On Algorithm-Dependent Boundary Case Identification for Problem Classes*

Chao Qian, Yang Yu, and Zhi-Hua Zhou

National Key Laboratory for Novel Software Technology,
Nanjing University, Nanjing 210046, China,
{qianc,yuy,zhouzh}@lamda.nju.edu.cn

Abstract. Running time analysis of metaheuristic search algorithms has attracted a lot of attention. When studying a metaheuristic algorithm over a problem class, a natural question is what are the easiest and the hardest cases of the problem class. The answer can be helpful for simplifying the analysis of an algorithm over a problem class as well as understanding the strength and weakness of an algorithm. This algorithm-dependent boundary case identification problem is investigated in this paper. We derive a general theorem for the identification, and apply it to a case that the (1+1)-EA with mutation probability less than 0.5 is used over the problem class of pseudo-Boolean functions with a unique global optimum.

1 Introduction

Metaheuristic search algorithms such as simulated annealing (SA) [11], particle swarm optimization (PSO) [10], evolutionary algorithms (EA) [2], etc., have been widely and successfully applied to real-world optimization problems. An advantage of metaheuristic algorithms is their problem independence, i.e., they can be applied to a very large range of optimization problems. A natural theoretical question is, therefore, how well the metaheuristic algorithms perform on classes of problems.

A commonly used quality measure of a metaheuristic algorithm is its expected running time, i.e., the expected number of steps that it takes to find an optimum. Several approaches, e.g., drift analysis [7] and convergence-based approach [17], have been developed for running time analysis of metaheuristic algorithms. The running time of several metaheuristic algorithms has been studied on some simple pseudo-Boolean problems, e.g., [4, 7, 5], and later, on some combinatorial optimization problems, e.g., [1, 12]. In most of these studies, the analysis was over restricted problem classes where the problem cases have similar structures. While, for large problem classes, a large variety of structures of problem cases can obstruct the analysis.

One possible way to simplify the analysis over a problem class is to characterize the class by its easiest and hardest cases, and then analyze on these boundary cases. By the well-known no free lunch theorem [14], we know that no single problem is intrinsically harder than another, until an algorithm is involved into the consideration. Therefore,

* This research was supported by the National Science Foundation of China (60903103, 61105043)

this paper studies the algorithm-dependent boundary case identification problem. Given a metaheuristic algorithm, the identification of the boundary cases of a problem class can not only help to study the performance of the algorithm over the problem class, but also provide concrete cases to reveal the strength and weakness of the metaheuristic algorithm.

For this purpose, we derive a general theorem for algorithm-dependent boundary case identification, which gives a sufficient condition for identifying the easiest and hardest cases of a problem class for an algorithm. We then prove that in the pseudo-Boolean function class with a unique global optimum, the OneMax and the Trap problem are the easiest and the hardest case for the (1+1)-EA with mutation probability less than 0.5, respectively.

There are a few previous studies concerning problem classes. The running time of EAs on linear pseudo-Boolean function class, which is a relatively small problem class, was analyzed in [5, 8, 9, 3]. Yu and Zhou [17] provided a general idea on why EAs can fail over a complex problem class; Fournier and Teytaud [6] provided a general lower bound for the performance of EAs over problem classes with VC-dimension measured complexity. However, these studies did not concern the boundary problem cases. Recently, Doerr et al. [3] used the lower bound of running time of the (1+1)-EA with mutation probability $\frac{1}{n}$ on the OneMax problem as that on the pseudo-Boolean function class with a unique global optimum by proving that the OneMax problem is the easiest case for the (1+1)-EA in this class, comparing to which this paper derives a more general result for the easiest case as well as the hardest case. Note that, the easiest case derived in this paper has also been proved by Witt [13], but we give a different and more compact proof.

The rest of this paper is organized as follows. Section 2 introduces some preliminaries. Section 3 presents the main theorem, which is then used to identify the boundary cases in the pseudo-Boolean function class for the (1+1)-EA in Section 4. Section 5 concludes.

2 Preliminaries

Most metaheuristic search algorithms generate solutions only based on their maintained solutions, but not the historical ones, therefore, they can be modeled and analyzed as Markov chains, e.g., [7, 17]. In this paper, we only consider the algorithms that can be modeled by Markov chains. A Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ modeling the metaheuristic algorithm is constructed by taking the algorithm's state space $\mathcal{X}$ as the chain's state space, i.e. $\xi_t \in \mathcal{X}$. Let $\mathcal{X}^* \subset \mathcal{X}$ denote the set of all optimal states. The goal of the algorithm is to reach $\mathcal{X}^*$ from an arbitrary initial state. Thus, the process of an algorithm seeking $\mathcal{X}^*$ can be analyzed by studying the corresponding Markov chain.

A Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ is a random process, where, for all $t \geq 0$, ξ_t is defined in the state space $\mathcal{X}$ and ξ_{t+1} depends only on ξ_t. Let $\mathcal{X}^* \subset \mathcal{X}$ be the target space. A Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ is said to be absorbing, if $\forall t \geq 0 : P(\xi_{t+1} \in \mathcal{X}^* | \xi_t \in \mathcal{X}^*) = 1$.

