Running Time Analysis: Convergence-based Analysis Reduces to Switch Analysis

Yang Yu and Chao Qian
National Key Laboratory for Novel Software Technology
Nanjing University, Nanjing 210023, China
Email: {yuy,qianc}@lamda.nju.edu.cn

***Abstract*—Evolutionary algorithms (EAs) are general purpose optimization tools that can be applied in various situations, therefore, general analysis approaches are appealing for facilitating the analysis of EAs in different problems. Expected running time is a key theoretical issue of evolutionary algorithms (EAs). Several general analysis approaches for the running time analysis of EAs have been proposed and have stimulated the theoretical development. Recently, *switch analysis* was proposed, which derives the running time of an EA process by comparing it with a simpler EA process. It has been proven that *drift analysis* and *fitness level method* are *reducible* to switch analysis, which means that switch analysis can derive at least as tight results as the two approaches. In this paper, we further prove that another analysis approach, *convergence-based analysis,* is reducible to switch analysis. We also show in a case study that switch analysis leads to a tighter result than convergence-based analysis.**

I. INTRODUCTION

Expected running time is among the central theoretical properties of evolutionary algorithms (EAs), as it discloses the time complexity of EAs. Many studies have devoted themselves to the running time analysis of EAs in the recent decades, resulting in progresses on the theoretical understanding of EAs [1], [6], [7]. Among the progresses, one interesting kind of work is the development of analysis approaches. An analysis approach gives a guideline on how to analyze EAs: where to look, which quantities to calculate, and what procedure to follow. This is particularly important for EAs. EAs are general purpose optimization techniques, which could come across all kinds of problems. Thus we cannot presume that EAs will be used to solve only previously investigated problems. When engaged in a new problem with no previous knowledge, the analysis approach can lead to an analysis result much easier.

Several analysis approaches have been developed with different ideas and principles. *Fitness level method* [9], [10] organizes all solutions into level sets, according to their fitness values. It then requires to estimate the probabilities that solutions from one level transits to other levels. The estimation of the probabilities leads to the estimation of the expected running time. *Drift analysis* [2]–[4], [8] requires to estimate the one-step progress toward the optimal solution, where the progress is measured by a given distance function. The expected running time is then derived by dividing the total distance by the one-step progress. *Convergence-based analysis* [12] requires to estimate the probabilities that the algorithm could achieve the optimal solution in every step, or the one-step successful probability for short. The expected running time is then calculated from that probabilities.

Recently, a new method, *switch analysis* [11] was developed. Unlike the previous analysis approaches that analyze an evolutionary process[1] from scratch, switch analysis compares two evolutionary processes. It requires to estimate the one-step differences of the two evolutionary processes. The one-step differences cumulate to be the difference of the expected running time of the two evolutionary processes. Thus when one evolutionary process is well analyzed, the expected running time of the other can be derived.

More interestingly, it has been proven that the fitness level method and drift analysis[2] are *reducible* to switch analysis, which means that switch analysis can derive at least as tight analysis results as the other two. Moreover, the proofs are constructive, and thus can help us view the analysis approaches from a unified perspective, as the two approaches are equivalent with particular configurations of switch analysis. However, it was unknown if convergence-based analysis is reducible to switch analysis.

In this paper, we prove that convergence-based analysis is reducible to switch analysis. We also provide a case study where switch analysis leads to a tighter result than the previously obtained by convergence-based analysis. Moreover, from a constructive proof of reducibility, we can observe that an approach is equivalent with a configuration of switch analysis. Then through comparing the equivalent configurations, we discuss the differences among convergence-based analysis, drift analysis, and the fitness-level method.

This work was supported by National Science Foundation of China (61375061,61223003), Jiangsu Science Foundation (BK2012303), Foundation for the Author of National Excellent Doctoral Dissertation of China (201451), and Fundamental Research Funds for the Central Universities.
978-1-4799-7492-4/15/$31.00

[1]By evolutionary process, we mean the process that an EA solves a problem, or in other words, the underlying Markov chain process.

[2]Drift analysis has many versions, and we mean in this paper the additive version investigated in [11]. However the discussion is also suitable for any version that can be derived from the additive version.

The rest of the paper is organized as follows. Section II introduces background knowledge, including preliminaries, the convergence-based analysis, the switch analysis, and the reducibility between approaches. Section III proves that the convergence-based analysis reduces to the switch analysis. Section IV shows that the switch analysis derives a better result than the previous one by the convergence-based analysis. Section V discusses and Section VI concludes the paper.

II. Background

A. Preliminaries

A Markov chain is a random process, $\{\xi_t\}_{t=0}^{+\infty}$, where the variable ξ_t is in the state space $\mathcal{X}$ and ξ_{t+1} depends only on the variable ξ_t, i.e., $P(\xi_{t+1} \mid \xi_t, \xi_{t-1}, \ldots, \xi_0) = P(\xi_{t+1} \mid \xi_t)$. To model an EA, let $\mathcal{S}$ be the solution space of a problem, then the EA with population size m is modeled by a Markov chain with state space $\mathcal{X} \subseteq \mathcal{S}^m$ and transition probability $P(\xi_{t+1} \mid \xi_t)$ defined by its reproduction operators and its selection. Let $\mathcal{X}^* \subset \mathcal{X}$ denote the optimal region, in which a population contains at least one optimal solution. It should be clear that a Markov chain models an EA process, i.e., the process of the running of an EA on a problem instance. In the rest of the paper, we will describe a Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ with state space $\mathcal{X}$ as "$\xi \in \mathcal{X}$" for simplicity. Also note that we do not assume homogeneous Markov chains, i.e., we allow $P(\xi_{t+1} \mid \xi_t)$ depends on t, since both switch analysis and convergence-based analysis apply to non-homogeneous Markov chains.

We consider the performance measure *expected first hitting time* defined below, which is the average number of iterations of an EA for finding the optimal solution:

Definition 1 (Conditional first hitting time, CFHT)
Given a Markov chain $\xi \in \mathcal{X}$ and a target subspace $\mathcal{X}^* \subset \mathcal{X}$, starting from time t_0 where $\xi_{t_0} = x$, let τ be a random variable that denotes the hitting events:

$$\tau = \begin{cases} 0, \text{ if } \xi_{t_0} \in \mathcal{X}^* \\ i, \text{ if } \xi_{t_0+i} \in \mathcal{X}^* \wedge (\xi_{t_0} \notin \mathcal{X}^* \wedge \ldots \wedge \xi_{t_0+i-1} \notin \mathcal{X}^*) \end{cases}$$

The conditional expectation of τ,

$$\mathbb{E}[\![\tau \mid \xi_{t_0} = x]\!] = \sum\nolimits_{i=0}^{+\infty} i \cdot P(\tau = i),$$

is called the conditional first hitting time (CFHT) of the Markov chain from $t = t_0$ and $\xi_{t_0} = x$.

