Theoretical Perspective of Complexity of Evolutionary Algorithms Adopting Optimal Mixing

Yu-Fan Tung
Tian-Li Yu

TEIL Technical Report No. 2015001
February, 2015
Theoretical Perspective of Complexity of Evolutionary Algorithms Adopting Optimal Mixing

Yu-Fan Tung, Tian-Li Yu
National Taiwan University
r02921044@ntu.edu.tw, tianliyu@cc.ee.ntu.edu.tw

Abstract

The optimal mixing evolutionary algorithms (OMEAs) have recently drawn much attention for their robustness, small size of required population, and efficiency in terms of number of function evaluations (NFE). In this paper, the performances and behaviors of OMEAs are studied by investigating the mechanism of optimal mixing (OM), the variation operator in OMEAs, under two scenarios—one-layer and two-layer masks. For the case of one-layer masks, the required population size is derived from the viewpoint of initial supply, while the convergence time is derived by analyzing the progress of sub-solution growth. NFE is then asymptotically bounded with rational probability by estimating the probability of performing evaluations. For the case of two-layer masks, empirical results indicate that the required population size is proportional to both the degree of cross competition and the results from the one-layer-mask case. The derived models also indicate that population sizing is decided by initial supply when disjoint masks are adopted, that the high selection pressure imposed by OM makes the composition of sub-problems impact little on NFE, and that the population size requirement for two-layer masks increases with the reverse-growth probability.

1 Introduction

The importance of convergence analysis of evolutionary algorithms cannot be overemphasized. Without theoretical support, the empirical findings cannot be generalized, and the development of new mechanism lacks direction. However, due to the stochastic nature, the analyses of evolutionary algorithms (EAs) are difficult, and hence are either inaccurate and incomplete or come much later than the debut of the algorithm. For example, the concept of the simple GA was proposed around the late 1960s. The accurate facet-wise models for the population sizing and the convergence time were not proposed until the late 1990s (Goldberg, Deb, & Clark, 1991; Thierens & Goldberg, 1993; Harik, Cantú-Paz, Goldberg, & Miller, 1999; Goldberg, Sastry, & Latoza, 2001). Similarly, the technique of model building in estimation of distribution algorithms (EDAs) has been developed since the late 1990s, and its analysis came in the filed about ten years later (Pelikan, Sastry, & Goldberg, 2002; Yu, Sastry, Goldberg, & Pelikan, 2007).

Recently, the optimal mixing operator (OM), proposed by Thierens and Bosman, has drawn much attention for its robustness and performance (Bosman & Thierens, 2012; Thierens & Bosman, 2013). Utilizing the intermediate results during variation, mixing with OM is considered noise-free decision making, which greatly reduces the required population size and boosts the performance. The performance of OM significantly depends on the linkage sets (Bosman & Thierens, 2012; Goldman & Tauritz, 2012). Recent researches involve using the CP index (Wang, Tung, & Yu, 2014) and filtering/combining linkage hierarchies (Bosman & Thierens, 2013) to decide whether a given mask is promising during optimization. However, those techniques have been developed with little theoretic support.

The goal of this paper is to analyze the convergence complexity of EAs adopting OM, also called OMEAs, when various linkage models are adopted. In particular, we focus on deriving analytical models for the required population size, the convergence time, and the number of function evaluations (NFE). Many variants of OMEA are proposed, and one of them called GOMEA most suits our study since we use the probabilistic model.
For the rest of this paper, the background is first briefly introduced in Section 2. In the scenario of one-layer masks, complexity models are proposed in Section 3. The results are then extended to two-layer masks in Section 4, followed by conclusion.

2 Background

In this section, background knowledge of this paper is provided. Since the behavior of OMEAs depends on the linkage sets, we first formally define the family of subsets and the associated notations. Using the notations, the test problems used in this paper are then described. We then give a brief introduction to GOMEA, which is the research scenario of this paper. Finally, since the concept of initial supply is highly related to our work, it is also addressed.

2.1 Family of Subsets

In OMEAs, variables are mixed according to linkage sets to ensure adequate mixing while preventing loss of promising partial solutions. The family of subsets (Thierens, 2010), denoted by $\mathcal{F}$, is a general model of linkage sets. Before defining the FOS we introduce the necessary notations. We denote the length of chromosome by $\ell$, and the population size by $n$. Variables in one chromosome, also called genes, can be expressed in a vector $\vec{x} = \langle x_1, \ldots, x_\ell \rangle$. The notation $|\vec{x}|$ denotes the number of elements in $\vec{x}$. The natural number set is denoted by $\mathbb{N}$. Using these notations the FOS is defined as follows.