Given a Markov chain $\{\xi_t\}_{t=0}^{+\infty}$, and $\xi_{\tilde{t}} = x$ for arbitrary $\tilde{t} \geq 0$, we define $\tau_{\tilde{t}}$ as a random variable such that $\tau_{\tilde{t}} = \min\{t | \xi_{\tilde{t}+t} \in \mathcal{X}^*, t \geq 0\}$. That is, $\tau_{\tilde{t}}$ is the number of steps needed to reach the target space for the first time from $\tilde{t}$. The mathematical expectation of $\tau_{\tilde{t}}$, $\mathbb{E}[\![\tau_{\tilde{t}}|\xi_{\tilde{t}}=x]\!]=\sum_{i=0}^{\infty} iP(\tau_{\tilde{t}}=i)$, is called the *conditional first hitting time* (CFHT) of the Markov chain from $\tilde{t}$ and $\xi_{\tilde{t}}=x$. If $\xi_{\tilde{t}}$ is drawn from a distribution $\pi_{\tilde{t}}$, the expectation of the CFHT over $\pi_{\tilde{t}}$, $\mathbb{E}[\![\tau_{\tilde{t}}|\xi_{\tilde{t}}\sim\pi_{\tilde{t}}]\!]=\sum_{x\in\mathcal{X}}\pi_{\tilde{t}}(x)\mathbb{E}[\![\tau_{\tilde{t}}|\xi_{\tilde{t}}=x]\!]$, is called the *distribution-conditional first hitting time* (DCFHT) of the Markov chain from $\tilde{t}$ and $\xi_{\tilde{t}}\sim\pi_{\tilde{t}}$. If $\tilde{t}=0$, $\mathbb{E}[\![\tau_0|\xi_0\sim\pi_0]\!]$ is also called the expected running time of the corresponding algorithm.

Switch analysis is a recently proposed approach [16] that compares two Markov chains for their first hitting time. By modeling EAs as Markov chains, it has been used to derive running time bounds of EAs [16] and investigate if one EA runs faster than another EA for a problem [15, 16].

Theorem 1 (Switch Analysis [16]). *Given two absorbing Markov chains $\{\xi_t\}_{t=0}^{+\infty}(\xi_t\in\mathcal{X})$ and $\{\xi'_t\}_{t=0}^{+\infty}(\xi'_t\in\mathcal{Y})$, let $\mathcal{X}^*$ and $\mathcal{Y}^*$ denote the optimal state space of ξ_t and ξ'_t, respectively, let τ and τ' denote the hitting events of ξ_t and ξ'_t, respectively, let π_t denote the distribution of ξ_t. Let $\{\rho_t\}_{t=0}^{+\infty}$ be a series of numbers whose sum converges to ρ. If there exists a mapping $\phi:\mathcal{X}\to\mathcal{Y}$, $\phi(x)\in\mathcal{Y}^*$ if and only if $x\in\mathcal{X}^*$; and it satisfies that $\mathbb{E}[\![\tau_0|\xi_0\sim\pi_0]\!]$ is finite, and for all $t\geq 0$,*

$$\begin{aligned}&\sum\nolimits_{x\in\mathcal{X},y\in\mathcal{Y}}\pi_t(x)P(\xi_{t+1}\in\phi^{-1}(y)|\xi_t=x)\mathbb{E}[\![\tau'_{t+1}|\xi'_{t+1}=y]\!]\\&\leq(\geq)\sum\nolimits_{y_1,y_2\in\mathcal{Y}}\pi'_t(y_1)P(\xi'_{t+1}=y_2|\xi'_t=y_1)\mathbb{E}[\![\tau'_{t+1}|\xi'_{t+1}=y_2]\!]+\rho_t,\end{aligned}\tag{1}$$

where $\phi^{-1}(y)=\{x\in\mathcal{X}|\phi(x)=y\}$ and $\pi'_t(y)=\pi_t(\phi^{-1}(y))$, it will hold that

$$\mathbb{E}[\![\tau_0|\xi_0\sim\pi_0]\!]\leq(\geq)\mathbb{E}[\![\tau'_0|\xi'_0\sim\pi'_0]\!]+\rho.$$

3 A Theorem for Boundary Problem Identification

We assume that the studied problem class is homogeneous, as in Definition 1, which means that all the problem cases in the class have the same solution space and the same optimal solutions when fixing the problem dimensionality. At the first glance, the requirement of the same optimal solutions is a strong restriction. However, since most metaheuristic algorithms do not rely on the meaning of the solution, in the analysis we can commonly switch the optimal solutions. For example, the solution 10011 in a binary space can be shifted as 11111 if we switch the meaning of 1 and 0 for the 2nd and the 3rd bits.

Definition 1 (Homogeneous Problem Class). *A problem class is homogeneous if, for each problem dimensionality, all the problem cases have the same solution space and the same optimal solutions.*

To characterize the algorithm-dependent structure of the problem cases, we partition the state space according to the CFHT of a given algorithm, as in Definition 2. We also define a jumping probability in Definition 3.