Definition 2 (Distribution-CFHT, DCFHT)
Given a Markov chain $\xi \in \mathcal{X}$ and a target subspace $\mathcal{X}^* \subset \mathcal{X}$, starting from time t_0 where ξ_{t_0} is drawn from a state distribution π, the expectation of the CFHT,

$$\begin{aligned} \mathbb{E}[\![\tau \mid \xi_{t_0} \sim \pi]\!] &= \mathbb{E}_{x\sim\pi}[\![\tau \mid \xi_{t_0} = x]\!] \\ &= \sum\nolimits_{x\in\mathcal{X}} \pi(x) \mathbb{E}[\![\tau \mid \xi_{t_0} = x]\!], \end{aligned}$$

is called the distribution-conditional first hitting time (DCFHT) of the Markov chain from $t = t_0$ and $\xi_{t_0} \sim \pi$.

Definition 3 (Expected first hitting time, EFHT)
Given a Markov chain $\xi \in \mathcal{X}$ and a target subspace $\mathcal{X}^* \subset \mathcal{X}$, the DCFHT of the chain from $t = 0$ and uniform distribution π_u,

$$\begin{aligned} \mathbb{E}[\![\tau]\!] &= \mathbb{E}[\![\tau \mid \xi_0 \sim \pi_u]\!] = \mathbb{E}_{x\sim\pi_u}[\![\tau \mid \xi_0 = x]\!] \\ &= \frac{1}{|\mathcal{X}|} \sum_{x\in\mathcal{X}} \mathbb{E}[\![\tau \mid \xi_0 = x]\!], \end{aligned}$$

is called the expected first hitting time (EFHT) of the Markov chain.

To reflect the computational time complexity of an EA, we count the number of evaluations to solutions, i.e., EFHT $\times$ the number of offspring solutions in each iteration, which is called the *expected running time* of the EA.

We call a chain *absorbing* (with a slight abuse of the term) if all states in $\mathcal{X}^*$ are absorbing states. The definition of absorbing is only for the analysis purpose, and is not a restriction of EAs. This is because that, given any non-absorbing chain, we can construct a corresponding absorbing chain that simulates the non-absorbing chain but stays in the optimal state once it has been found. The EFHT of the constructed absorbing chain is the same as that of the non-absorbing chain. Thus it is sufficient to study only the absorbing chain. We assume all chains considered in this paper are absorbing.

Definition 4 (Absorbing Markov Chain)
Given a Markov chain $\xi \in \mathcal{X}$ and a target subspace $\mathcal{X}^* \subset \mathcal{X}$, ξ is said to be an absorbing chain, if

$$\forall x \in \mathcal{X}^*, \forall t \geq 0 : P(\xi_{t+1} \neq x \mid \xi_t = x) = 0.$$

The following basic property of Markov chains will be used in this paper.

Lemma 1
Given an absorbing Markov chain $\xi \in \mathcal{X}$ and a target subspace $\mathcal{X}^* \subset \mathcal{X}$, we have that $\mathbb{E}[\![\tau | \xi_t \in \mathcal{X}^*]\!] = 0$ and

$$\begin{aligned} &\forall x \notin \mathcal{X}^* : \mathbb{E}[\![\tau \mid \xi_t = x]\!] \\ &= 1 + \sum_{y\in\mathcal{X}} P(\xi_{t+1} = y \mid \xi_t = x) \mathbb{E}[\![\tau \mid \xi_{t+1} = y]\!]. \end{aligned}$$

B. Convergence-based Analysis

Yu and Zhou [12] derived convergence-based analysis for running time analysis. It utilizes the relationship between the convergence rate, i.e., $1 - \pi_0(\mathcal{X}^*)$, and the hitting time. The convergence rate is bounded using the *normalized success probability* as in Lemma 2, where α_t and β_t are the lower and upper bounds, respectively. Converting from the convergence rate, convergence-based analysis is in Theorem 1. Using this approach, as long as we can bound the normalized one-step success probability, we can arrive at the bounds of the DCEFHT. Note that the theorem is presented with a bit difference with that in [12] since we take the initial state distribution into consideration.

Lemma 2 ([5])
Given an absorbing Markov chain $\xi \in \mathcal{X}$ and a target subspace $\mathcal{X}^* \subset \mathcal{X}$, if two sequences $\{\alpha_t\}_{t=0}^{+\infty}$ and $\{\beta_t\}_{t=0}^{+\infty}$ satisfy

$$\prod_{t=0}^{+\infty}(1-\alpha_t)=0$$

and

$$\beta_t \geq \sum_{x\notin\mathcal{X}^*} P(\xi_{t+1}\in\mathcal{X}^* \mid \xi_t = x)\frac{\pi_t(x)}{1-\pi_t(\mathcal{X}^*)} \geq \alpha_t,$$

then the chain converges to $\mathcal{X}^*$, and that[3]

$$1-\pi_t(\mathcal{X}^*) \geq (1-\pi_0(\mathcal{X}^*))\prod_{i=0}^{t-1}(1-\beta_i),$$

$$1-\pi_t(\mathcal{X}^*) \leq (1-\pi_0(\mathcal{X}^*))\prod_{i=0}^{t-1}(1-\alpha_i).$$

Theorem 1 (Convergence-based Analysis [12])
Given an absorbing Markov chain $\xi \in \mathcal{X}$ and a target subspace $\mathcal{X}^* \subset \mathcal{X}$, let τ denote the hitting event of ξ, and let π_t denote the distribution of ξ_t. If two sequences $\{\alpha_t\}_{t=0}^{+\infty}$ and $\{\beta_t\}_{t=0}^{+\infty}$ satisfy

$$\prod_{t=0}^{+\infty}(1-\alpha_t)=0$$

and

$$\beta_t \geq \sum_{x\notin\mathcal{X}^*} P(\xi_{t+1}\in\mathcal{X}^* \mid \xi_t = x)\frac{\pi_t(x)}{1-\pi_t(\mathcal{X}^*)} \geq \alpha_t,$$

we have

$$\mathbb{E}[\![\tau \mid \xi_0 \sim \pi_0]\!] \leq (1-\pi_0(\mathcal{X}^*))(\sum_{t=1}^{+\infty} t\alpha_{t-1}\prod_{i=0}^{t-2}(1-\alpha_i))$$

and

$$\mathbb{E}[\![\tau \mid \xi_0 \sim \pi_0]\!] \geq (1-\pi_0(\mathcal{X}^*))(\sum_{t=1}^{+\infty} t\beta_{t-1}\prod_{i=0}^{t-2}(1-\beta_i)).$$

C. Switch Analysis

Yu *et al.* [11] derived switch analysis for bounding EFHT of EAs. This approach compares two Markov chains, and by bounding step-wise difference, it bounds the difference between the DCEFHTs of the two chains. As presented in Theorem 2, by using an aligned mapping ϕ defined in Definition 5, the states of one chain are mapped to the states of the other chain, then ρ_t records the *switching difference* at step t.

