**On the Usefulness of Infeasible Solutions in Evolutionary Search: A Theoretical Study**

Yang Yu, and Zhi-Hua Zhou, *Senior Member, IEEE*

National Key Laboratory for Novel Software Technology
Nanjing University, Nanjing 210093, China

{yuy, zhouzh}@lamda.nju.edu.cn

---

**Abstract**—Evolutionary algorithms (EAs) have been widely used in optimization, where infeasible solutions are often encountered. Some EAs regard infeasible solutions as useless individuals while some utilize infeasible solutions based on heuristic ideas. It is not clear yet that whether infeasible solutions are helpful or not in the evolutionary search. This paper theoretically analyzes that under what conditions infeasible solutions are beneficial. A sufficient condition and a necessary condition are derived and discussed. Then, the paper theoretically shows that the use of infeasible solutions could change the hardness of the task. For example, an EA-hard problem can be transformed to EA-easy by exploiting infeasible solutions. While, the conditions derived in the paper can be used to judge whether to use infeasible solutions or not.

---

**I. INTRODUCTION**

Evolutionary algorithms (EAs) are a kind of optimization technique, inspired by the natural evolution process. Despite many different implementations [1], e.g. Genetic Algorithm, Genetic Programming and Evolutionary Strategies, traditional evolutionary algorithms can be summarized below by four steps:

1. Generate an initial population of random solutions;
2. Reproduce new solutions from the current population;
3. Remove relatively bad solutions in the population;
4. Repeat from Step 2 until the stop criterion is met.

Real-world optimization problems usually come with constraints, and many approaches have been developed to enable EAs to handle constrained optimization problems [2; 15; 10]. Among those approaches, infeasible solutions, i.e. individuals that do not satisfy the constraints, are usually avoided, repaired or penalized [2]. Approaches which avoid infeasible solutions ensure that no infeasible solutions will be generated during the evolutionary process, such as through the homomorphous mapping [9]: Approaches which repair infeasible solutions transform infeasible solutions into feasible ones through some heuristic rules, e.g. a binary coded infeasible solution for knapsack problem can be repaired to be feasible by twice scanning [12]: Approaches which penalize infeasible solutions try to transform the constrained problem into an unconstrained problem by changing the fitness function or the selection function, e.g. death penalty assigns infeasible solutions the worst fitness [2], superior penalty selects the feasible solution in comparison between feasible and infeasible individuals [4], Stochastic Ranking introduces randomness into the penalty of infeasible individuals [15].

While ‘keeping the search space small’ is an intuitive reason to kick infeasible solutions away, it was reported that death penalty approach where infeasible solutions are unlikely to join the evolution process, is worse than ‘softer’ penalty approaches [3; 11] and other methods such as [15] where infeasible solutions have chances to be involved. It was intuitively explained [2] that, when there are many disjointed feasible regions, infeasible solutions would bridge these regions hence might be helpful to the search, and including infeasible solutions would help to explore the boundary between feasible and infeasible regions. It was also noticed that the relationship between an infeasible solution and a feasible region plays a significant role in penalizing such a solution [2; 13]. It was also hypothesized that infeasible solutions can cause ‘shortcuts’ [14], although no significant support found [14]. These findings suggest that infeasible solutions might be useful in evolutionary search. However, theoretically, how infeasible solutions can affect the search and when infeasible solutions are helpful are unknown yet.

Very recently, a few initial works on infeasible solution concerned theoretical analysis have emerged [17; 8]. In [17] the penalty coefficient of a penalty-based approach is studied, and in [8] penalty-based approach is compared with repair-based approach on a knapsack problem whose capacity is restricted to be half of the sum of all weights.

This paper presents a step towards theoretical analysis of the usefulness of infeasible solutions. We theoretically study the problem that when infeasible solutions are helpful. We derive a sufficient condition, under which an EA will reach the optimal solution faster when infeasible individuals are included, and a necessary condition, which specifies the requirement that must be met for infeasible solutions to be helpful. Then, for illustrating the usefulness of the derived conditions, we apply them to two problems to judge that whether exploiting infeasible solutions is helpful or not. The judgements are verified through the estimation of the expected first hitting time of EAs on these two problems. We find that infeasible solutions could significantly affect the performance of EAs, which could be a switch between EA-hard and EA-easy problems, while the conditions derived in this paper can be well useful in judging the helpfulness of infeasible solutions.