Definition 2 (CFHT-Partition). *For a Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ with state space $\mathcal{X}$, the CFHT-Partition at time t is a partition of $\mathcal{X}$ into non-empty spaces $\{\mathcal{X}_0^t,\mathcal{X}_1^t,\ldots,\mathcal{X}_m^t\}$ such that $\forall x,y\in\mathcal{X}_i^t$, $\mathbb{E}[\![\tau_{t+1}|\xi_{t+1}=x]\!]=\mathbb{E}[\![\tau_{t+1}|\xi_{t+1}=y]\!]$ and $\mathbb{E}[\![\tau_{t+1}|\xi_{t+1}\in\mathcal{X}_m^t]\!]>\ldots>\mathbb{E}[\![\tau_{t+1}|\xi_{t+1}\in\mathcal{X}_1^t]\!]>\mathbb{E}[\![\tau_{t+1}|\xi_{t+1}\in\mathcal{X}_0^t]\!]=0$.*

Definition 3. *For a Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ with the state space $\mathcal{X}$, $P_\xi^t(x,\mathcal{X}')$ is the probability of jumping from state x to state space $\mathcal{X}' \subseteq \mathcal{X}$ in one step at time t.*

We then derive Theorem 2, which is a general sufficient condition for identifying the easiest and the hardest problem cases. Note that, the easiest (hardest) problem case of a problem class for an algorithm means that the expected running time of the algorithm on the problem case is the smallest (largest).

Theorem 2. *Given a homogeneous problem class $\mathcal{F}$ and an algorithm $\mathcal{A}$, with dimensionality n, let $\{\xi'_t\}_{t=0}^{+\infty}$ model $\mathcal{A}$ running on a problem $f^* \in \mathcal{F}_n$, of which $\{\mathcal{X}_0^t, \mathcal{X}_1^t, \ldots, \mathcal{X}_m^t\}$ is the CFHT-Partition at time t. If for all problem $f \in \mathcal{F}_n - \{f^*\}$, for all $t \geq 0$, and for all $x \in \mathcal{X} - \mathcal{X}_0^t$, denoting $\{\xi_t\}_{t=0}^{+\infty}$ as the chain modeling $\mathcal{A}$ running on f, there exists an integer $k \in [0,m]$,*

$$\forall j \leq k, P_\xi^t(x,\mathcal{X}_j^t) \leq (\geq) P_{\xi'}^t(x,\mathcal{X}_j^t), \quad \forall j > k, P_\xi^t(x,\mathcal{X}_j^t) \geq (\leq) P_{\xi'}^t(x,\mathcal{X}_j^t), \tag{2}$$

then f^* is the easiest (hardest) case in $\mathcal{F}_n$ for the algorithm $\mathcal{A}$.

Proof. We use the switch analysis approach to show the easiest problem case identification of this theorem by proving that the expected running time of the algorithm $\mathcal{A}$ on the problem f^* is at most as large as that on any other problem. The hardest case identification can be proved similarly. Note that both Markov chains $\{\xi_t\}_{t=0}^{+\infty}$ and $\{\xi'_t\}_{t=0}^{+\infty}$ can be transformed to be absorbing by letting them always stay at the optimal state once an optimal state has been found, and this transformation does not affect their running time by the definition of CFHT/DCFHT.

The two chains $\{\xi_t\}_{t=0}^{+\infty}$ and $\{\xi'_t\}_{t=0}^{+\infty}$ have the same state space $\mathcal{X}$ and the same optimal state space $\mathcal{X}^*$, since the studied problem class is homogeneous. For the clearness of the proof, we denote the state space and the optimal space of $\{\xi'_t\}_{t=0}^{+\infty}$ by $\mathcal{Y}$ and $\mathcal{Y}^*$, respectively. Obviously, $\mathcal{Y} = \mathcal{X}$ and $\mathcal{Y}^* = \mathcal{X}^*$. Then, we construct the mapping $\phi : \mathcal{X} \to \mathcal{Y}$ as that $\forall x \in \mathcal{X} : \phi(x) = x$. It is obvious that $\phi(x) \in \mathcal{Y}^*$ iff $x \in \mathcal{X}^*$.

Then, we investigate Eq. 1 in switch analysis. For an optimal state $x \in \mathcal{X}^* = \mathcal{X}_0^t$, since $\phi(x) = x$ and both Markov chains are absorbing, we have

$$\begin{aligned} &\sum\nolimits_{y\in\mathcal{Y}} P(\xi_{t+1} \in \phi^{-1}(y) \mid \xi_t = x)\mathbb{E}[\![\tau'_{t+1} \mid \xi'_{t+1} = y]\!] \\ &= \sum\nolimits_{y\in\mathcal{Y}} P(\xi'_{t+1} = y \mid \xi'_t = \phi(x))\mathbb{E}[\![\tau'_{t+1} \mid \xi'_{t+1} = y]\!] = 0. \end{aligned} \tag{3}$$

For a non-optimal state $x \in \mathcal{X}_i^t$ $(i \geq 1)$, since $\phi(x) = x$, we have

$$\sum_{y\in\mathcal{Y}} P(\xi'_{t+1} = y|\xi'_t = \phi(x))\mathbb{E}[\![\tau'_{t+1}|\xi'_{t+1} = y]\!] = \sum_{j=0}^{m} P_{\xi'}^t(x,\mathcal{X}_j^t)\mathbb{E}[\![\tau'_{t+1}|\xi'_{t+1} \in \mathcal{X}_j^t]\!];$$

$$\sum_{y\in\mathcal{Y}} P(\xi_{t+1} \in \phi^{-1}(y)|\xi_t = x)\mathbb{E}[\![\tau'_{t+1}|\xi'_{t+1} = y]\!] = \sum_{j=0}^{m} P_{\xi}^t(x,\mathcal{X}_j^t)\mathbb{E}[\![\tau'_{t+1}|\xi'_{t+1} \in \mathcal{X}_j^t]\!].$$