Definition 5 (Aligned Mapping [11])
Given two spaces $\mathcal{X}$ and $\mathcal{Y}$ with target subspaces $\mathcal{X}^*$ and $\mathcal{Y}^*$, respectively, a function $\phi : \mathcal{X} \to \mathcal{Y}$ is called
(a) a left-aligned mapping *if $\forall x \in \mathcal{X}^* : \phi(x) \in \mathcal{Y}^*$;*
(b) a right-aligned mapping *if $\forall x \in \mathcal{X}-\mathcal{X}^* : \phi(x) \notin \mathcal{Y}^*$;*
(c) an optimal-aligned mapping *if it is both left-aligned and right-aligned.*

Theorem 2 (Switch Analysis [11])
Given two absorbing Markov chains $\xi \in \mathcal{X}$ and $\xi' \in \mathcal{Y}$, let τ and τ' denote the hitting events of ξ and ξ', respectively, and let π_t denote the distribution of ξ_t. Given a series of values $\{\rho_t \in \mathbb{R}\}_{t=0}^{+\infty}$ with $\rho = \sum_{t=0}^{+\infty}\rho_t$ and a right (or left)-aligned mapping $\phi : \mathcal{X} \to \mathcal{Y}$, if $\mathbb{E}[\![\tau \mid \xi_0 \sim \pi_0]\!]$ is finite and

$$\forall t: \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \pi_t(x)P(\xi_{t+1}\in\phi^{-1}(y) \mid \xi_t = x)\mathbb{E}[\![\tau' \mid \xi'_0 = y]\!]$$
$$\leq (\text{or } \geq) \sum_{u,y\in\mathcal{Y}} \pi_t^{\phi}(u)P(\xi'_1 = y \mid \xi'_0 = u)\mathbb{E}[\![\tau' \mid \xi'_1 = y]\!] + \rho_t, \quad (1)$$

where $\pi_t^{\phi}(y) = \pi_t(\phi^{-1}(y)) = \sum_{x\in\phi^{-1}(y)}\pi_t(x)$, we have

$$\mathbb{E}[\![\tau \mid \xi_0 \sim \pi_0]\!] \leq (\text{or } \geq)\mathbb{E}[\![\tau' \mid \xi'_0 \sim \pi_0^{\phi}]\!] + \rho.$$

We can find that switch analysis does not estimate the DCEFHT of an evolutionary process directly. To do so, we need to choose a *reference chain*, and compare the evolutionary process with the reference chain by switch analysis. We then obtain the difference of their DCEFHTs. If we choose a reference chain that can be easily analyzed, switch analysis can be helpful to estimate the DCEFHT of the concerning evolutionary process. In the rest of the paper, we will call the Markov chain used to help analyze the concerning evolutionary process as the reference chain.

D. Reducibility

In [11], reducibility was proposed in order to compare analysis approaches. To study reducibility, the analysis approaches are firstly formally characterized as three parts, the input, the parameters and the output as in Definition 6. The input is a variable assignment derived from the concerning EA process; the parameters are variable assignments which should rely on no more information of the EA process than the input; and the output is a lower and upper bound of the running time. In this way, switch analysis is characterized in Characterization 1.

Definition 6 (EA Analysis Approach [11])
A procedure $\mathfrak{A}$ is called an EA analysis approach if for any EA process $\xi \in \mathcal{X}$ with initial state ξ_0 and transition probability P, $\mathfrak{A}$ provided with $\Theta = g(\xi_0, P)$ for some function g and a set of parameters $\Omega(\Theta)$ outputs a lower running time bound of ξ notated as $\mathfrak{A}^l(\Theta;\Omega)$ and/or an upper bound $\mathfrak{A}^u(\Theta;\Omega)$.

Characterization 1 (Switch Analysis [11])
For an EA process $\xi \in \mathcal{X}$, switch analysis approach $\mathfrak{A}_{SA}$ is defined by its parameters, input and output:
Parameters: a reference process $\xi' \in \mathcal{Y}$ with bounds of its transition probabilities $P(\xi'_1|\xi'_0)$ and CFHT $\mathbb{E}[\![\tau' \mid \xi'_t = y]\!]$ for all $y \in \mathcal{Y}$ and $t \in \{0,1\}$, and a right-aligned mapping $\phi^u : \mathcal{X} \to \mathcal{Y}$ or a left-aligned mapping $\phi^l : \mathcal{X} \to \mathcal{Y}$.
Input: bounds of one-step transition probabilities $P(\xi_{t+1} \mid \xi_t)$.
Output: denoting $\pi_t^{\phi}(y) = \pi_t(\phi^{-1}(y))$ for all $y \in \mathcal{Y}$, $\mathfrak{A}_{SA}^u = \mathbb{E}[\![\tau' \mid \xi'_0 \sim \pi_0^{\phi}]\!] + \rho^u$ where $\rho^u = \sum_{t=0}^{+\infty}\rho_t^u$ and $\rho_t^u \geq \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \pi_t(x)P(\xi_{t+1}\in\phi^{-1}(y)|\xi_t = x)\mathbb{E}[\![\tau'|\xi'_0 = y]\!] - \sum_{u,y\in\mathcal{Y}} \pi_t^{\phi}(u)P(\xi'_1 = y \mid \xi'_0 = u)\mathbb{E}[\![\tau' \mid \xi'_1 = y]\!]$ for all t;

[3] In this paper, we define $\prod_{i=a}^{b}(\cdot) = 1$ if $b < a$

$\mathfrak{A}^l_{SA} = \mathbb{E}[\![\tau' \mid \xi'_0 \sim \pi^\phi_0]\!] + \rho^l$ *where* $\rho^l = \sum_{t=0}^{+\infty} \rho^l_t$ *and* $\rho^l_t \leq \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \pi_t(x) P(\xi_{t+1} \in \phi^{-1}(y) | \xi_t = x) \mathbb{E}[\![\tau' | \xi'_0 = y]\!] - \sum_{u,y \in \mathcal{Y}} \pi^\phi_t(u) P(\xi'_1 = y \mid \xi'_0 = u) \mathbb{E}[\![\tau' \mid \xi'_1 = y]\!]$ *for all* t.

When analysis approaches are characterized, reducibility defined in Definition 7 compares two approaches. If approach A takes the same input as (or transformed input from) approach B, while A outputs expected running time bounds no worse than B, we call B is reducible to A. It means that A is at least as theoretically powerful as B on estimating the expected running time.

Definition 7 (Reducible [11])
For two EA analysis approaches $\mathfrak{A}_1$ *and* $\mathfrak{A}_2$*, if for any input* Θ *and any parameter* Ω_A*, there exist a transformation* T *and parameter* Ω_B *(which possibly depends on* Ω_A*) such that*
(a) $\mathfrak{A}^u_1(\Theta; \Omega_A) \geq \mathfrak{A}^u_2(T(\Theta); \Omega_B)$*, then* $\mathfrak{A}_1$ *is* upper-bound reducible *to* $\mathfrak{A}_2$*;*
(b) $\mathfrak{A}^l_1(\Theta; \Omega_A) \leq \mathfrak{A}^l_2(T(\Theta); \Omega_B)$*, then* $\mathfrak{A}_1$ *is* lower-bound reducible *to* $\mathfrak{A}_2$*.*
Moreover, $\mathfrak{A}_1$ *is* reducible *to* $\mathfrak{A}_2$ *if it is both upper-bound reducible and lower-bound reducible.*

III. Convergence-based Analysis Reduces to Switch Analysis

To study reducibility, we also need to characterize convergence-based analysis as in Characterization 2. Note that, in the original theorem of [12], $P(\xi_t = x)$ and $\mu_t = \sum_{x \in \mathcal{X}^*} P(\xi_t = x)$ are used; we equivalently replace them by $\pi_t(x)$ and $\pi_t(\mathcal{X}^*)$ respectively for the sake of clearness. We have also multiplied their original bounds by $1 - \pi_0(\mathcal{X}^*)$ since we do not assume starting from non-optimal solutions.