The rest of the paper starts with a section that introduces some preliminaries. Then, Section III derives the sufficient and necessary conditions on the usefulness of infeasible solutions. Section IV applies the conditions in a case study.
Section V concludes.

II. PRELIMINARIES

EAs evolve solutions from generation to generation. Each generation stochastically depends on the very pervious one, except the initial generation which is randomly generated. This essential can be modeled by the Markov chain model naturally [7; 16].

The key of constructing a Markov chain that models an EA is to map the populations of the EA to the states of the Markov chain. A popular mapping [7; 16] considers each population as a set and lets a state of Markov chain correspond to a possible population of EAs. However, we can equivalently consider each population as an ordered set, which simplifies the calculation later. Suppose a solution is encoded into a vector with length $L$, each component of the vector is drawn from a space $B$, and each population contains $M$ solutions. So the solution space is $S = B^L$, and the population space $X = S^M$. A Markov chain models the EA is constructed by taking $X$ as the state space, i.e. building a chain $\{\xi_t\}_{t=0}^{\infty}$ where $\xi_t \in X$.

A population is called the optimal population if it contains at least one optimal solution. Let $X^* \subset X$ denotes the set of all optimal populations. The goal of EAs is to reach any population in $X^*$ from an initial population. Thus, the process of an EA seeking $X^*$ can be analyzed by studying the corresponding Markov chain [7; 16].

First, given a Markov chain $\{\xi_t\}_{t=0}^{\infty}$ ($\xi_t \in X$) and a target subspace $X^* \subset X$, let $\mu_t (t = 0, 1, \ldots )$ denotes the probability of $\xi_t$ in $X^*$, that is,

$$\mu_t = \begin{cases} \sum_{x \in X^*} P(\xi_t = x), & S \text{ is a discrete space} \\ \int_{X^*} p(\xi_t = x) dx, & S \text{ is a continuous space} \end{cases}$$

**Definition 1 (Convergence)** Given a Markov chain $\{\xi_t\}_{t=0}^{\infty}$ ($\xi_t \in X$) and a target subspace $X^* \subset X$, $\{\xi_t\}_{t=0}^{\infty}$ is said to converge to $X^*$ if $$\lim_{t \to +\infty} \mu_t = 1$$

**Definition 2 (Absorbing Markov Chain)** Given a Markov chain $\{\xi_t\}_{t=0}^{\infty}$ ($\xi_t \in X$) and a target subspace $X^* \subset X$, $\{\xi_t\}_{t=0}^{\infty}$ is said to be an absorbing chain, if

$$\forall t = 0, 1, \ldots : P(\xi_{t+1} \notin X^* | \xi_t \in X^*) = 0$$

An EA can be modeled by an absorbing Markov chain, if it never loses the optimal solution once been found. Many EAs for real problems can be modeled by an absorbing Markov chain. This is because if the optimal solution to the concerned problem can be identified, then an EA will stop when it finds the optimal solution; while if the optimal solution to the concerned problem cannot be identified, then an EA will keep the best-so-far solution in each generation.

**Definition 3 (Expected first hitting time)** Given a Markov chain $\{\xi_t\}_{t=0}^{\infty}$ ($\xi_t \in X$) and a target subspace $X^* \subset X$, let $\tau$ be a random variable that denotes the events:

- $\tau = 0 : \xi_0 \in X^*$
- $\tau = 1 : \xi_1 \in X^* \land \xi_t \notin X^* (\forall t = 0)$
- $\tau = 2 : \xi_2 \in X^* \land \xi_t \notin X^* (\forall t = 0, 1)$

... then the mathematical expectation of $\tau$, $E[\tau]$, is called the expected first hitting time (EFHT) of the Markov chain.

The expected first hitting time is the average time that EAs find the optimal solution, which is an important property of EAs since it implies the average computational time complexity of EAs [5; 16]. Generally, the smaller the EFHT, the better the EA. General theoretical approaches for estimating the EFHT of general EAs have been developed in [5; 16].

Given an EA, a problem is **EA-Easy** if the EA solves the problem in polynomial time, while a problem is **EA-Hard** if the EA solves the problem in exponential time [5; 6].