**Definition 1** For any integer $\ell \in \mathbb{N}$, $S_\ell$ is an index set consists of all integers from 1 to $\ell$. A subset of the index set is a mask.

$$S_\ell = \{i \in \mathbb{N} | 1 \leq i \leq \ell\}.$$

**Definition 2** For an index set $S_\ell$, $\mathcal{F}$ is a family of subsets (FOS) with the following two properties.

1. $\mathcal{F}$ is an ordered set consists of masks.

   $$\mathcal{F} = \langle F^1, F^2, \ldots, F^{\mathcal{|F|}} \rangle,$$

   where $F^i \subseteq S_\ell$, $1 \leq i \leq \mathcal{|F|}$.

2. Every element in $S_\ell$ must be contained in at least one of the masks in $\mathcal{F}$.

   $$\forall i \in S_\ell, \exists F^j \in \mathcal{F} : i \in F^j.$$

Take $\ell = 3$ as an example. The index set $S_3 = \{1, 2, 3\}$. Sets such as $\{1\}, \{2, 3\}$, and $\{1, 2, 3\}$ are all masks of $S_3$. $\langle \{1, 2\}, \{2, 3\}\rangle$, $\langle \{2, 3\}, \{1, 2\}\rangle$, and $\langle \{1, 2\}, \{2, 3\}, \{1, 2\}\rangle$ are three different FOSs of $S_3$. $\langle \{1, 2\}, \{2\}\rangle$ is not a valid FOS of $S_3$, since no mask contains 3. Note that the masks in FOS are in order, and there may be duplicated masks.

During OM, variables are mixed according to the elements in FOS. The operation can be viewed as a variable-wise mask operation, so elements in FOS are called masks. For a chromosome $C$, the variables selected according to a mask $M$ are denoted by $C_M$.

Some special FOSs are further defined as follows.

**Definition 3** For an index set $S_\ell$ and an integer $k$ dividing $\ell$, a homogeneous FOS $\mathcal{F}_k$ is defined as

$$\mathcal{F}_k = \langle F^1, F^2, \ldots, F^{\ell/k} \rangle,$$

where $F^i = \{\pi_j | (i-1) \cdot k < j \leq i \cdot k\}$.

$\bar{\pi} = \langle \pi_1, \pi_2, \ldots, \pi_\ell \rangle$ can be any permutation of the index set $S_\ell$. In other words, $\mathcal{F}_k$ is an FOS consisting of $\ell/k$ disjoint masks, and every mask contains $k$ indexes. Since the locations of variables should not affect the behavior of optimization, we use the permutation notation to eliminate the dependency of variable locations.
Definition 4 The concatenated homogeneous FOS can be defined below.

\[ F_{k_1, \ldots, k_a} = F_{k_1} \| F_{k_2} \| \cdots \| F_{k_a}, \]

where \( \| \) is the concatenation operator, which concatenates the elements in the sets while preserving the order.

Take the case \( \ell = 6 \) as an example. \( \langle \{1, 5\}, \{2, 4\}, \{3, 6\} \rangle, \langle \{2, 4, 5\}, \{1, 3, 6\} \rangle, \) and \( \langle \{1, 2, 3, 4, 5, 6\} \rangle \) are three valid homogeneous FOSs. They can be denoted by \( F_2, F_3, \) and \( F_6 \) respectively. A valid concatenation of them is

\[ F_{3,6,2} = \langle \{2, 4, 5\}, \{1, 3, 6\}, \{1, 2, 3, 4, 5, 6\}, \{1, 5\}, \{2, 4\}, \{3, 6\} \rangle. \]

2.2 Test Problems

The analyses and experiments are based on three maximization problems. By successful optimization we means that the global optimum is found. The first problem is onemax:

\[ f_{\text{onemax}}(\vec{x}) = u(\vec{x}), \]

where \( u(\vec{x}) \) is the number of 1s in \( \vec{x} \). Onemax is trivial, but optimization performance on onemax indicates how well an algorithm deals with fully decomposable problems.

The second problem is one instance of the Royal Road functions (Mitchell, Forrest, & Holland, 1992), the structure of which can be characterized by an FOS \( F \):

\[ f_{\text{royal}}(\vec{x}) = \sum_{\forall F_i \in F} R(\vec{x}_{F_i}), \tag{1} \]

where

\[ R(\vec{x}) = \begin{cases} 1 & \text{if } u(\vec{x}) = |\vec{x}|, \\ 0 & \text{otherwise}. \end{cases} \]

Note that \( \vec{x}_F \) is the part of \( \vec{x} \) indicated by a mask \( F \).