Characterization 2 (Convergence-based Analysis)
For an EA process $\xi \in \mathcal{X}$*, convergence-based analysis approach* $\mathfrak{A}_{CA}$ *is defined by its input and output:*
Input:
$\alpha_t \leq \sum_{x \notin \mathcal{X}^*} P(\xi_{t+1} \in \mathcal{X}^* \mid \xi_t = x) \frac{\pi_t(x)}{1 - \pi_t(\mathcal{X}^*)}$ *for all* $t \geq 0$*, with a restriction that* $\prod_{t=0}^{+\infty}(1 - \alpha_t) = 0$*;*
$\beta_t \geq \sum_{x \notin \mathcal{X}^*} P(\xi_{t+1} \in \mathcal{X}^* \mid \xi_t = x) \frac{\pi_t(x)}{1 - \pi_t(\mathcal{X}^*)}$ *for all* $t \geq 0$*.*
Output:
$\mathfrak{A}^u_{CA} = (1 - \pi_0(\mathcal{X}^*))(\sum_{t=1}^{+\infty} t\alpha_{t-1} \prod_{i=0}^{t-2}(1 - \alpha_i))$*;*
$\mathfrak{A}^l_{CA} = (1 - \pi_0(\mathcal{X}^*))(\sum_{t=1}^{+\infty} t\beta_{t-1} \prod_{i=0}^{t-2}(1 - \beta_i))$*.*

From the characterization, we can find that convergence-based analysis approach does not have parameters. We then prove Theorem 3 that convergence-based analysis reduces to switch analysis.

Theorem 3
$\mathfrak{A}_{CA}$ *is reducible to* $\mathfrak{A}_{SA}$*.*

Before proving the theorem, we introduce a simple Markov chain called OneJump-fix, which will be used as the reference chain in switch analysis. In OneJump-fix, a non-optimal state either jumps to an optimal state or stays as it is in one-step with a fixed probability.

Definition 8 (OneJump-fix)
OneJump-fix chain with state space $\mathcal{X}$ *and target subspace* $\mathcal{X}^*$ *is a homogeneous Markov chain* $\xi \in \mathcal{X}$ *with a parameter* p_{fix}*. Its initial state is selected from* $\mathcal{X}$ *uniformly at random, and its transition probability is defined as, for any* $x \in \mathcal{X}$ *and any* t*,*

$$P(\xi_{t+1} = y \mid \xi_t = x) = \begin{cases} p_{fix}, & y \in \mathcal{X}^* \\ 1 - p_{fix}, & y = x \\ 0, & \textit{otherwise} \end{cases}.$$

Theorem 3 is proved by combining Lemma 3 and Lemma 4, which respectively prove the upper bound and lower bound reducibility.

Lemma 3
$\mathfrak{A}_{CA}$ *is upper-bound reducible to* $\mathfrak{A}_{SA}$*.*

Proof: The proof is to find the parameters of $\mathfrak{A}_{SA}$ and the input of $\mathfrak{A}_{SA}$ from that of $\mathfrak{A}_{CA}$, and show that $\mathfrak{A}^u_{SA} \leq \mathfrak{A}^u_{CA}$.

Denote $\xi \in \mathcal{X}$ as the EA process we are going to analyze. We know the variable α_t as in Characterization 2, which is the input of $\mathfrak{A}_{CA}$.

We choose the reference chain ξ' as the OneJump-fix chain in the same space of ξ, i.e., the state space $\mathcal{X}$ and the target subspace $\mathcal{X}^*$. The parameter of ξ' is set as

$$p_{fix} = \frac{1}{\sum_{t=0}^{+\infty} \prod_{i=0}^{t-1}(1 - \alpha_i)}.$$

Then, we have $\forall t \geq 0, x \notin \mathcal{X}^* : \mathbb{E}[\![\tau' | \xi'_t = x]\!] = \sum_{t=0}^{+\infty} \prod_{i=0}^{t-1}(1 - \alpha_i)$. We construct the mapping function $\phi : \mathcal{X} \to \mathcal{Y}$ as that $\phi(x) = x$. It is easy to verify that ϕ is an optimal-aligned mapping.

We then calculate the upper bound output of $\mathfrak{A}_{SA}$ using the input of $\mathfrak{A}_{CA}$ and the reference process. For the left part of Eq.(1), we have

$$\begin{aligned} &\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \pi_t(x) P(\xi_{t+1} \in \phi^{-1}(y) | \xi_t = x) \mathbb{E}[\![\tau' | \xi'_0 = y]\!] \\ &= \sum_{x \in \mathcal{X}} \pi_t(x)(1 - P(\xi_{t+1} \in \mathcal{X}^* \mid \xi_t = x)) \sum_{t=0}^{+\infty} \prod_{i=0}^{t-1}(1 - \alpha_i) \\ &= (1 - \pi_{t+1}(\mathcal{X}^*)) \sum_{t=0}^{+\infty} \prod_{i=0}^{t-1}(1 - \alpha_i). \end{aligned}$$

For the right part of Eq.(1), we have

$$\begin{aligned} &\sum_{u,y \in \mathcal{Y}} \pi^\phi_t(u) P(\xi'_1 = y \mid \xi'_0 = u) \mathbb{E}[\![\tau' \mid \xi'_1 = y]\!] \\ &= \sum_{x \in \mathcal{X} - \mathcal{X}^*} \pi_t(x)(\sum_{t=0}^{+\infty} \prod_{i=0}^{t-1}(1 - \alpha_i) - 1) \quad \text{(by Lemma 1)} \\ &= (1 - \pi_t(\mathcal{X}^*))(\sum_{t=0}^{+\infty} \prod_{i=0}^{t-1}(1 - \alpha_i) - 1). \end{aligned}$$

Thus, for all $t \geq 0$,

$$\begin{aligned}
&\sum_{x\in\mathcal{X},y\in\mathcal{Y}} \pi_t(x)P(\xi_{t+1}\in\phi^{-1}(y)|\xi_t=x)\mathbb{E}[\![\tau'|\xi'_0=y]\!] \\
&\quad-\sum_{u,y\in\mathcal{Y}}\pi_t^{\phi}(u)P(\xi'_1=y|\xi'_0=u)\mathbb{E}[\![\tau'|\xi'_1=y]\!] \\
&=(\pi_t(\mathcal{X}^*)-\pi_{t+1}(\mathcal{X}^*))\sum_{t=0}^{+\infty}\prod_{i=0}^{t-1}(1-\alpha_i)+(1-\pi_t(\mathcal{X}^*)) \\
&\leq(\pi_t(\mathcal{X}^*)-\pi_{t+1}(\mathcal{X}^*))\sum_{t=0}^{+\infty}\prod_{i=0}^{t-1}(1-\alpha_i) \\
&\quad+(1-\pi_0(\mathcal{X}^*))\prod_{i=0}^{t-1}(1-\alpha_i)\quad\text{(by Lemma 2)}
\end{aligned}$$