We say $f(n) = O(g(n))$ iff $\lim_{n \to +\infty} f(n)/g(n) < +\infty$, $f(n) = \Omega(g(n))$ iff $g(n) = O(f(n))$, and we use $\sim$ to denote asymptotic equal to.

III. SUFFICIENT AND NECESSARY CONDITIONS FOR EXPLOITATION OF INFEASIBLE SOLUTIONS

The solution space $S$ consists of a feasible subspace $S^F$ and an infeasible subspace $S^I$. In $S^F$, there is an optimal solution $x^*$, which is the goal of the optimization. We want to compare between two EAs, $EAF^*$ and $EAI$, $EAF^*$ kicks off infeasible solutions in search, while $EAI$ exploits infeasible solutions.

Note that there are many possible ways to kick off infeasible solutions, thus we characterize the behavior of kicking off by constraint the space that $EAF^*$ searches. Formally, given the populations size $M$, let $X$ be the population space on $S$, $EAF^*$ searches in a subspace $X^F = (S^F)^M (\subset X)$. We denote $X^{F*} = X^F \cap X^*$ as its reachable target region. $EAF^*$ searches in the whole population space $X$, and $X^*$ is its reachable target subspace. We denote $X^I = X - X^F$ as the extra space searched by $EAI$, and $X^{I*}$ as its extra target subspace.

**Theorem 1 (Sufficient Condition)** Let two absorbing Markov chains $\{\xi_t^F\}_{t=0}^{\infty}$ and $\{\xi_t^I\}_{t=0}^{\infty}$ correspond to the processes of $EAF^*$ and $EAI$ respectively. If it satisfies for all $t \geq 0$ that

$$\sum_{x \in X^F - X^{F*}} P(\xi_{t+1} \notin X^F | \xi_t = x) \tilde{p}(\xi_t^F = x) \leq \sum_{x \in X^I - X^{I*}} P(\xi_{t+1} \notin X^I | \xi_t = x) \tilde{p}(\xi_t^I = x) + \sum_{x \in X^I - X^{I*}} P(\xi_{t+1} \notin X^I | \xi_t = x) \tilde{p}(\xi_t^I = x)$$

(1)

we will have $E[\tau^F] \geq E[\tau^I]$ where $\tilde{p}$ is

$$\tilde{p}(\xi_t^F = x) = \frac{P(\xi_t^F = x)}{1 - P(\xi_t^F \in X^{F*})}$$

$$\tilde{p}(\xi_t^I = x) = \frac{P(\xi_t^I = x)}{1 - P(\xi_t^I \in X^*)}$$
Proof. First, denote $\alpha_t$ as
\[
\alpha_t = \sum_{x \in X^F - X^F_t} P(\xi_{t+1}^F \in X^F_t | \xi_t^F = x) \bar{p}(\xi_t^F = x),
\]
and denote $\beta_t$ as
\[
\beta_t = \sum_{x \in X - X^*} P(\xi_{t+1}^F \in X^* | \xi_t^F = x) \bar{p}(\xi_t^F = x).
\]
Assuming the evolution starts from non-optimal populations, the EFHT of $EA^F$ and $EA^I$ can be obtained from $\alpha_t$ and $\beta_t$, respectively [16]:
\[
\begin{align*}
E[\tau^F] &= \alpha_0 + \sum_{t=2}^{+\infty} t \alpha_{t-1} \prod_{i=0}^{t-2} (1 - \alpha_i) \\
E[\tau^I] &= \beta_0 + \sum_{t=2}^{+\infty} t \beta_{t-1} \prod_{i=0}^{t-2} (1 - \beta_i)
\end{align*}
\]
Therefore, Eq.1 derives $\forall t \geq 0 : \alpha_t \leq \beta_t$. Since it is assumed that the evolution starts from a non-optimal population, that is
\[
\begin{align*}
\left\{ \begin{array}{l}
  P(\xi_t^F \in X^F) = 0 \\
  P(\xi_{t+1}^F \in X^F) = P(\xi_t^F \in X^F) + \alpha_t
  \\
  P(\xi_t^I \in X^I) = 0 \\
  P(\xi_{t+1}^I \in X^I) = P(\xi_t^I \in X^I) + \beta_t
\end{array} \right.
\end{align*}
\]
we get $\forall t \geq 0 : P(\xi_t^F \in X^F) \leq P(\xi_t^I \in X^I)$. According to Definition 3, it can be rewritten as
\[
\forall t \geq 0 : P(\tau^F = t) \leq P(\tau^I = t).
\]
Then, according to the Lemma 2 in [16], the inequality $\forall t \geq 0 : P(\tau^F = t) \leq P(\tau^I = t)$ derives
\[
E[\tau^F] \geq E[\tau^I],
\]
which means, by average, $EA^F$ needs more time than $EA^I$ to find the optimal solution. \[ \square \]

