acknowledgements. The author thanks Ingo Wegener for helpful discussions.

References

1. Beier, R., Vöcking, B.: Random knapsack in expected polynomial time. In: Proc. of 35th STOC. (2003) 232–241
2. Spielman, D.A., Teng, S.H.: Smoothed analysis: Why the simplex algorithm usually takes polynomial time. In: Proc. of 33rd STOC. (2001) 296–305
3. Hochbaum, D.S., ed.: Approximation Algorithms for NP-Hard Problems. Wadsworth Publishing Company (1997)
4. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220** (1983) 1087–1092
5. Baeck, T., Fogel, D.B., Michalewicz, Z., eds.: Handbook of Evolutionary Computation. Institute of Physics Publishing (1997)
6. Droste, S., Jansen, T., Wegener, I.: On the analysis of the (1+1) evolutionary algorithm. Theoretical Computer Science **276** (2002) 51–81
7. Wegener, I., Witt, C.: On the optimization of monotone polynomials by simple randomized search heuristics. Combinatorics, Probability and Computing (2005) in press.
8. Witt, C.: An analysis of the (μ+1) EA on simple pseudo-boolean functions. In: Proc. of GECCO 2004, LNCS 3102 (2004) 761–773
9. Giel, O., Wegener, I.: Evolutionary algorithms and the maximum matching problem. In: Proc. of 20th STACS, LNCS 2607 (2003) 415–426
10. Neumann, F., Wegener, I.: Randomized local search, evolutionary algorithms, and the minimum spanning tree problem. In: Proc. of GECCO 2004, LNCS 3102 (2004) 713–724
11. Frenk, J.B.G., Rinnooy Kan, A.H.G.: The rate of convergence to optimality of the LPT rule. Discrete Applied Mathematics **14** (1986) 187–197
12. Coffman, Jr., E.G., Whitt, W.: Recent asymptotic results in the probabilistic analysis of schedule makespans. In Chrètienne, P., Coffman, Jr., E.G., Lenstra, J.K., Liu, Z., eds.: Scheduling Theory and its Applications. Wiley (1995) 15–31
13. Graham, R.L.: Bounds on multiprocessing timing anomalies. SIAM Journal on Applied Mathematics **17** (1969) 263–269
14. Motwani, R., Raghavan, P.: Randomized Algorithms. Cambridge University Press (1995)
15. David, H.A., Nagaraja, H.N.: Order Statistics. 3rd edn. Wiley (2003)