Therefore, we find an assignment of ρ_t in Theorem 2, and thus

$$\begin{aligned}
\rho&=\sum_{t=0}^{+\infty}(\pi_t(\mathcal{X}^*)-\pi_{t+1}(\mathcal{X}^*))\sum_{t=0}^{+\infty}\prod_{i=0}^{t-1}(1-\alpha_i) \\
&\quad+\sum_{t=0}^{+\infty}(1-\pi_0(\mathcal{X}^*))\prod_{i=0}^{t-1}(1-\alpha_i) \\
&=(\pi_0(\mathcal{X}^*)-\lim_{t\to+\infty}\pi_t(\mathcal{X}^*)+1-\pi_0(\mathcal{X}^*))\sum_{t=0}^{+\infty}\prod_{i=0}^{t-1}(1-\alpha_i) \\
&=0.\quad\text{(since the chain converges to } \mathcal{X}^* \text{ by Lemma 2)}
\end{aligned} \tag{2}$$

We then can calculate the upper bound output, noticing $\pi_0^{\phi}(y)=\pi_0(y)$,

$$\begin{aligned}
\mathfrak{A}_{SA}^u&=\mathbb{E}[\![\tau'|\xi'_0\sim\pi_0^{\phi}]\!]+0 \\
&=(1-\pi_0(\mathcal{X}^*))\sum_{t=0}^{+\infty}\prod_{i=0}^{t-1}(1-\alpha_i) \\
&=(1-\pi_0(\mathcal{X}^*))\big(1+\sum_{t=2}^{+\infty}t(\prod_{i=0}^{t-2}(1-\alpha_i)-\prod_{i=0}^{t-1}(1-\alpha_i)) \\
&\qquad-(1-\alpha_0)\big) \\
&=(1-\pi_0(\mathcal{X}^*))(\sum_{t=1}^{+\infty}t\alpha_{t-1}\prod_{i=0}^{t-2}(1-\alpha_i))=\mathfrak{A}_{CA}^u,
\end{aligned} \tag{3}$$

which proves the lemma. ■

Lemma 4
$\mathfrak{A}_{CA}$ is lower-bound reducible to $\mathfrak{A}_{SA}$.

The proof for Lemma 4 is similar to that for Lemma 3 except some minor changes: α and $\leq$ are replaced with β and $\geq$ respectively; the last "$=$" of Eq.(2) and the third "$=$" of Eq.(3) are replaced with "$\geq$" since the convergence cannot be derived using only the input β.

IV. Trap Problem Revisit

As we have proved that convergence-based analysis reduces to switch analysis, we compare them in a case study that solving the Trap problem by the (1+1)-EA.

Definition 9 (Trap Problem)
Given a set of n positive values, i.e., $W=\{w_i\}_{i=1}^n$, and a capacity value c, it is to find x^* such that

$$x^*=\arg\max_{x\in\{0,1\}^n}\sum_{i=1}^n w_ix_i$$
$$\text{s.t.}\quad \sum_{i=1}^n w_ix_i\leq c,$$

where $w_1=w_2=\ldots=w_{n-1}>1$, $w_n=\sum_{i=1}^{n-1}w_i+1$ and $c=w_n$.

The fitness function for solving the Trap problem is defined as

$$\forall x\in\{0,1\}^n: f(x)=\theta\sum_{i=1}^n w_ix_i-c,$$

where $\theta=1$ if x is a feasible solution, i.e., $\sum_{i=1}^n w_ix_i\leq c$, and $\theta=0$ otherwise. It is easy to see that the optimal solution is $0^{n-1}1$, and other solutions with 1 on the last bit are infeasible.

This fitness function is deceptive, as it will lead the evolution towards a sub-optimum, which is far from the global optimum.

The (1+1)-EA is a simplified evolutionary algorithms that has been frequently used in theoretical studies. It resembles most EAs in the structure: starting from a randomly generated solution, and iteratively improving the solution by reproduction and selection. It uses a bit-wise mutation operator in line 2 to generate a new solution from the current one. The bit-wise mutation randomly flips each bit of the current solution with a probability. After that, the new solution replaces the current one if the new one is not worse. We know an upper bound of EFHT of (1+1)-EA on the Trap problem, $\frac{1}{1-p}(\frac{1}{p})^{n-1}$, which is the worst case that needs to flip all the leading $n-1$ number of 0-bits of the local minimum $0^{n-1}1$, and $O(n^{n-1})$ when $p=\frac{1}{n}$.

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

1. *$x:=$choose a solution from $\mathcal{X}=\{0,1\}^n$ uniformly at random.*
2. *$x':=$ mutation x by randomly flippling each bit (0 to 1 and vice versa) with a probability $p\in(0,0.5]$.*
3. *If $f(x')\geq f(x)$, $x:=x'$.*
4. *Terminate if x is optimal.*
5. *Goto step 2.*

where mutation$(\cdot):\mathcal{X}\to\mathcal{X}$ is a mutation operator.

Before proving the EFHT of (1+1)-EA with different mutation operators on the Trap problem, we introduce a simple Markov chain, TrapJump, that will be used as the reference chain for switch analysis. The TrapJump chain is a very simple Markov chain with only $n+1$ states, named 0 to n. For non-optimal state $i\in\{1,2,\ldots,n\}$, there can be only three possible one-step transitions: jump to the optimal state 0, move one step farther from the optimal state (i.e., state $i+1$), or stay where it was (i.e., state i). And the farther from the optimal state, the lower the probability that it transits to the optimal state in one step. This chain reflects our intuition on the deceptiveness of the Trap problem.

Definition 10 (TrapJump)
TrapJump chain with size n and parameter $p \in (0, 0.5]$ is a homogeneous Markov chain $\xi \in \mathcal{X}$ with the state space $\mathcal{X} = \{0, 1, \dots, n\}$ and the optimal state 0. Its transition probability satisfies that, for all $t \geq 0$,

$$\forall i > 1, P(\xi_{t+1} = y \mid \xi_t = i) = \begin{cases} p^i(1-p)^{n-i}, & y = 0 \\ (n-i)p(1-p)^{n-1}, & y = i+1 \\ 1 - p^i(1-p)^{n-i} - (n-i)p(1-p)^{n-1}, & y = i \\ 0, & \textit{otherwise} \end{cases}; \tag{4}$$

$$P(\xi_{t+1} = y \mid \xi_t = 1) = \begin{cases} p(1-p)^{n-1}, & y = 0 \\ p(1-p)^{n-1}, & y = 2 \\ 1 - 2p(1-p)^{n-1}, & y = 1 \\ 0, & \textit{otherwise} \end{cases}. \tag{5}$$

Let $\mathbb{E}_{tj}(i)$ denote the CFHT of TrapJump when starting from state i. It is easy to see that $\mathbb{E}_{tj}(0) = 0$, which implies the optimal state. Moreover, we have Lemma 5.