**Theorem 2 ( Necessary Condition)** Let two absorbing Markov chains $\{\xi_t^F\}_{t=0}^{+\infty}$ and $\{\xi_t^I\}_{t=0}^{+\infty}$ correspond to the processes of $EA^F$ and $EA^I$ respectively. If we have
\[
\forall t \geq 0 : P(\tau^F = t) \leq P(\tau^I = t)
\]
there must be some $t \geq 0$ such that
\[
\sum_{x \in X^F - X^F_t} P(\xi_{t+1}^F \in X^F_t | \xi_t^F = x) \bar{p}(\xi_t^F = x) \leq \sum_{x \in X^F - X^F_t} P(\xi_{t+1}^I \in X^I_t | \xi_t^I = x) \bar{p}(\xi_t^I = x) + \sum_{x \in X^I - X^I_t} P(\xi_{t+1}^I \in X^I_t | \xi_t^I = x) \bar{p}(\xi_t^I = x)
\]
where $\bar{p}$ is
\[
\begin{align*}
\bar{p}(\xi_t^F = x) &= \frac{P(\xi_t^F = x)}{1 - P(\xi_t^F \in X^F_t)} \\
\bar{p}(\xi_t^I = x) &= \frac{P(\xi_t^I = x)}{1 - P(\xi_t^I \in X^I)}
\end{align*}
\]
The proof to Theorem 2 is not difficult to obtain based on the techniques used in the proof to Theorem 1. So, due to the page limitation, here we do not present the proof.

Theorems 2 and 1 are quite general, let’s consider a more specific situation: $EA^F$ assigns all infeasible solutions the worst fitness, such that these solutions will be killed by the selection process, and the reproduction operator can be expressed by the following property:
\[
\forall s, s' \in S : P_r(s' | s) = \delta(\|s' - s\|)
\]
where $P_r$ is the transition probability from $s$ to $s'$; $\| \cdot \|$ is a distance function and $\delta(\cdot)$ is a monotonic distance-to-probability conversion function that satisfy
\[
\forall a \geq b > 0 : \delta(a) \leq \delta(b) \quad (3)
\]
and
\[
\forall s' \in X : \sum_{x \in X} \prod_{i=1}^{M} \delta(\|x_i' - x_i\|) = 1. \quad (4)
\]
Not that many mutation operators satisfy Eq.3 and Eq.4. Therefore, for any given population $x \in X$:
\[
\forall x' \in X : P(\xi_{t+1}^F = x' | \xi_t^F = x) = \prod_{i=1}^{M} \delta(\|x_i' - x_i\|) \quad (5)
\]
where $\xi_{t+1}^F$ denote the population produced from $\xi_t$ by the reproduction operators, $M$ is the population size. Note that here the population is represented by an ordered set.

According to Eq.5, the probability that any non-optimal population, $x \notin X^*$, transits to be optimal is
\[
P(\xi_{t+1}^F \in X^* | \xi_t^F = x) = \sum_{x' \in X^*} P(\xi_{t+1}^F = x' | \xi_t^F = x) = \sum_{x' \in X^*} P(\xi_{t+1}^F = x' | \xi_t^F = x) = \sum_{x' \in X^*} \prod_{i=1}^{M} \delta(\|x_i' - x_i\|) = 1 - \prod_{i=1}^{M} (1 - \delta(\|x_i^* - x_i\|)),
\]
no matter whether $\xi_t$ here represents $\xi_t^F$ or $\xi_t^I$.