By comparing the above two equalities, since the condition Eq. 2 holds, and furthermore, $\mathbb{E}[\![\tau'_{t+1}|\xi'_{t+1} \in \mathcal{X}_j^t]\!]$ increases with j, we have

$$\sum\nolimits_{y\in\mathcal{Y}} P(\xi_{t+1} \in \phi^{-1}(y)|\xi_t = x)\mathbb{E}[\![\tau'_{t+1}|\xi'_{t+1} = y]\!] \tag{4}$$

$$\geq \sum\nolimits_{y \in \mathcal{Y}} P(\xi'_{t+1} = y | \xi'_t = \phi(x)) \mathbb{E}[\![\tau'_{t+1} | \xi'_{t+1} = y]\!].$$

Then, by combining Eq. 3 with Eq. 4, we have for all $t \geq 0$,

$$\sum\nolimits_{x \in \mathcal{X}, y \in \mathcal{Y}} \pi_t(x) P(\xi_{t+1} \in \phi^{-1}(y) | \xi_t = x) \mathbb{E}[\![\tau'_{t+1} | \xi'_{t+1} = y]\!]$$
$$\geq \sum\nolimits_{x \in \mathcal{X}, y \in \mathcal{Y}} \pi_t(x) P(\xi'_{t+1} = y | \xi'_t = \phi(x)) \mathbb{E}[\![\tau'_{t+1} | \xi'_{t+1} = y]\!]$$
$$= \sum_{y_1, y_2 \in \mathcal{Y}} \pi'_t(y_1) P(\xi'_{t+1} = y_2 | \xi'_t = y_1) \mathbb{E}[\![\tau'_{t+1} | \xi'_{t+1} = y_2]\!]. \quad (\text{by } \pi'_t(y) = \pi_t(\phi^{-1}(y)))$$

Thus, Eq. 1 holds with $\rho_t = 0$. By switch analysis, $\mathbb{E}[\![\tau_0 | \xi_0 \sim \pi_0]\!] \geq \mathbb{E}[\![\tau'_0 | \xi'_0 \sim \pi'_0]\!]$. Since $\pi_0 = \pi'_0$, we have $\mathbb{E}[\![\tau_0 | \xi_0 \sim \pi_0]\!] \geq \mathbb{E}[\![\tau'_0 | \xi'_0 \sim \pi_0]\!]$. The two sides of this inequality are the expected running time of $\mathcal{A}$ on a problem $f \in \mathcal{F}_n - \{f^*\}$ and that on the problem f^*, respectively. This inequality holds for each $f \in \mathcal{F}_n - \{f^*\}$ and f^*. Thus, f^* is the easiest problem in $\mathcal{F}_n$ for the algorithm $\mathcal{A}$. □

4 Pseudo-Boolean Function Class and (1+1)-EA

In this section, we use the proved theorem to identify the easiest and the hardest function in the pseudo-Boolean function class with a unique global optimum for the (1+1)-EA with mutation probability less than 0.5.

The pseudo-Boolean function class in Definition 4 is a large function class which only requires the solution space to be $\{0,1\}^n$ and the objective space to be $\mathbb{R}$. The pseudo-Boolean function class with a unique global optimum is a subset of the pseudo-Boolean function class, where each function has a unique global optimum. Here, we consider maximization problems since minimizing f is equivalent to maximizing $-f$. The OneMax problem in Definition 5 is to maximize the number of 1 bits of a solution. The Trap problem in Definition 6 is to maximize the number of 0 bits of a solution except the global optimum 1^n.

Definition 4 (Pseudo-Boolean Function Class). *A function in the pseudo-Boolean function class has the form:* $f : \{0,1\}^n \to \mathbb{R}$.

Definition 5 (OneMax Problem). *OneMax Problem of size n is to find an n bits binary string x^* such that* $x^* = \arg\max_{x \in \{0,1\}^n} \{f(x) | f(x) = \sum_{i=1}^n x_i\}$.

Definition 6 (Trap Problem). *Trap Problem of size n is to find an n bits binary string x^* such that* $x^* = \arg\max_{x \in \{0,1\}^n} \{f(x) | f(x) = \sum_{i=1}^n (1 - x_i) + (n+1) \prod_{i=1}^n x_i\}$.

The (1+1)-EA [4] in Algorithm 1 is a randomized heuristic algorithm for maximizing pseudo-Boolean functions, which is often involved in theoretical analysis of EAs.

Algorithm 1 ((1+1)-EA) *Given solution length n and objective function f, (1+1)-EA consists of the following steps:*

1. $x :=$ *randomly selected from* $\{0,1\}^n$.
2. Repeat until the termination condition is met
3. $\quad x' :=$ *flip each bit of* x *with probability* p*;*
4. $\quad$ *if* $f(x') \geq f(x)$
5. $\qquad x := x'$*;*
where $p \in (0,1)$ *is the mutation probability.*

For a pseudo-Boolean function with a unique global optimum, we assume without loss of generality that the optimal solution is 1^n. This is because the (1+1)-EA treats the bits 0 and 1 symmetrically, and thus the 0 bits in an optimal solution can be interpreted as 1 bits without affecting the behavior of the (1+1)-EA. The optimization time of the (1+1)-EA for maximizing a pseudo-Boolean function is computed as the number of iterations until a global optimum has been found for the first time.