Lemma 5
$\forall i \geq 1 : \mathbb{E}_{tj}(i) \geq \frac{1}{p^i(1-p)^{n-i}}$*, and* $\mathbb{E}_{tj}(i-1) \leq \mathbb{E}_{tj}(i)$.

Proof: We first prove $\forall i \geq 1 : \mathbb{E}_{tj}(i) \geq \frac{1}{p^i(1-p)^{n-i}}$ inductively on i.

(a) Initialization is to prove $\mathbb{E}_{tj}(n) \geq \frac{1}{p^n}$. Since $\mathbb{E}_{tj}(n) = 1 + p^n\mathbb{E}_{tj}(0) + (1-p^n)\mathbb{E}_{tj}(n)$, we have $\mathbb{E}_{tj}(n) = \frac{1}{p^n}$.

(b) Inductive Hypothesis assumes that

$$\forall\ i > K (K \leq n-1) : \mathbb{E}_{tj}(i) \geq \frac{1}{p^i(1-p)^{n-i}}.$$

Then, we consider $i = K$. For $K \geq 2$, by combining Eq.(4) with Lemma 1, we have

$$\begin{aligned} \mathbb{E}_{tj}(K) &= 1 + p^K(1-p)^{n-K}\mathbb{E}_{tj}(0) \\ &\quad + (n-K)p(1-p)^{n-1}\mathbb{E}_{tj}(K+1) \\ &\quad + (1 - p^K(1-p)^{n-K} - (n-K)p(1-p)^{n-1})\mathbb{E}_{tj}(K) \\ &= \frac{1 + (n-K)p(1-p)^{n-1}\mathbb{E}_{tj}(K+1)}{p^K(1-p)^{n-K} + (n-K)p(1-p)^{n-1}} \\ &\geq \frac{p^K(1-p)^{n-K} + (n-K)p(1-p)^{n-1}\frac{1-p}{p}}{p^K(1-p)^{n-K} + (n-K)p(1-p)^{n-1}} \\ &\quad \cdot \frac{1}{p^K(1-p)^{n-K}} \\ &\geq \frac{1}{p^K(1-p)^{n-K}}, \end{aligned}$$

where the first inequality is since $\mathbb{E}_{tj}(K+1) \geq \frac{1}{p^{K+1}(1-p)^{n-K-1}}$ by inductive hypothesis, and the last inequality is by $p \leq 0.5$.

For $K = 1$, by combining Eq.(5) with Lemma 1, we have

$$\begin{aligned} \mathbb{E}_{tj}(1) =& 1 + p(1-p)^{n-1}\mathbb{E}_{tj}(0) + p(1-p)^{n-1}\mathbb{E}_{tj}(2) \\ &+ (1 - 2p(1-p)^{n-1})\mathbb{E}_{tj}(1) \\ =& \frac{1 + p(1-p)^{n-1}\mathbb{E}_{tj}(2)}{2p(1-p)^{n-1}} \\ \geq& \frac{p(1-p)^{n-1} + (1-p)^n}{2p(1-p)^{n-1}} \cdot \frac{1}{p(1-p)^{n-1}} \\ \geq& \frac{1}{p(1-p)^{n-1}}. \end{aligned}$$

According to (a) and (b), we can conclude that

$$\forall i \geq 1 : \mathbb{E}_{tj}(i) \geq \frac{1}{p^i(1-p)^{n-i}}.$$

Then, we are to show that $\forall i \geq 1 : \mathbb{E}_{tj}(i-1) \leq \mathbb{E}_{tj}(i)$. First, it trivially holds that $\mathbb{E}_{tj}(0) = 0 < \mathbb{E}_{tj}(1)$. Then, from the above proof,

$$\begin{aligned} \mathbb{E}_{tj}(1) &= \frac{1 + p(1-p)^{n-1}\mathbb{E}_{tj}(2)}{2p(1-p)^{n-1}} \\ &\leq \frac{p^2(1-p)^{n-2} + p(1-p)^{n-1}}{2p(1-p)^{n-1}} \cdot \mathbb{E}_{tj}(2) \\ &\leq \mathbb{E}_{tj}(2), \end{aligned}$$

and for $i \geq 2$,

$$\begin{aligned} \mathbb{E}_{tj}(i) &= \frac{1 + (n-i)p(1-p)^{n-1}\mathbb{E}_{tj}(i+1)}{p^i(1-p)^{n-i} + (n-i)p(1-p)^{n-1}} \\ &\leq \frac{p^{i+1}(1-p)^{n-i-1} + (n-i)p(1-p)^{n-1}}{p^i(1-p)^{n-i} + (n-i)p(1-p)^{n-1}} \\ &\quad \cdot \mathbb{E}_{tj}(i+1) \\ &\leq \mathbb{E}_{tj}(i+1). \end{aligned}$$

■

Proposition 1
For the process that the (1+1)-EA on the Trap problem with size n under uniform initial distribution, $\mathfrak{A}^l_{SA} = \Omega((\frac{1}{2p(1-p)})^n)$.

Proof: Let the chain $\xi \in \mathcal{X}$ model the analyzed process. Then, $\mathcal{X} = \{0,1\}^n$ and $\mathcal{X}^* = \{0^{n-1}1\}$. Let $\xi' \in \mathcal{Y}$ model the reference process that is the TrapJump chain with size n and parameter p. Then, $\mathcal{Y} = \{0, 1, \dots, n\}$ and $\mathcal{Y}^* = \{0\}$.

We divide $\mathcal{X}$ into $\{\mathcal{X}^*, \mathcal{X}^F_0, \dots, \mathcal{X}^F_{n-1}, \mathcal{X}^I\}$, where $\mathcal{X}^F_i$ contains feasible solutions which have Hamming distance $n-i$ with the optimal solution $0^{n-1}1$, and $\mathcal{X}^I$ contains all the infeasible solutions. That is, $\mathcal{X}^F_i = \{s0 \mid s \in \{0,1\}^{n-1}, |s| = n-1-i\}$ and $\mathcal{X}^I = \{s1 \mid s \in \{0,1\}^{n-1}, |s| > 0\}$, where $|s|$ is the number of 1 bits of s. We construct the mapping function $\phi : \mathcal{X} \to \mathcal{Y}$ that

$$\phi(x) = \begin{cases} 0, & x \in \mathcal{X}^* \\ n-i, & x \in \mathcal{X}^F_i \\ 1, & x \in \mathcal{X}^I \end{cases}.$$

It is easy to see that ϕ is an optimal-aligned mapping because $\phi(x) \in \mathcal{Y}^* = \{0\}$ iff $x \in \mathcal{X}^*$.