The third problem is the deceptive-trap problem (Deb & Goldberg, 1992), which can also be characterized by an FOS \( F \):

\[ f_{\text{trap}}(\vec{x}) = \sum_{\forall F_i \in F} T(\vec{x}_{F_i}), \]

where

\[ T(\vec{x}) = \begin{cases} 1 & \text{if } u(\vec{x}) = |\vec{x}|, \\ 0.9 \cdot \frac{|\vec{x}| - u(\vec{x})}{|\vec{x}| - 1} & \text{otherwise}. \end{cases} \]

The deceptive-trap function is also commonly used for benchmark and is known for its deceptive nature. Local hill climbing does not lead to the global optimum when solving trap problems.

Note that although these problems are all binary-encoded, our results can be applied to \( \chi \)-ary problems.

2.3 GOMEA

GOMEA is one of the first proposed OMEAs and can be integrated with arbitrary FOS. In this paper, we focus on the variation operator of GOMEA, which is named gene-pool optimal mixing (GOM). The pseudocode of GOM can be expressed in Algorithm 1. Utilizing this operator, the pseudocode of GOMEA can be expressed in Algorithm 2. The population is denoted by \( P = \{P_1, \ldots, P_n\} \), and the offspring are denoted by \( \{O_1, \ldots, O_n\} \). Optimization halts if the population converges to one instance. In this paper, all experiments are done with an implementation of GOMEA. For the experiments that need to determine the minimal required population size, a bisection procedure (Sastry, 2001) is adopted.

In practice, linkage detection techniques are applied to generate FOS. In our scenario, this step is omitted. We manually determine the FOS before optimization to prevent unnecessary noise. We also disable the function of mutation in all of our experiments, which is the same as the scenario in the canonical GOMEA.
Algorithm 1: Genepool Optimal Mixing

**Input**: $f$: fitness function, $R$: receiver, $P$: population, $F$: FOS

**Output**: $O$: offspring

$O \leftarrow R$

$B \leftarrow R$

for $i = 1$ to $|F|$ do

$r \leftarrow$ random number from 1 to $|P|$

$D \leftarrow P_r$

$B_F \leftarrow D_F$

if $f(B) \geq f(O)$ then

$O_F \leftarrow B_F$

else

$B_F \leftarrow O_F$

end if

end for

return $O$

Algorithm 2: GOMEA

**Input**: $f$: fitness function, $n$: population size, $F$: FOS

**Output**: optimization solution

randomly generate $n$ instances for $P$

while ¬ SHOULD TERMINATE($P$) do

for $i = 1$ to $n$ do

$O_i \leftarrow$ GenepoolOptimalMixing($f$, $P_i$, $P$, $F$)

end for

for $i = 1$ to $n$ do

$P_i \leftarrow O_i$

end for

end while

return the best instance in $P$

2.4 Population Sizing

In genetic algorithms (GAs), a population-sizing model was proposed based on the concept of ensuring an adequate supply of partial solutions. GAs find the global optimum by mixing the segments of it, which must exist in the initial population. Consider a $\chi$-ary problem which is decomposed into $m$ parts, and each of them contains $k$ variables. Since GAs are stochastic processes, we can never guarantee the existence of all partial solutions. Instead, we consider a tolerable probability $\alpha$ of not having a partial solution in the population. Goldberg et al. derived a minimal required population size in this scenario (Goldberg, Sastry, & Latoza, 2001):

$$n = \chi^k (k \log \chi - \log \alpha).$$

To further simplify this equation, a reasonable assumption is that $\alpha$ is inversely proportional to $m$. In other words, substitute $\alpha$ with $1/m$. Hence the required population size can be expressed in

$$n = \chi^k (k \log \chi + \log m).$$  \hspace{1cm} (2)

There are other population-sizing models for GAs. Goldberg et al. proposed one by considering that larger population leads to higher probability of making correct decisions (Goldberg, Deb, & Clark, 1991). Harik et al. proposed another model by making an analogy between selection in GAs and the gambler’s ruin problem (Harik, Cantú-Paz, Goldberg, & Miller, 1999). These models give different estimations of the required population size. In this paper, we find out that the supply model is a good estimator for OMEAs.
3 One-Layer Masks

To begin our study, we first focus on the scenario so-called one-layer masks where FOS consists of disjoint masks. We assume that the problem can be decomposed into independent subproblems according to the FOS. In most applications, the problem may not be decomposable, and the problem structure is unknown. Various techniques were proposed to detect the problem structure. In this paper, we focus on the mixing operator rather than linkage detection mechanisms, so we just assume the structure is already perfectly detected. We also make two reasonable assumptions. The global optimum is unique, and the population is randomly initialized. In this scenario, we proposed the population-sizing model and the convergence-time model. Combination of these two models leads to the NFE model.