Denote $D(x)$ as the probability that $x$ transits to be optimal, i.e.
\[
D(x) = P(\xi_{t+1}^F \in X^* | \xi_t^F = x) = 1 - \prod_{i=1}^{M} (1 - \delta(\|x_i^* - x_i\|)),
\]
then Eqs.1 and 2 can be rewritten as
\[
\sum_{x \in X^F - X^F_t} D(x)(\bar{p}(\xi_t^F = x) - \bar{p}(\xi_t^I = x)) \leq \sum_{x \in X^I - X^I_t} D(x)\bar{p}(\xi_t^F = x)
\]
Eq.6 can help to understand the intuitive meaning of Theorems 1 and 2. We can regard $D(x)$ as the goodness of population $x$. The $\bar{p}$ in Eqs.1 and 2 is exactly the probability on $x$ given that optimal solution has not been found yet. Therefore, $\bar{p}(\xi_t^F = x) - \bar{p}(\xi_t^I = x)(x \in X^F - X^F_t)$ is the difference of care on feasible solutions, and the left-hand side of Eq.6 means deconcentration from good populations in feasible population space when searching infeasible solutions; while the right-hand side of Eq.6 means benefit from good populations in infeasible population space. Eq.6 expresses that when the loss by the deconcentration is less than the benefit, it is worth exploiting infeasible solutions.
In [14] it hypothesizes that infeasible solutions are useful because they cause 'shortcuts' to good solutions, i.e., passing through the infeasible region, the search path goes from one feasible solution with bad fitness to another with good fitness. Eq.6 shows a revision to this assumption that the goodness of solutions should be measured by $D(x)$, i.e., success probability, instead of fitness.

IV. Case Study

To illustrate the usefulness of the conditions derived in the previous section, here we apply them to analyze that in two concrete problems. The two problems we studied come from a general optimization problem, the Subset Sum Problem.

Definition 4 (Subset Sum Problem) Given $n$ positive integers $W = (w_1, \ldots, w_n)$ and a constant $C$, solve a vector $\mathbf{x}^* = \arg\max_{\mathbf{x} \in \{0,1\}^n} \sum_{i=1}^{n} x_i \cdot w_i$ subject to $\sum_{i=1}^{n} x_i \cdot w_i \leq C$.

Solutions that don’t satisfy the constraint are infeasible solutions. By restricting the weights and the constant, specific Subset Sum problems can be obtained. We study the following two specific Subset Sum problems:

Problem 1: let $w_1, \ldots, w_{n-1} > 1$, $w_n = 1 + 2 \sum_{i=1}^{n-1} w_i$ and $C = w_n$.

Problem 2: let $w_1, \ldots, w_{n-1} > 1$, $w_n = 1 + \sum_{i=1}^{n-1} w_i$ and $C = \sum_{i=1}^{n} w_i$.

The optimal solution for Problem 1 is $x_1^* = (0, 0, \ldots, 0, 1)$, which means that only $w_n$ is included in the subset. For Problem 1, solutions containing more than $w_n$ will be summed above $C$, hence are infeasible solutions. The optimal solution for Problem 2 is $x_2^* = (1, 1, \ldots, 1, 0)$, which means only $w_n$ is excluded from the subset. For Problem 2, the solutions containing $w_n$ are infeasible solutions.

For both Problem 1 and 2, firstly, it holds that

$$\forall x, y \in S_F \exists k \geq 2 \exists z_1, \ldots, z_k \in S_F : z_1 = x$$
$$\land \land \land z_k = y \land (\forall j \in [1, k-1] : \|z_j - z_{j+1}\|_H \leq 1),$$

which means there is no disjunctive feasible regions, thus the heuristic “infeasible solutions can help connect disjunctive feasible regions” [2] does not work. Secondly, it also holds that

$$\forall x \in S_F \lor S_F \exists y \in S_F \lor S_F : \|x - y\|_H = 0,$$

which means every solution is at the edge between the feasible and the infeasible regions, thus the edge is exponentially long and the heuristic “infeasible solutions can help to explore the edge” [2] does not work.

Two $(n + n)$-EAs, EA-1 and EA-2, are employed to solve this problem.

Definition 5 ($(n + n)$-EA) The common settings of the two EAs are:

- Reproduction consists of 1-bit mutation only. Given a solution $x$, the 1-bit mutation randomly chooses an index $i$ and flip $x_i$ (from 1 to 0, or inverse). Reproduce $n$ new solutions from the current population.
- Select the best (larger fitness) $n$ solutions from the $2n$ solutions (current population + new generated solutions), and let it be the next generation.
- Stop until the largest fitness in the population is 0.