Then, we investigate the condition Eq.(1) of switch analysis. We first calculate the left part of Eq.(1). For ξ, any solution $x \in \mathcal{X}_i^F$ can only jump to $\mathcal{X}^* \cup \mathcal{X}_0^F \cup \cdots \cup \mathcal{X}_{i-1}^F$, because the EA only accepts better offspring solutions and

$$f(x \in \mathcal{X}^*) > f(x \in \mathcal{X}_0^F) > \cdots > f(x \in \mathcal{X}_{n-1}^F) = f(x \in \mathcal{X}^I).$$

For any $x \in \mathcal{X}_i^F$, $P(\xi_{t+1} \in \mathcal{X}^* \mid \xi_t = x) = p^{n-i}(1-p)^i$ since it has Hamming distance $n-i$ with the optimal solution; and $P(\xi_{t+1} \in \mathcal{X}_{i-1}^F \mid \xi_t = x) \geq (1-p) \cdot ip(1-p)^{n-2}$ since it is sufficient to keep the last 0 bit unchanged (i.e., keep the solution feasible), flip one of the other i 0 bits and keep the remaining bits unchanged. Thus, for any $x \in \mathcal{X}_i^F$, we have

$$\begin{aligned} &\sum\nolimits_{y \in \mathcal{Y}} P(\xi_{t+1} \in \phi^{-1}(y) \mid \xi_t = x) \mathbb{E}[\![\tau' \mid \xi_0' = y]\!] \\ &\geq p^{n-i}(1-p)^i \mathbb{E}_{tj}(0) + ip(1-p)^{n-1} \mathbb{E}_{tj}(n-i+1) \\ &\quad + (1 - p^{n-i}(1-p)^i - ip(1-p)^{n-1}) \mathbb{E}_{tj}(n-i), \end{aligned}$$

where the inequality is since $\mathbb{E}_{tj}(i)$ increases with i by Lemma 5.
For any $x \in \mathcal{X}^I$, $P(\xi_{t+1} \in \mathcal{X}^* \mid \xi_t = x) \leq p(1-p)^{n-1}$ since the Hamming distance with the optimal solution is at least 1; and $P(\xi_{t+1} \in \mathcal{X}_0^F \cup \cdots \cup \mathcal{F}_{n-2}^F \mid \xi_t = x) \geq p(1-p)^{n-1}$ since it is sufficient to flip the last 1 bit and keep the other bits unchanged. Thus, for any $x \in \mathcal{X}^I$,

$$\begin{aligned} &\sum\nolimits_{y \in \mathcal{Y}} P(\xi_{t+1} \in \phi^{-1}(y) \mid \xi_t = x) \mathbb{E}[\![\tau' \mid \xi_0' = y]\!] \\ &\geq p(1-p)^{n-1} \mathbb{E}_{tj}(0) + p(1-p)^{n-1} \mathbb{E}_{tj}(2) \\ &\quad + (1 - 2p(1-p)^{n-1}) \mathbb{E}_{tj}(1). \end{aligned}$$

We then calculate the right part of Eq.(1). For any $x \in \mathcal{X}_i^F (i < n-1)$, by $\phi(x) = n - i > 1$ and Eq.(4),

$$\begin{aligned} &\sum\nolimits_{y \in \mathcal{Y}} P(\xi_1' = y \mid \xi_0' = \phi(x)) \mathbb{E}[\![\tau' \mid \xi_1' = y]\!] \\ &= p^{n-i}(1-p)^i \mathbb{E}_{tj}(0) + ip(1-p)^{n-1} \mathbb{E}_{tj}(n-i+1) \\ &\quad + (1 - p^{n-i}(1-p)^i - ip(1-p)^{n-1}) \mathbb{E}_{tj}(n-i). \end{aligned}$$

For any $x \in \mathcal{X}_{n-1}^F \cup \mathcal{X}^I$, by $\phi(x) = 1$ and Eq.(5),

$$\begin{aligned} &\sum\nolimits_{y \in \mathcal{Y}} P(\xi_1' = y \mid \xi_0' = \phi(x)) \mathbb{E}[\![\tau' \mid \xi_1' = y]\!] \\ &= p(1-p)^{n-1} \mathbb{E}_{tj}(0) + p(1-p)^{n-1} \mathbb{E}_{tj}(2) \\ &\quad + (1 - 2p(1-p)^{n-1}) \mathbb{E}_{tj}(1). \end{aligned}$$

According to the above calculations, we now know

$$\begin{aligned} &\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \pi_t(x) P(\xi_{t+1} \in \phi^{-1}(y) | \xi_t = x) \mathbb{E}[\![\tau' | \xi_0' = y]\!] \\ &\geq \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \pi_t(x) P(\xi_1' = y | \xi_0' = \phi(x)) \mathbb{E}[\![\tau' | \xi_1' = y]\!], \end{aligned}$$

thus we have found a proper $\rho_t = 0$ in Eq.(1), and therefore,

$$\mathbb{E}[\![\tau | \xi_0 \sim \pi_0]\!] \geq \mathbb{E}[\![\tau' | \xi_0' \sim \pi_0^{\phi}]\!].$$

Then, we investigate $\mathbb{E}[\![\tau' | \xi_0' \sim \pi_0^{\phi}]\!]$. By the uniform initial distribution over $\mathcal{X} = \{0,1\}^n$, $\pi_0(\mathcal{X}^*) = \frac{1}{2^n}$, $\pi_0(\mathcal{X}_i^F) = \frac{\binom{n-1}{n-1-i}}{2^n}$ and $\pi_0(\mathcal{X}^I) = \frac{1}{2} - \frac{1}{2^n}$. Thus,

$$\begin{aligned} &\mathbb{E}[\![\tau' | \xi_0' \sim \pi_0^{\phi}]\!] = \sum\nolimits_{i=0}^{n} \pi_0^{\phi}(i) \mathbb{E}_{tj}(i) \\ &= \pi_0(\mathcal{X}_{n-1}^F \cup \mathcal{X}^I) \mathbb{E}_{tj}(1) + \sum\nolimits_{i=2}^{n} \pi_0(\mathcal{X}_{n-i}^F) \mathbb{E}_{tj}(i) \\ &\geq \frac{1}{2p(1-p)^{n-1}} + \sum_{i=2}^{n} \frac{\binom{n-1}{i-1}}{2^n} \frac{1}{p^i(1-p)^{n-i}} \quad \text{(by Lemma 5)} \\ &= \frac{1}{2^n p^n (1-p)^{n-1}} + \frac{1}{2p(1-p)^{n-1}} - \frac{1}{2^n p(1-p)^{n-1}} \\ &\in \Omega\left(\left(\frac{1}{2p(1-p)}\right)^n\right). \end{aligned}$$

■

In [12], convergence-based analysis is used to derive the EFHT of the (1+1)-EA: when the mutation probability p is a constant in $(0, 0.5]$, $\mathfrak{A}_{CA}^l = \Omega((\frac{1}{1-p})^n)$, and when the mutation probability $p = \frac{1}{n}$, $\mathfrak{A}_{CA}^l = \Omega(2^n)$. By Proposition 1, we know that when p is a constant in $(0, 0.5]$, $\mathfrak{A}_{SA}^l = \Omega((\frac{1}{2p(1-p)})^n)$, and when $p = \frac{1}{n}$, $\mathfrak{A}_{SA}^l = \Omega((\frac{n}{2})^n)$. In both cases of p, switch analysis achieves tighter lower bounds.