3.1 Population Sizing

The required population size can be estimated from the viewpoint of initial supply. For a set contains $k$ variables which are $\chi$-ary, in correspondence to a mask with $k$ indexes, the optimal state for a variable set is unique, namely being part of the global optimum. A set is called a correct one if it is optimal, and is considered incorrect if not. The probability that the set is correct after uniformly random initialization is $\chi^{-k}$. Since the variables are exchanged using the corresponding mask, they are always tied together during mixing. Therefore no new pattern is created after mixing.

Since the masks reflects the decomposed subproblems, correct sets are never replaced, and eventually dominate the population. Hence the population converges to the optimal state for the mask if and only if at least one correct set exists initially. For a mask of size $k$, the probability that at least one set of initial variables is correct among the population is

$$p_{\text{correct}} = 1 - (1 - \chi^{-k})^n.$$  

EA finds the global optimum if and only if all sets are in optimal state, the probability of which is

$$p_{\text{success}} = \prod_{\forall F_i \in \mathcal{F}} \left(1 - (1 - \chi^{-|F_i|})^n\right).$$  \hspace{1cm} (3)

If the tolerable failure rate of EA is $\alpha$, solving $p_{\text{success}} = 1 - \alpha$ yields the required population size $n$.

A special case of Equation (3) is that all masks are equal-sized, which means $|F_i| = k$, $i = 1, 2, \ldots, m$. Therefore the equation can be rewritten as

$$1 - \alpha = p = \left(1 - (1 - \chi^{-k})^n\right)^m.$$  \hspace{1cm} (4)

Thus the minimal required population $n$ can be expressed in a closed form:

$$n = \frac{\ln \left(1 - (1 - \alpha)^\frac{1}{m}\right)}{\ln (1 - \chi^{-k})}.$$  \hspace{1cm} (5)

From this equation the complexity can be derived. Since $m$ is a positive integer, we have $0 < \frac{1}{m} \leq 1$. Let $f(t) = (1 - \alpha)^t$, where $t \in (0, 1]$. Since $f(t)$ is exponential, it is convex for $t \in \mathbb{R}$. Two inequalities can be derived:

$$f(t) \geq f(0) + tf'(0) = 1 - \ln \frac{1}{1 - \alpha} \cdot t,$$

and

$$f(t) \leq (1 - t) f(0) + tf(1) = 1 - \alpha t.$$  

Combining these inequalities and the fact that the natural logarithm is strictly increasing yields

$$1 - \frac{\ln \frac{1}{1 - \alpha}}{m} \leq (1 - \alpha)^\frac{1}{m} \leq 1 - \alpha \frac{1}{m}. $$
Figure 1: Success rate for various population size with $\ell = 500$ and $k = 5$. Experiments are repeated $10^4$ times. The maximum absolute error is 1.03%.

\[
\ln \left( \frac{1}{m} \right) \geq \ln \left( 1 - (1 - \alpha)^{\frac{1}{m}} \right) \geq \ln \left( \frac{\alpha}{m} \right),
\]

\[
\ln \left( \frac{\alpha}{m} \right) \geq \ln \left( 1 - (1 - \alpha)^{\frac{1}{m}} \right) \geq \ln \left( \frac{c_2}{m} \right),
\]

where $c_1$ and $c_2$ are some constants.

To compare the optimization cost of problems with different sizes, we assume that the problems have similar structures, which implies they are encoded in the same way ($\chi$ is constant), and the adopted masks are similar ($k$ is constant). The tolerable failure rate $\alpha$ should not vary with the problem size. By fixing $\alpha$, $\chi$, and $k$, asymptotically tight bound for Equation 5 with regarding to $m$ can be derived:

\[
\Theta (n) = \Theta \left( \ln \left( 1 - (1 - \alpha)^{\frac{1}{m}} \right) \right) = \Theta (\ln (m)).
\]

The model is verified empirically in Figure 1 where the theoretical values are given by Equation 4. The results show that the successful rate only depends on the supply model. Therefore the population size only depends on the initial supply instead of other aspects, such as decision-making.

3.2 Convergence Time

The study of convergence time involves two parts. The case of a single mask is first studied, and the result of which is then extended to two-layer masks.

For the one-mask case, denote the mask size by $k$ and the generation number by $t$, starting with $t = 0$. After each generation, $t$ is increased by 1. Define $p_t$ as the proportion of correct sets among all chromosomes. $p_0$ can be approximated by its expected value, $\chi^{-k}$. We estimate the convergence time by modeling the growth of $p_t$.