EA-1 and EA-2 are different only in the strategy of how to deal with the infeasible solutions.

**EA-1**: In initialization, it drops any infeasible solutions to generate $n$ feasible solutions. Selection uses the function $F_1(x) = -(C - \sum_{i=1}^{n} x_i w_i)$ to evaluate feasible solutions and drops all infeasible solutions.

**EA-2**: In initialization, it does not mind whether the solutions are feasible or infeasible. Selection uses the function $F_2(x) = -(C - \sum_{i=1}^{n} x_i w_i)$ to evaluate all feasible and infeasible solutions.

Note that EA-1 kills all infeasible solutions, while EA-2 lets infeasible solutions compete with feasible solutions.

Then by Theorem 1, we find that exploiting infeasible solutions is desirable for the $(n + n)$-EA to solve Problem 1, as proved in Proposition 1.

**Proposition 1** Solving Problem 1 by the $(n + n)$-EA, exploiting infeasible solutions using fitness function $F_2$ is better that kicking off infeasible solutions, measured by $EFHT$.

**Proof.** It is natural to use Hamming distance, denoted as $\| \cdot \|_H$, to measure the distance of two solutions. For the 1-bit-flipping mutation, it holds (recall the $\delta$ in Eq.3):

$$\forall s, s' \in S : \delta(\|s - s'\|_H) = \begin{cases} 0 & \|s - s'\|_H > 1 \\ 1/n & \|s - s'\|_H = 1 \\ 0 & \|s - s'\|_H = 0 \end{cases}$$

which satisfies Eqs.3 and 4.

Then, for the EA that kicks off infeasible solutions, we have

$$\forall s \in S_{F'} : P(\xi_{t+1} \in X_F | s \in \xi_t) > 0$$
and
$$\forall s \in S_F - S_{F'} : P(\xi_{t+1} \in X_F | s \in \xi_t) = 0,$$

where $S_{F'}$ is the set of feasible solutions that has only one bit different from $s^*$. Let $X_{F'}$ be the set of populations containing at least one solution in $S_{F'}$. So,

$$\sum_{x \in X_{F' \setminus X_F}} \left(1 - \prod_{i=1}^{M} (1 - \delta(\|s^* - x_i\|)) \right) \tilde{p}(\xi_{t+1}^F = x)$$
$$= \sum_{x \in X_{F'}, D(x)} D(x) \tilde{p}(\xi_{t+1}^F = x)$$
And for the EA that exploits infeasible solutions, we have

$$\forall s \in S' : P(\xi_{t+1} \in X_F | s \in \xi_t) > 0$$
and
$$\forall s \in S - S' : P(\xi_{t+1} \in X_F | s \in \xi_t) = 0.$$

1 Due to the page limit, hints of proofs are given instead. Complete proofs will be presented in a longer version.
where $S'$ is the set of solutions that have only one bit different from $s^*$. Note $S' = S'^F \cup S'^I$, let $X'^I$ be the set of populations containing at least one solution in $S'^I$. So,

$\sum_{x \in X'^F - X'^F_*} \left( 1 - \prod_{i=1}^M (1 - \delta(|s^* - x_i|)) \right) \tilde{p}(\xi^f_i = x) + \sum_{x \in X'^I - X'^F_*} \left( 1 - \prod_{i=1}^M (1 - \delta(|s^* - x_i|)) \right) \tilde{p}(\xi^i_i = x)
= \sum_{x \in X'^F} D(x)\tilde{p}(\xi^f_i = x) + \sum_{x \in X'^I} D(x)\tilde{p}(\xi^i_i = x)
$

Since $|S'^F| = 1$ and $|S'^I| = n - 1$, it is easy to get that

$\sum_{x \in X'^F} D(x)\tilde{p}(\xi^f_i = x) < \sum_{x \in X'^I} D(x)\tilde{p}(\xi^i_i = x)
< \sum_{x \in X'^F} D(x)\tilde{p}(\xi^f_i = x) + \sum_{x \in X'^I} D(x)\tilde{p}(\xi^i_i = x)
$