V. Discussion

Since the proofs of reducibility for the fitness level method, drift analysis, and convergence-based analysis are constructive, we can find configurations of switch analysis that are equivalent with these approaches. By comparing these configurations, we have an opportunity of comparing these approaches in a unified framework.

In [11], the proofs of reducibility of both the fitness level method and drift analysis employed the OneJump chain in Definition 11 as the reference process.

Definition 11 (OneJump [11])
OneJump chain with state space dimension n is a homogeneous Markov chain $\xi \in \{0,1\}^n$ with $n+1$ parameters $\{p_0, \ldots, p_n\}$ each is in $[0,1]$ and target state 1^n. Its initial state is selected from $\{0,1\}^n$ uniformly at random, and its transition probability is defined as, for any $x \in \{0,1\}^n$ and any t,

$$P(\xi_{t+1} = y \mid \xi_t = x) = \begin{cases} p_{|x|_1}, & y = 1^n \\ 1 - p_{|x|_1}, & y = x \\ 0, & \textit{otherwise} \end{cases},$$

where $|x|_1$ is the number of 1 bits in x.

The OneJump chain is similar with the OneJump-fix chain, as both of them are simply jumping to optimal or staying where they were. In the OneJump chain the transition probability depends on the solution, but the transition probability in the OneJump-fix is a constant. Besides the transition probability, they are different in the state space. The OneJump chain is in a binary state space, while the OneJump-fix chain is in the same state space of the concerning evolutionary process. In [11], for the fitness level method, the solution was partitioned into m subsets, and the state space of the

TABLE I. COMPARISON OF THE ANALYSIS APPROACHES FROM THE UPPERBOUND REDUCTION TO SWITCH ANALYSIS. THE v IS THE INPUT OF THE FITNESS LEVEL METHOD, THE V IS THE DISTANCE FUNCTION OF THE DRIFT ANLAYSIS AND $\mathcal{V}$ IS THE ORDERED SET OF V VALUES (DETAILS IN [12]).

Approach	Reference chain	Transition probability	Aligned mapping
fitness level	OneJump in $\{0,1\}^{m-1}$ with m being the number of levels	$p_i = \frac{1}{\frac{1}{v_{i+1}} + \chi_u \sum_{j=i+2}^{m-1} \frac{1}{v_j}}$	$\phi(x) = 1^{i-1}0^{m-i}$
drift analysis	OneJump in $\{0,1\}^m$ with $m+1$ being the number of distinct distance values	$p_i = \frac{1}{V_{m-i}}$	$\phi(x) = 1^{m-\mathcal{V}_x}0^{\mathcal{V}_x}$
convergence-based	OneJump-fix in $\mathcal{X}$ i.e. the original state space	$p_{fix} = \frac{1}{\sum_{t=0}^{+\infty} \prod_{i=0}^{t-1}(1-\alpha_i)}$	$\phi(x) = x$

reference chain was set to $\{0,1\}^{m-1}$. For drift analysis, $\mathcal{V}$ denoted the set of all distinct values of the distance function V, of which the size was $m+1$, and the state space of the reference chain was set to $\{0,1\}^m$. Moreover, $\mathcal{V}_x$ was used to denote the ordered index of the distance value for the solution x, which is used in the aligned mapping.

Table I lists the major components of the reductions of the analysis approaches to switch analysis. This allow us to make comparisons of these approaches. We can observe a limitation of the fitness level method: the aligned mapping must make some different solutions collapse into one binary string. This causes the method cannot distinguish some solutions that need to be handled differently. The limitation of convergence-based analysis is in the other end: the transition probability is the same for all solutions, and thus can not be configured to handle different solutions differently. This may make the approach less flexible. For drift analysis, when extremely the distance function has a different value on every solution, it can distinguish every solution by the aligned mapping, while it can assign different transition probabilities to every solution. But it also reminds us that the choosing of a good distance function is important for drift analysis yet not straightforward.

Nevertheless, the above discussion on the reducibility and the relationship among different approaches is only about the theoretical ability of these approaches, but not about their "goodness" or other aspects. For example, the hardness of using an analysis approach may also depends on the background knowledge, problem understanding, and personal preference of the analyst.

VI. CONCLUSION

In this paper, we prove that convergence-based analysis [12] is reducible to switch analysis [11]. This is the third approach, following the fitness level method and drift analysis, that have been proven to be reducible to switch analysis. We also apply switch analysis on the Trap problem to achieve a tighter lower EFHT bound than the previous result by convergence-based analysis. Moreover, the fitness level method, drift analysis, and convergence-based analysis are compared, from the perspective of how they are reduced to switch analysis. The comparison reveals some different focuses and limitations of different approaches.

ACKNOWLEDGMENT

We thank the anonymous reviewers for their valuable comments.

REFERENCES

[1] A. Auger and B. Doerr, *Theory of Randomized Search Heuristics - Foundations and Recent Developments*. Singapore: World Scientific, 2011.

[2] B. Hajek, "Hitting-time and occupation-time bounds implied by drift analysis with applications," *Advances in Applied Probability*, vol. 14, no. 3, pp. 502–525, 1982.

[3] J. He and X. Yao, "Drift analysis and average time complexity of evolutionary algorithms," *Artifical Intelligence*, vol. 127, no. 1, pp. 57–85, 2001.

[4] ——, "A study of drift analysis for estimating computation time of evolutionary algorithms," *Natural Computing*, vol. 3, no. 1, pp. 21–35, 2004.

[5] J. He and X. Yu, "Conditions for the convergence of evolutionary algorithms," *Journal of Systems Architecture*, vol. 47, no. 7, pp. 601–612, 2001.

[6] T. Jansen, *Analyzing Evolutionary Algorithms*. Berlin, Germany: Springer-Verlag, 2013.

[7] F. Neumann and C. Witt, *Bioinspired Computation in Combinatorial Optimization - Algorithms and Their Computational Complexity*. Berlin, Germany: Springer-Verlag, 2010.

[8] G. Sasaki and B. Hajek, "The time complexity of maximum matching by simulated annealing," *Journal of the ACM*, vol. 35, no. 2, pp. 387–403, 1988.

[9] D. Sudholt, "A new method for lower bounds on the running time of evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 3, pp. 418–435, 2013.

[10] I. Wegener, "Methods for the analysis of evolutionary algorithms on pseudo-Boolean functions," in *Evolutionary Optimization*, M. M. Ruhul A. Sarker and X. Yao, Eds. Kluwer, 2002.

[11] Y. Yu, C. Qian, and Z.-H. Zhou, "Switch analysis for running time analysis of evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, 2015, DOI:10.1109/TEVC.2014.2378891.

[12] Y. Yu and Z.-H. Zhou, "A new approach to estimating the expected first hitting time of evolutionary algorithms," *Artificial Intelligence*, vol. 172, no. 15, pp. 1809–1832, 2008.