In GOMEA, two candidates are selected as the donor and receiver respectively, and the former tries to donate part of the bit pattern to the latter. The receiver only takes the change if its fitness does not decrease, so a correct set never accepts an incorrect one. On the contrary, incorrect receiver becomes correct once the donor is correct. Since the expected number of incorrect chromosomes in generation $t$ is $n (1 - p_t)$, the expected gain in the number of correct sets after one generation is $n (1 - p_t) \cdot P (\text{donor is correct}) = np_t (1 - p_t)$. Note that the change are made on the offspring instead of receiver, so $p_t$ remains constant until the end of the current generation. Thus we get the iterative equation:

\[
n p_{t+1} - np_t = n \cdot p_t \cdot (1 - p_t).
\]

For $t \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, denote $1 - p_t$ by $q_t$, and we have

\[
q_t - q_{t+1} = (1 - q_t) q_t,
\]

\[
q_t^2 = q_{t+1}.
\]
Along with \( q_0 = 1 - p_0 = 1 - \chi^{-k} \), we get
\[
p_t = 1 - q_t = 1 - \left(1 - \chi^{-k}\right)^{2^t}, \ t \in \mathbb{N}_0.
\] (7)

Although the solution to Equation (4) is derived, there is no \( t \) satisfying \( p_t = 1 \), because we approximate \( p_t \) by its expected value. Since \( p_t \) represents the proportion of correct instances in the population with \( n \) candidates, the greatest value of \( p_t \) less than 1 is \( 1 - 1/n \). Hence the estimation of convergence time is decomposed into two parts. First, we estimate the time \( t_1 \), the required time for \( p_t \) to grow to \( 1 - 1/n \). Second, we estimate the time \( t_2 \), which means how long it takes for the remaining incorrect chromosome to converge to the correct one. Hence the estimation of convergence time equals to \( t_1 + t_2 \).

To estimate \( t_1 \), we approximate \( p_t \) by a continuous function \( p(t) \), which takes \( t \) as a continuous variable. From Equation (7) we set
\[
p(t) = 1 - q(t) = 1 - \left(1 - \chi^{-k}\right)^{2^t}, \ t \geq 0.
\] (8)
Note that \( p(t) = p_t \) for \( t \in \mathbb{N}_0 \). Solving \( p(t_1) = 1 - 1/n \) yields
\[
t_1 = \lg \left(\frac{\ln \frac{1}{n}}{\ln (1 - \chi^{-k})}\right),
\] (9)
where \( \lg \) is the binary logarithm.

When only one incorrect set is left, it becomes correct in the next generation if and only if the donor is another set rather than itself. Therefore after one generation, the status remains unchanged with a probability \( 1/n \). Estimating \( t_2 \) by its expected value yields
\[
t_2 = 1 + \frac{1}{n} \cdot t_2,
\]
\[
t_2 = \frac{n}{n - 1}.
\] (10)
Combining \( t_1 \) and \( t_2 \) yields
\[
t_{\text{conv}} = t_1 + t_2 = \lg \left(\frac{\ln \frac{1}{n}}{\ln (1 - \chi^{-k})}\right) + \frac{n}{n - 1}.
\] (11)

The result can be further extended to multiple disjoint masks. Since the problem is decomposable, convergence in one position does not interfere with that in another position. Convergence time is thus dominated by the variable set with the longest convergence time.

Assume the chromosome consists of \( m \) sets of variables, convergence time of which are \( t_{\text{conv}, 1}, \ldots, t_{\text{conv}, m} \) respectively. We have \( t_{\text{conv}} = \max \{t_{\text{conv}, i}\} \). By inspecting Equation (11) smaller initial proportion leads to longer convergence time. For the case that each set contains \( k \) variables, we can approximate the maximum of \( t_{\text{conv}, i} \) by finding the convergence time of the set with least initial correct instances. Let \( X_1, \ldots, X_m \) be \( m \) independent and identically distributed random variables following the binomial distribution \( B(n, \chi^{-k}) \). Denote the first order statistic by \( X_{(1)} = \min_i (X_i) \). Let \( E[X_{(1)}] = x_{(1)} \). Hence for multiple masks we have
\[
t_{\text{conv}} = t_1 + t_2 = \lg \left(\frac{\ln \frac{1}{n}}{\ln (1 - \frac{x_{(1)}}{n})}\right) + \frac{n}{n - 1}.
\] (12)

The approximation we derived is verified with the following experiments. Figures 2 shows the case of one mask, where the theoretical values are from Equation (11).