Considering $X'^F = \bigcup_{i=1}^n X'^F_i$ and $X'^I = \bigcup_{i=1}^n X'^I_i$, where $X'^F_i$ and $X'^I_i$ are sets of populations containing $i$ number of solutions from $S'^F$ and $S'^I$, respectively, it holds

$$\forall i \forall x \in X'^F_i : y \in X'^I_i : P(\xi^f_{t+1} \in X'^F_i | \xi^f_i = x) = P(\xi^i_{t+1} \in X^* | \xi^i_i = y)$$

and

$$P(\xi^f_{t+1} \in X'^F_i | \xi^f_i \in X^F - X'^F - X'^F_i) = 0
P(\xi^i_{t+1} \in X'^I_i | \xi^i_i \in X^I - X'^I - X'^I_i) > 0$$

Then it holds

$$\sum_{x \in X'^F} D(x)\tilde{p}(\xi^f_i = x) < \sum_{x \in X'^I} D(x)\tilde{p}(\xi^i_i = x)
< \sum_{x \in X'^F} D(x)\tilde{p}(\xi^f_i = x) + \sum_{x \in X'^I} D(x)\tilde{p}(\xi^i_i = x).$$

By Theorem 1, we have $E[r^F] > E[r^I]$, which means that it is better to exploit infeasible solutions.

By Theorem 2, we find exploiting infeasible solutions is undesirable for the $(n + n)$-EA to solve Problem 2 (proved in Proposition 2).

**Proposition 2** Solving Problem 2 by the $(n + n)$-EA, kicking off infeasible solutions is better than exploiting infeasible solutions using fitness function $F_2$, measured by $EFHT$.

**Proof.** Same to the proof of Proposition 1. Eqs.3 and 4 are satisfied. Using notations $S'^F, S'^I, X'^F$ and $X'^I$ as in the proof of Proposition 1, we have

$$\sum_{x \in X'^F - X'^F_*} \left( 1 - \prod_{i=1}^M (1 - \delta(|s^* - x_i|)) \right) \tilde{p}(\xi^f_i = x)
= \sum_{x \in X'^F} D(x)\tilde{p}(\xi^f_i = x)
$$

and

$$\sum_{x \in X'^F - X'^F_*} \left( 1 - \prod_{i=1}^M (1 - \delta(|s^* - x_i|)) \right) \tilde{p}(\xi^i_i = x)
+ \sum_{x \in X'^I - X'^F_*} \left( 1 - \prod_{i=1}^M (1 - \delta(|s^* - x_i|)) \right) \tilde{p}(\xi^i_i = x)
= \sum_{x \in X'^F} D(x)\tilde{p}(\xi^f_i = x) + \sum_{x \in X'^I} D(x)\tilde{p}(\xi^i_i = x).$$

Since $|S'^F| = n - 1$ and $|S'^I| = 1$, it is easy to get that

$$\sum_{x \in X'^F} D(x)\tilde{p}(\xi^f_i = x)
> \sum_{x \in X'^F} D(x)\tilde{p}(\xi^i_i = x) + \sum_{x \in X'^I} D(x)\tilde{p}(\xi^i_i = x).$$

Denote $S'^I$ and $S_i$ as subspaces of $S'^F$ and $S$, respectively, where solutions have $i$ bits different from the optimal solution. Since for the mutation we used can only generate solutions that have at most one bit different from the current solutions. For any $x$, denote

$$\{x^+ = \{x' \in X \mid \forall i : \delta(|x_i - x'_i|) \in \{0, 1\} \land |x_i - s^*| \leq |x_i - x'_i| \}
\{x^- = \{x' \in X \mid \forall i : \delta(|x_i - x'_i|) \in \{0, 1\} \} - \{x^+\}$$

For kicking off infeasible solutions, $P(\xi^f_{t+1} + |\xi^f_i = x) = 1$, and for exploitation of infeasible solutions, $P(\xi^i_{t+1} \in \{x^+ | \xi^i_i = x) + P(\xi^i_{t+1} \in \{x^- | \xi^i_i = x) = 1$ Therefore, it is easy to get

$$\forall x \in X'^F, y \in X : P(\xi^f_i \in X'^F | \xi^f_i = x) \geq P(\xi^f_i \in X'^F \cup X'^I | \xi^f_i = y)$$