Theorem 3. *In the pseudo-Boolean function class with a unique global optimum, the OneMax and the Trap problem are the easiest and the hardest problem case for the (1+1)-EA with* $0 < p < 0.5$*, respectively.*

Before proving Theorem 3, we first prove the order of the CFHT of the (1+1)-EA on the OneMax problem in Lemma 1 as well as that on the Trap problem in Lemma 2. Since the bits of the OneMax problem are independent and their weights are same, it is not hard to see that the CFHT $\mathbb{E}[\![\tau'_t|\xi'_t = x]\!]$ of the (1+1)-EA on the OneMax problem only depends on the number of 1 bits of the solution x, i.e., $\|x\|$. Thus, we denote $\mathbb{E}(j)$ as the CFHT $\mathbb{E}[\![\tau'_t|\xi'_t = x]\!]$ with $\|x\| = n - j$. Then, it is obvious that $\mathbb{E}(0) = 0$, which implies the optimal solution.

Lemma 1. *For* $0 < p < 0.5$*, it holds that* $\mathbb{E}(0) < \mathbb{E}(1) < \mathbb{E}(2) < \ldots < \mathbb{E}(n)$.

Proof. We prove $\forall\, 0 \leq j < n : \mathbb{E}(j) < \mathbb{E}(j+1)$ inductively on j.

(a) Initialization is to prove $\mathbb{E}(0) < \mathbb{E}(1)$. Since $\mathbb{E}(1) = 1 + p(1-p)^{n-1}\mathbb{E}(0) + (1 - p(1-p)^{n-1})\mathbb{E}(1)$, we have $\mathbb{E}(1) = 1/(p(1-p)^{n-1}) > 0 = \mathbb{E}(0)$.

(b) Inductive Hypothesis assumes that

$$\forall\, 0 \leq j < K (K \leq n-1) : \mathbb{E}(j) < \mathbb{E}(j+1).$$

Then, we consider $j = K$. Let x and x' be a solution with $K+1$ number of 0 bits and that with K number of 0 bits, respectively. Then, we have $\mathbb{E}(K+1) = \mathbb{E}[\![\tau'_t|\xi'_t = x]\!]$ and $\mathbb{E}(K) = \mathbb{E}[\![\tau'_t|\xi'_t = x']\!]$. For a Boolean string of length $n-1$ with K number of 0 bits, we denote P_i $(0 \leq i \leq n-1)$ as the probability that the number of 0 bits changes to be i after bit-wise mutation on this string with mutation probability p.

For the solution x, we divide the mutation on x into two parts: mutation on one 0 bit and mutation on the $n-1$ remaining bits. The $n-1$ remaining bits contain K number of 0 bits since $n - \|x\| = K + 1$. Then, by considering the mutation and selection behavior of the (1+1)-EA on the OneMax problem, we have

$$\begin{aligned}\mathbb{E}(K+1) = 1 + p \cdot (&\textstyle\sum_{i=0}^{K+1} P_i\mathbb{E}(i) + \sum_{i=K+2}^{n-1} P_i\mathbb{E}(K+1)) \\ &+ (1-p) \cdot (\textstyle\sum_{i=0}^{K} P_i\mathbb{E}(i+1) + \sum_{i=K+1}^{n-1} P_i\mathbb{E}(K+1)),\end{aligned}$$

where the term p is the probability that the 0 bit in the first mutation part is flipped.

For the solution x', we also divide the mutation on x' into two parts: mutation on one 1 bit and mutation on the $n-1$ remaining bits. The $n-1$ remaining bits also contain K number of 0 bits since $n - \|x'\| = K$. Then, we have

$$\begin{aligned}\mathbb{E}(K) = 1 + p \cdot (&\sum\nolimits_{i=0}^{K-1} P_i \mathbb{E}(i+1) + \sum\nolimits_{i=K}^{n-1} P_i \mathbb{E}(K)) \\ + (1-p) \cdot (&\sum\nolimits_{i=0}^{K} P_i \mathbb{E}(i) + \sum\nolimits_{i=K+1}^{n-1} P_i \mathbb{E}(K)),\end{aligned}$$

where the term p is the probability that the 1 bit in the first mutation part is flipped.