The case of multiple disjoint masks is shown in Figures 3 and 4, where theoretical results are from Equation (12).

We find out that the convergence time decreases while \( n \) increases when \( n \) is small. This is mainly because of the \( \frac{1}{n} \) term in Equation (12). According to the central limit theorem, the first order statistics of the proportion of correct instances increases and approaches \( \chi^{-k} \) when \( n \) grows. When \( n \) is large, the \( \frac{1}{n} \) term dominates the estimation, making it increase when \( n \) grows. This phenomenon is verified in Figure 4.

\* \( B(n, p) \) denotes the distribution with the pmf \( P(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \quad x \in \{0, 1, \ldots, n\}, \)
otherwise.
Figure 2: Convergence time with one mask of size 5. For population size ranging from 100 to 25600, the maximum relative error is 4.99%; the maximum absolute error is 0.42 generations.

Figure 3: Convergence time for $n = 500$. For problem size ranging from 100 to 3200, the maximum relative error is 3.21%; the maximum absolute error is 0.35 generations.

Figure 4: Convergence time for $\ell = 100$. For population size ranging from 200 to 102400, the maximum relative error is 4.22%; the maximum absolute error is 0.38 generations. Note that for very large $n$, the absolute error is still less than half a generation.
3.3 Number of Function Evaluations

The function evaluations of OMEAs consist of two parts. The first is those during the initialization, and the second is those during OM. Therefore the total NFE is

\[ n_{fe} = n + n_{fe,OM}. \]  

We then model \( n_{fe,OM} \) for the case of a single size-\( k \) mask.

In generation \( t \), denote the proportion of correct solutions among the population by \( p_t \). Under the assumption that there is only one optimal solution, there are remaining \( \chi^k - 1 \) possible solutions which are not optimal. Denote each of their densities in the population by \( q_t^{(1)}, \ldots, q_t^{(\chi^k-1)} \), respectively. In this generation, a receiver only needs to be evaluated when it differs from the donor. Therefore the probability that a receiver needs the evaluation is

\[ e_t = 1 - p_t^2 - \left( \frac{\sum_{i=1}^{\chi^k-1} q_t^{(i)}}{\chi^k - 1} \right)^2 \geq e_t \geq 1 - p_t^2 - (1 - p_t)^2. \]  

We already have the approximation of \( p_t \) in Equation 7. Since \( q_t^{(1)}, \ldots, q_t^{(\chi^k-1)} \in [0, 1] \), we have

\[ \sum_{i=1}^{\chi^k-1} q_t^{(i)} = 1 - p_t, \]

Equation 14 leads to

\[ 1 - p_t^2 - \frac{(1 - p_t)^2}{\chi^k - 1} \leq e_t \leq 1 - p_t^2 - (1 - p_t)^2. \]  

The required part of NFE can be approximated by

\[ n_{fe,OM} = \sum_{t=0}^{\infty} n \cdot e_t = n \sum_{t=0}^{\infty} e_t. \]

By expanding the summation, we can derive the upper and lower bound of \( n_{fe} \). For the upper bound, there is no closed form. Since \( \sum_{t=0}^{\infty} \left( 1 - p_t^2 - \frac{(1 - p_t)^2}{\chi^k - 1} \right) \) is a function of \( k \) if \( p_t \) is approximated by Equation 7 we denote it by a function \( U(k) \), where \( U(k) \) is a function from \( \mathbb{N} \) to \( \mathbb{R} \). \( \mathbb{N} \) is the natural number set, and \( \mathbb{R} \) is the real number set. Here the variable \( \chi \) is considered constant. So we have

\[ n_{fe,OM} \leq n \cdot U(k). \]

There is a closed form of the lower bound:

\[ \sum_{t=0}^{\infty} \left( 1 - p_t^2 - (1 - p_t)^2 \right) = \sum_{t=0}^{\infty} (2q_t - 2q_t^2) = \sum_{t=0}^{\infty} (2q_t - 2q_{t+1}) = 2q_0. \]

This yields

\[ n_{fe,OM} \geq n \cdot 2q_0 = n \cdot 2 \left( 1 - \chi^{-k} \right) = n \cdot L(k), \]

where \( L(k) = 2 \left( 1 - \chi^{-k} \right) \).

Hence the NFE can be bounded by

\[ n \cdot U(k) \geq n_{fe,OM} \geq n \cdot L(k). \]
Figure 5: \( n_{fe} \) with different sizes of masks. Each FOS contains 100 masks, and the population size is fixed at \( 10^4 \). Note that the Royal Road, the onemax, and the trap function are identical for \( k = 1 \).