Considering that $X'^F = \bigcup_{i=1}^n X'^F_i$ and $X'^I = \bigcup_{i=1}^n X'^I_i$, where $X'^F_i$ and $X'^I_i$ are sets of populations containing $i$ number of solutions from $S'^F$ and $S'^I$, respectively, it holds

$$\forall i \forall x \in X'^F_i, y \in X'^I_i : P(\xi^f_{t+1} \in X'^F_i | \xi^f_i = x)
= P(\xi^i_{t+1} \in X^* | \xi^i_i = y)$$

Therefore, we have

$$P(\xi^f_{t+1} \in X'^F) = P(\xi^f_i \in X'^F)
- P(\xi^f_i \in X^F | |\xi^f_i \in X'^F)
+ P(\xi^f_i \in X^F | |\xi^f_i \in X'^F - X'^F)$$

is larger than

$$P(\xi^f_i \in X'^I) = P(\xi^i_i \in X'^I)
- P(\xi^i_i \in X^* | |\xi^f_i \in X'^F \cup X'^I)
+ P(\xi^f_i \in X^F_i | \xi^f_i \in X^F - X'^F - X'^F)$$

Then it holds

$$\sum_{x \in X'^F} D(x)\tilde{p}(\xi^f_i = x)
> \sum_{x \in X'^F} D(x)\tilde{p}(\xi^i_i = x) + \sum_{x \in X'^I} D(x)\tilde{p}(\xi^i_i = x).$$
So, Eq.2 is not satisfied. By Theorem 2, we have that the inequality $\mathbb{E}[\tau^F] > \mathbb{E}[\tau^I]$ can not hold, which means that kicking off infeasible solutions is better. \qed

The implication of Propositions 1 and 2 can be verified through the estimation of the EFHT of EA-1 and EA-2. Solving Problem 1 by the two EAs, we have the following proposition.

**Proposition 3** Solving Problem 1, the EFHT of EA-1 is lower bounded as $\mathbb{E}[\tau] = \Omega(2^n/n^2)$, while the EFHT of EA-2 is upper bounded as $\mathbb{E}[\tau] = O(n^2)$.

The proof is sketched as follows. For EA-1, at first, we can prove the mass of probability that transfers from non-optimal populations to optimal populations is $\Omega(2^n/n^2)$ at $t = 0$ and will not increase for $t > 0$. Then the lower bound of the EFHT for EA-1 can be obtained by the Theorem 1 in [16]. For EA-2, we first decompose the population space to $n+1$ subspace according to the fitness of the best solution in the population. Then the EFHT that EA-2 in a subspace moves to any better subspace requires $O(n)$ steps can be obtained by the Theorem 1 in [16]. Finally the sum-up is $O(n^2)$ for the worst case.

Solving Problem 2 by the two EAs, we have the following proposition, which is proved similarly to Proposition 3.

**Proposition 4** Solving Problem 2, the EFHT of EA-1 is upper bounded as $\mathbb{E}[\tau] = O(n^2)$, while the EFHT of EA-2 is lower bounded as $\mathbb{E}[\tau] = \Omega(2^n)$.

Propositions 3 and 4 verify that for Problem 1, infeasible solutions should be involved, which could transfer the EA-hard problem to EA-easy; for Problem 2, infeasible solutions should not be involved, otherwise the EA-easy problem will become EA-hard. These results suggest that infeasible solutions could play a role as a switch between EA-easy and EA-hard. Also, these results confirm the helpfulness of the sufficient and necessary conditions derived in this paper in judging whether to use infeasible solutions or not.

V. CONCLUSION

This paper presents a first theoretical study on the usefulness of infeasible solutions in evolutionary search. We derive a sufficient condition and a necessary condition, and as an illustration, we show that these conditions can be helpful for judging the usefulness of infeasible solutions in concrete problems. Moreover, we find that by exploiting infeasible solutions in the search process, an EA-Hard problem can be transformed to be EA-Easy and the reverse, which might link to the open question “what makes a problem hard for EAs” [6]. Note that the theoretical conditions derived in the paper may be helpful in designing powerful evolutionary algorithms that could exploit infeasible solutions well, which is an issue to be explored in the future.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation of China under Grant No. 60635030 and 60721002.

REFERENCES