From the above two equalities, we have

$$\begin{aligned}&\mathbb{E}(K+1) - \mathbb{E}(K) \\ &= p \cdot \big(\sum\nolimits_{i=0}^{K-1} P_i(\mathbb{E}(i) - \mathbb{E}(i+1)) + \sum\nolimits_{i=K+1}^{n-1} P_i(\mathbb{E}(K+1) - \mathbb{E}(K))\big) \\ &+ (1-p) \cdot \big(\sum\nolimits_{i=0}^{K} P_i(\mathbb{E}(i+1) - \mathbb{E}(i)) + \sum\nolimits_{i=K+1}^{n-1} P_i(\mathbb{E}(K+1) - \mathbb{E}(K))\big) \\ &= (1-2p) \cdot \big(\sum\nolimits_{i=0}^{K-1} P_i(\mathbb{E}(i+1) - \mathbb{E}(i))\big) \\ &+ \big((1-p)P_K + \sum\nolimits_{i=K+1}^{n-1} P_i\big) \cdot \big(\mathbb{E}(K+1) - \mathbb{E}(K)\big) \\ &> \big((1-p)P_K + \sum\nolimits_{i=K+1}^{n-1} P_i\big) \cdot \big(\mathbb{E}(K+1) - \mathbb{E}(K)\big),\end{aligned}$$

where the inequality is by $0 < p < 0.5$ and inductive hypothesis.

Since $(1-p)P_K + \sum_{i=K+1}^{n-1} P_i < 1$, we have $\mathbb{E}(K+1) > \mathbb{E}(K)$. $\square$

For the Trap problem, it is not hard to see that the CFHT $\mathbb{E}[\![\tau'_t | \xi'_t = x]\!]$ of the (1+1)-EA on the Trap problem also only depends on the number of 1 bits of the solution x. Thus, we denote $\mathbb{E}'(j)$ as the CFHT $\mathbb{E}[\![\tau'_t | \xi'_t = x]\!]$ with $\|x\| = n - j$. Then, it is obvious that $\mathbb{E}'(0) = 0$, which implies the optimal solution.

Lemma 2. *For* $0 < p < 0.5$*, it holds that* $\mathbb{E}'(0) < \mathbb{E}'(1) < \mathbb{E}'(2) < \ldots < \mathbb{E}'(n)$.

Proof. First, $\mathbb{E}'(0) < \mathbb{E}'(1)$ trivially holds, since $\mathbb{E}'(0) = 0$ and $\mathbb{E}'(1) > 0$. Then, we prove $\forall\, 0 < j < n : \mathbb{E}'(j) < \mathbb{E}'(j+1)$ inductively on j.

(a) Initialization is to prove $\mathbb{E}'(n-1) < \mathbb{E}'(n)$. For $\mathbb{E}'(n)$, since only the offspring 0^n or 1^n will be accepted, we have $\mathbb{E}'(n) = 1 + p^n \mathbb{E}'(0) + (1-p^n)\mathbb{E}'(n)$, then, $\mathbb{E}'(n) = 1/p^n$. For $\mathbb{E}'(n-1)$, since the accepted offsprings are 0^n, the solutions with $n-1$ number of 0 bits and 1^n, we have $\mathbb{E}'(n-1) = 1 + p^{n-1}(1-p)\mathbb{E}'(0) + p(1-p)^{n-1}\mathbb{E}'(n) + (1 - p^{n-1}(1-p) - p(1-p)^{n-1})\mathbb{E}'(n-1)$, then, $\mathbb{E}'(n-1) = (1 + (1-p)^{n-1}/p^{n-1})/(p^{n-1}(1-p) + p(1-p)^{n-1})$. Thus, we have

$$\frac{\mathbb{E}'(n)}{\mathbb{E}'(n-1)} = \frac{p^{n-1}(1-p) + p(1-p)^{n-1}}{p^n + (1-p)^{n-1}p} > 1,$$

where the inequality is by $0 < p < 0.5$.

(b) Inductive Hypothesis assumes that

$$\forall\ K < j \leq n-1 (K \geq 1) : \mathbb{E}'(j) < \mathbb{E}'(j+1).$$

Then, we consider $j = K$. When comparing $\mathbb{E}'(K+1)$ with $\mathbb{E}'(K)$, we use the same analysis method as that in the proof of Lemma 1. By additionally considering the selection behavior of the (1+1)-EA on the Trap problem which is different from that on the OneMax problem, we can get

$$\begin{aligned}\mathbb{E}'(K+1) = 1 + p \cdot (P_0\mathbb{E}'(0) + \sum\nolimits_{i=1}^{K} P_i\mathbb{E}'(K+1) + \sum\nolimits_{i=K+1}^{n-1} P_i\mathbb{E}'(i))\\ + (1-p) \cdot (\sum\nolimits_{i=0}^{K-1} P_i\mathbb{E}'(K+1) + \sum\nolimits_{i=K}^{n-1} P_i\mathbb{E}'(i+1)),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}'(K) = 1 + p \cdot (\sum\nolimits_{i=0}^{K-2} P_i\mathbb{E}'(K) + \sum\nolimits_{i=K-1}^{n-1} P_i\mathbb{E}'(i+1))\\ + (1-p) \cdot (P_0\mathbb{E}'(0) + \sum\nolimits_{i=1}^{K-1} P_i\mathbb{E}'(K) + \sum\nolimits_{i=K}^{n-1} P_i\mathbb{E}'(i)).\end{aligned}$$