The results can be extended to arbitrary number of masks, \( m \). Since all of our masks are disjoint, evaluations of them are independent to each other. So the total NFE is the summation of all required NFES for each mask. Therefore we get

\[
n \cdot \sum_{\forall F^i \in F} U(|F^i|) \geq n_{fe, OM} \geq n \cdot \sum_{\forall F^i \in F} L(|F^i|),
\]

or equivalently,

\[
n + n \cdot \sum_{\forall F^i \in F} U(|F^i|) \geq n_{fe} \geq n + n \cdot \sum_{\forall F^i \in F} L(|F^i|).
\]

For the special case that the sizes of all masks are equal, we have

\[
n \cdot (1 + mU(k)) \geq n_{fe} \geq n \cdot (1 + mL(k)),
\]

where \( k \) is the size of every mask, and \( m \) is the number of masks.

The problem nature affects the required number of evaluations. In Equation 17 the equality of upper bound holds when \( q_1^{(1)} = q_2^{(2)} = \cdots = q_t^{(k-1)} \).

Consider one of our test problems, the Royal Road function. When optimizing this function, one cannot tell two sets apart by fitness if none of them is optimal. Therefore the proportions of non-optimal instances are roughly equal. Since the equality of upper bound holds, this problem costs more function evaluations than others do.

The results are verified in following experiments. Figure 5 shows NFE with masks of various sizes. Figure 6 shows the result with various population sizes. The theoretical values are from Equation 17.

In these two experiments, we have made some observations. While all the experiment values are bounded by the lower bound, values of the Royal Road function slightly exceed the upper bound. In Figure 6 NFE exceeds the upper bound by 2.17% when \( n = 200 \) and by 0.02% when \( n = 12800 \). We infer that the reason is the assumption of sufficient population size. The estimation of \( p_t \) is derived from the expected value while \( p_t \) is discrete. For small \( n \), the approximation might be inaccurate. This inference is verified in Figure 7 which indicates that smaller population makes the population converge more slowly, implying additional consumption in function evaluations. This explains why NFE is slightly underestimated, especially when \( n \) is small.

We also notice that for the same FOS, NFE of the Royal Road problem is always the largest, and NFE of the trap problem is always the smallest. This verifies that the subproblem affects the NFE, yet the difference is subtle. The upper bound seems a good estimator for the NFE regardless of the problems, with relative error less than 5% in our results. We infer that the high mask-wise selection pressure of OM quickly filters out the incorrect sets, and therefore the convergence behavior is not much affected from the subproblem structures.
Figure 6: $n_{fe}$ with different sizes of population. $m = 100$ and $k = 5$ for the test problems.

Figure 7: $p_t$ for various population sizes. The problem is a 500-bit onemax problem, with masks of size 5. We can see smaller $n$ leads to slower convergence.
4 Two-layer Masks

In this section, we extend our study to two-layer masks as the first step to multi-layer masks, which are often adopted in OMEAs. Specifically, the FOS in the form $F_{k,1}$ with $k > 1$ is adopted. Here we focus on the onemax problem.

The population converges to the global optimum if and only if all the variables converge to 1s. For an index $i$, consider all the variables $x_i$ in every chromosomes. There are $n$ variables with this index. We use the notation $p$ to denote the proportion of 1s in these $n$ variables.

Since masks in a homogeneous FOS do not share common variables, exactly one mask in $F_k$ and one mask in $F_1$ contain $x_i$. When mixing with one-bit masks, 1s never change to 0s, but this is not the case of $k$-bit masks. Since the fitness is evaluated when all $k$ variables are exchanged, the overall fitness never decreases, but some variables can be ruined. For example, the pattern 101 overwrites 010 when they are mixed with a 3-bit mask, since the fitness increases by 1. However, the second variable in 010, which is one, is overwritten by a zero. For a certain index, if all $n$ variables with the index in the population become 0s, the optimization fails.

This phenomenon can be further quantified from the viewpoint of probability. By the assumption of random initialization, $p = 2^{-1}$ in binary encoded problems. For a set of $k$ variables covered by a $k$-bit mask, the number of 1s in the set can be expressed in a random variable. If the first variable is zero, the random variable $X_0$ follows the Binomial distribution $B(k - 1, 2^{-1})$. Recall that $k$ is an integer larger than one. If the first variable is given one, then the random variable is $X_1 = 1 + X_0$. Consider a variable $x_i$ equals 0 in the donor and equals 1 in the receiver. $x_i$ in the receiver becomes 0 with probability $P(X_1 \leq X_0)$, which we called the reverse-growth probability, $p_{rg}$. The probabilities are calculated using the Binomial distribution and shown in Table 1. We can see larger $k$ leads to larger $p_{rg}$.

The required population size varies when adopting various FOSs. The concept of cross competition explains this phenomenon (Goldberg, Deb, & Thierens, 1992). Cross competition is first introduced to derive an upper bound for the selection pressure in GA. An upper bound of the selection pressure is derived to prevent cross-competitive failure:

$$s < n \frac{\ln (1 - p_0)}{\ln \alpha},$$

where $s$ is the selection pressure, $p_0$ is the probability that a correct bit is preserved into the next generation, and $\alpha$ is a threshold, which means optimization is likely to fail when $\alpha$ variables are incorrectly converged. After some modification, we obtain

$$n > \frac{s \cdot \ln \alpha}{\ln (1 - p_0)}. \quad (18)$$

Although in the scenario of GOMEA, the situation is different, but the similar concept applies. Larger $k$ leads to larger $p_{rg}$, which then leads to lower probability that a correct bit is preserved. In other words, larger $k$ means smaller $p_0$. By assuming that $s$ and $\alpha$ do not vary much with $k$, the RHS of Equation (18) increases when $k$ increases. This means that the lower bound of population size increases. In other words, the required population increases if $k$ increases when adopting $F_{k,1}$.

In the following experiments our inference is verified. For various lengths of problems, different FOSs are adopted to solve the onemax problem. These FOSs are $F_{5,1}, F_{4,1}, F_{3,1}, F_{2,1}$, and $F_1$. The result is shown in Figure 5. It indicates that optimization using $F_{k,1}$ with larger $k$ requires larger population. Furthermore, a proportional relationship exists between the required population size for $F_{k,1}$ and $F_1$, while the latter can be approximated by Equation (18). From this observation,

<table>
<thead>
<tr>
<th>$k$</th>
<th>$P(X_1 &gt; X_0)$</th>
<th>$P(X_1 = X_0)$</th>
<th>$P(X_1 &lt; X_0)$</th>
<th>$p_{rg}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>75%</td>
<td>25%</td>
<td>0%</td>
<td>25%</td>
</tr>
<tr>
<td>3</td>
<td>69%</td>
<td>25%</td>
<td>6%</td>
<td>31%</td>
</tr>
<tr>
<td>4</td>
<td>66%</td>
<td>23%</td>
<td>11%</td>
<td>34%</td>
</tr>
<tr>
<td>5</td>
<td>64%</td>
<td>22%</td>
<td>14%</td>
<td>36%</td>
</tr>
</tbody>
</table>

Table 1: Comparison of $X_1$ and $X_0$.
Figure 8: Required population size for various FOS. Theoretical values of $F_{k,1}$ are estimated by curve fitting as multiple of that of $F_1$. The maximum relative error among all estimations are 2.0%.

the theoretical values of required population for $F_{k,1}$ are obtained by fitting the experiment values to a constant multiple of that of $F_1$. According to the results, masks which cover multiple separable subproblems are not beneficial from the viewpoint of population size.

5 Conclusion

In this paper, we derived the convergence models of OMEAs for one-layer and two-layers masks. For problems with separable structures, behaviors of GOMEA with masks corresponding to the structures are analyzed. The required population size was accurately estimated from the viewpoint of initial supply. Analyzing the growth of sub-solution led to the convergence-time model. The NFE was then estimated. These three models were verified empirically, and the values of relative error among all experiments are less than 5%. For multi-layer disjoint masks, a special case with two-layer masks was studied. We found that the concept of cross competition explains the growth in the required population sizes.

As for future work, we would like to extend our result to multi-layer masks by quantifying the effect of cross competition and by investigating the growth of subsolutions iteratively. Hopefully, we would be able to analyze the behavior of OMEAs with full linkage tree, which is a special case of multi-layer FOS.

The major contributions of this paper reside in the derivations of the population-sizing, the convergence-time, and the NFE models for OMEAs with one-layer and two-layer masks. Our models are empirically verified and the relative errors of our estimators are small. In addition, our models lead to the following insightful findings. First, for the case of one-layer masks, the required population size is decided by initial supply rather than decision-making. This explains why OMEAs generally require relatively small populations compared to EDAs. Second, NFE for the test problems is very close to the proposed upper bound, by which we infer that the mask-wise selection pressure of OM quickly filter out non-optimum subsolutions, making the subproblem composition insignificant. Third, for two-layer masks, the required population size is proportional to that of the one-layer masks, and the ratio is positively related to the reverse-growth probability.

References