From the above two equalities, we have

$$\begin{aligned}&\mathbb{E}'(K+1) - \mathbb{E}'(K) = p \cdot \big(P_0(\mathbb{E}'(0) - \mathbb{E}'(K)) + \sum\nolimits_{i=1}^{K-1} P_i(\mathbb{E}'(K+1) - \mathbb{E}'(K))\\ &+ \sum\nolimits_{i=K+1}^{n-1} P_i(\mathbb{E}'(i) - \mathbb{E}'(i+1))\big) + (1-p) \cdot \big(P_0(\mathbb{E}'(K+1) - \mathbb{E}'(0))\\ &+ \sum\nolimits_{i=1}^{K} P_i(\mathbb{E}'(K+1) - \mathbb{E}'(K)) + \sum\nolimits_{i=K+1}^{n-1} P_i(\mathbb{E}'(i+1) - \mathbb{E}'(i))\big)\\ &= P_0 \cdot \big((1-p)\mathbb{E}'(K+1) - p\mathbb{E}'(K)\big) + \big(\sum\nolimits_{i=1}^{K-1} P_i + (1-p)P_K\big)\\ &\cdot \big(\mathbb{E}'(K+1) - \mathbb{E}'(K)\big) + (1-2p) \cdot \big(\sum\nolimits_{i=K+1}^{n-1} P_i(\mathbb{E}'(i+1) - \mathbb{E}'(i))\big)\\ &> \big(\sum\nolimits_{i=1}^{K-1} P_i + (1-p)P_K + pP_0\big) \cdot \big(\mathbb{E}'(K+1) - \mathbb{E}'(K)\big),\end{aligned}$$

where the inequality is by $0 < p < 0.5$ and inductive hypothesis.

Since $\sum_{i=1}^{K-1} P_i + (1-p)P_K + pP_0 < 1$, we have $\mathbb{E}'(K+1) > \mathbb{E}'(K)$. □

Proof of Theorem 3. The pseudo-Boolean function class with a unique global optimum is homogeneous, since for each dimensionality n, the solution space and the optimal solution for any function are $\{0,1\}^n$ and 1^n, respectively. By the behavior of the (1+1)-EA, it is easy to see that the (1+1)-EA can be modeled as a Markov chain.

Let the OneMax problem correspond to f^* in Theorem 2. Then for the parameter m and $\mathcal{X}_i^t$ in Theorem 2, we have $m = n$ and $\mathcal{X}_i^t = \{x \mid \|x\| = n - i\}$ $(0 \leq i \leq n)$ by Lemma 1. For any non-optimal solution $x \in \mathcal{X}_k^t$ $(k > 0)$, we denote $P(j)$ $(0 \leq j \leq n)$ as the probability that the offspring generated by bit-wise mutation on x has j number of 0 bits. For $\{\xi'_t\}_{t=0}^{+\infty}$, since only the offspring solution with no more 0 bits will be accepted, we have

$$\begin{aligned}&\forall\ 0 \leq j \leq k-1 : P_{\xi'}^t(x, \mathcal{X}_j^t) = P(j); \quad P_{\xi'}^t(x, \mathcal{X}_k^t) = \sum\nolimits_{j=k}^{n} P(j);\\ &\forall\ k+1 \leq j \leq n : P_{\xi'}^t(x, \mathcal{X}_j^t) = 0.\end{aligned}$$

For $\{\xi_t\}_{t=0}^{+\infty}$, since the offspring solution with less 0 bits may be rejected and that with more 0 bits may be accepted, we have

$$P_\xi^t(x, \mathcal{X}_0^t) = P(0); \qquad \forall\, 1 \le j \le k-1 : P_\xi^t(x, \mathcal{X}_j^t) \le P(j);$$
$$\forall\, k+1 \le j \le n : P_\xi^t(x, \mathcal{X}_j^t) \ge 0.$$

Thus, if $P_{\xi'}^t(x, \mathcal{X}_k^t) \ge P_\xi^t(x, \mathcal{X}_k^t)$, we have

$$\forall\, 0 \le j \le k : P_\xi^t(x, \mathcal{X}_j^t) \le P_{\xi'}^t(x, \mathcal{X}_j^t),\ \forall\, k+1 \le j \le n : P_\xi^t(x, \mathcal{X}_j^t) \ge P_{\xi'}^t(x, \mathcal{X}_j^t);$$

otherwise, we have

$$\forall\, 0 \le j \le k-1 : P_\xi^t(x, \mathcal{X}_j^t) \le P_{\xi'}^t(x, \mathcal{X}_j^t),\ \forall\, k \le j \le n : P_\xi^t(x, \mathcal{X}_j^t) \ge P_{\xi'}^t(x, \mathcal{X}_j^t).$$

Note that the above two formulas hold for all $t \ge 0$, since the (1+1)-EA uses time-invariant operators. Therefore, by Theorem 2, we get that the OneMax problem is the easiest in the pseudo-Boolean function class with a unique global optimum for the (1+1)-EA with $p < 0.5$.

Let the Trap problem correspond to f^*. By Lemma 2, we have $m = n$ and $\mathcal{X}_i^t = \{x \mid \|x\| = n - i\}$ $(1 \le i \le n)$. For any non-optimal solution $x \in \mathcal{X}_k^t$ $(k > 0)$, we also denote $P(j)$ $(0 \le j \le n)$ as the probability that the offspring generated by bit-wise mutation on x has j number of 0 bits. For $\{\xi_t'\}_{t=0}^{+\infty}$, since only the optimal solution and the offspring solutions with no less 0 bits will be accepted, we have

$$P_{\xi'}^t(x, \mathcal{X}_0^t) = P(0); \qquad \forall\, 1 \le j \le k-1 : P_{\xi'}^t(x, \mathcal{X}_j^t) = 0;$$
$$P_{\xi'}^t(x, \mathcal{X}_k^t) = \sum\nolimits_{j=1}^{k} P(j); \quad \forall\, k+1 \le j \le n : P_{\xi'}^t(x, \mathcal{X}_j^t) = P(j).$$

For $\{\xi_t\}_{t=0}^{+\infty}$, since the offspring solution with less 0 bits may be accepted and that with more 0 bits may be rejected, we have

$$P_\xi^t(x, \mathcal{X}_0^t) = P(0); \qquad \forall\, 1 \le j \le k-1 : P_\xi^t(x, \mathcal{X}_j^t) \ge 0;$$
$$\forall\, k+1 \le j \le n : P_\xi^t(x, \mathcal{X}_j^t) \le P(j).$$

Thus, if $P_{\xi'}^t(x, \mathcal{X}_k^t) \ge P_\xi^t(x, \mathcal{X}_k^t)$, we have

$$\forall\, 0 \le j \le k-1 : P_\xi^t(x, \mathcal{X}_j^t) \ge P_{\xi'}^t(x, \mathcal{X}_j^t),\ \forall\, k \le j \le n : P_\xi^t(x, \mathcal{X}_j^t) \le P_{\xi'}^t(x, \mathcal{X}_j^t);$$

otherwise, we have

$$\forall\, 0 \le j \le k : P_\xi^t(x, \mathcal{X}_j^t) \ge P_{\xi'}^t(x, \mathcal{X}_j^t),\ \forall\, k+1 \le j \le n : P_\xi^t(x, \mathcal{X}_j^t) \le P_{\xi'}^t(x, \mathcal{X}_j^t).$$

By Theorem 2, we get that the Trap problem is the hardest in the pseudo-Boolean function class with a unique global optimum for the (1+1)-EA with $p < 0.5$. $\square$

5 Conclusion

In this paper, we derive a theorem to identify the easiest and the hardest problem cases of a problem class for an algorithm. Using the theorem, we prove that the OneMax and the Trap problem are the easiest and the hardest function in the pseudo-Boolean function class with a unique global optimum for the (1+1)-EA with mutation probability less than 0.5, respectively, which much extends the previous knowledge [3]. In the future, we will apply this theorem for more problem classes and more algorithms.

References

1. Auger, A., Doerr, B.: Theory of Randomized Search Heuristics: Foundations and Recent Developments. World Scientific, Singapore (2011)
2. Bäck, T.: Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms. Oxford University Press, Oxford, UK (1996)
3. Doerr, B., Johannsen, D., Winzen, C.: Drift analysis and linear functions revisited. In: Proceedings of the 2010 IEEE Congress on Evolutionary Computation. pp. 1–8. Barcelona, Spain (2010)
4. Droste, S., Jansen, T., Wegener, I.: A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. Evolutionary Computation 6(2), 185–196 (1998)
5. Droste, S., Jansen, T., Wegener, I.: On the analysis of the (1+1) evolutionary algorithm. Theoretical Computer Science 276(1-2), 51–81 (2002)
6. Fournier, H., Teytaud, O.: Lower bounds for comparison based evolution strategies using VC-dimension and sign patterns. Algorithmica 59(3), 387–408 (2011)
7. He, J., Yao, X.: Drift analysis and average time complexity of evolutionary algorithms. Artificial Intelligence 127(1), 57–85 (2001)
8. He, J., Yao, X.: A study of drift analysis for estimating computation time of evolutionary algorithms. Natural Computing 3(1), 21–35 (2004)
9. Jägersküpper, J.: A blend of Markov-chain and drift analysis. In: Proceedings of the 10th International Conference on Parallel Problem Solving from Nature. pp. 41–51. Dortmund, Germany (2008)
10. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of the 1995 IEEE International Conference on Neural Networks. pp. 1942–1948. Perth, Australia (1995)
11. Kirkpatrick, S.: Optimization by simulated annealing: Quantitative studies. Journal of Statistical Physics 34(5), 975–986 (1984)
12. Neumann, F., Witt, C.: Bioinspired Computation in Combinatorial Optimization: Algorithms and Their Computational Complexity. Springer-Verlag, Berlin, Germany (2010)
13. Witt, C.: Optimizing linear functions with randomized search heuristics-The robustness of mutation. In: Proceedings of the 29th Symposium on Theoretical Aspects of Computer Science. pp. 420–431. Paris, France (2012)
14. Wolpert, D., Macready, W.: No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation 1(1), 67–82 (1997)
15. Yu, Y., Qian, C., Zhou, Z.H.: Towards analyzing recombination operators in evolutionary search. In: Proceedings of the 11th International Conference on Parallel Problem Solving from Nature. pp. 144–153. Krakow, Poland (2010)
16. Yu, Y., Qian, C., Zhou, Z.H.: Towards analyzing crossover operators in evolutionary search via general Markov chain switching theorem. CORR abs/1111.0907 (2011)
17. Yu, Y., Zhou, Z.H.: A new approach to estimating the expected first hitting time of evolutionary algorithms. Artificial Intelligence 172(15), 1809–1832 (2008)