

An upper bound of the mutation probability in the genetic algorithm for general 0-1 knapsack problem

Yang Yang*

School of Mathematical Sciences, Xiamen University, Xiamen 361005, P.R. China

Abstract

As an important part of genetic algorithms (GAs), mutation operators is widely used in evolutionary algorithms to solve \mathcal{NP} -hard problems because it can increase the population diversity of individual. Due to limitations in mathematical tools, the mutation probability of the mutation operator is primarily empirically set in practical applications.

In this paper, we propose a novel reduction method for the 0-1 knapsack problem(0-1 KP) and an improved mutation operator (IMO) based on the assumption $\mathcal{NP} \neq \mathcal{P}$, along with the utilization of linear relaxation techniques and a recent result by Dey et al. (Math. Prog., pp 569-587, 2022). We employ this method to calculate an upper bound of the mutation probability in general instances of the 0-1 KP, and construct an instance where the mutation probability does not tend towards 0 as the problem size increases. Finally, we prove that the probability of the IMO hitting the optimal solution within only a single iteration in large-scale instances is superior to that of the traditional mutation operator.

Keywords: genetic algorithm; mutation operator; mutation probability; upper bound; reduction method

1 Introduction

Definition 1. [1, 2, 3, 4] Given a set N consisting of n items, where the j -th item has a profit and a weight, and denoted by p_j and w_j , respectively. The goal of the 0-1 knapsack

*Email: zhugemutian@outlook.com

problem(0-1 KP) is to find an optimal subset of the set N such that its total profit is maximized without exceeding the knapsack capacity C . The problem can be formulated as follows.

$$\max f(\mathbf{X}) = \sum_{j=1}^n x_j p_j, \quad (1.1)$$

s.t.

$$g(\mathbf{X}) = \sum_{j=1}^n x_j w_j \leq C, \quad (1.2)$$

$$x_j \in \{0, 1\}, \quad (1.3)$$

where $x_j = 1$ represents the j -th item is packed in the knapsack while $x_j = 0$ indicates not. The objective function (1.1) of the model aims to maximize the total profit of the selected items. Inequality (1.2) is a capacity constraint that ensures the total weight of selected items do not exceed the knapsack capacity C . The decision variables x_j are binary, as indicated in (1.3).

A substantial number of commercial decision-making and industrial manufacturing problems can be formulated as the 0-1 KP, such as budget control and cargo loading. This has resulted in a significant focus on developing efficient algorithms to tackle these problems. However, the finding for a polynomial-time algorithm to exactly solve the 0-1 KP, which is widely recognized as an \mathcal{NP} -hard problem[5, 6, 7], poses a substantial challenge. Existing algorithms like dynamic programming [8, 9] can only solve the problem exactly within a pseudo-polynomial time complexity of $O(nC)$, and even the fastest exact algorithms require $O(2^{n/2})$ [10, 11]. Additionally, a significant body of research indicates that unless one intends to demonstrate $\mathcal{NP} = \mathcal{P}$, we can assume $\mathcal{NP} \neq \mathcal{P}$ [12, 13]. If $\mathcal{NP} \neq \mathcal{P}$, for a general instance of the 0-1 KP, there are items whose selection in the optimal solution cannot be determined without exhaustively enumerating all feasible solutions.

Coincidentally, in the 1960s, Holland proposed a genetic algorithm(GA) [14] that simulates biological evolution, consisting of selection, crossover, and mutation operators, primarily used to search the binary solution space. Genetic algorithms(GAs) do not require the solution space to be continuous or need to calculate the gradient of the objective function during iteration, and as a result, the algorithm has been widely applied to solving various \mathcal{NP} -hard problems, such as the task scheduling problem [15], the Traveling Salesman Problem [16], and the Vehicle Routing Problems [17].

It is worth noting that the convergence analysis of GAs is primarily limited to convex problems or \mathcal{P} problems [18, 19, 20, 21, 22, 23]. However, there is a lack of theoretical

results for \mathcal{NP} -hard problems, especially for the optimal parameter control of operators such as the mutation operator (MO)[24, 25, 26]. Despite the widespread and efficient performance of GAs in practical industrial and application processes[27], there is still widespread skepticism among researchers regarding GA research[28, 29, 30, ?]. Moreover, the No Free Lunch (NFL) theorem states that there is no superior parameterization for every problem[27, 31, 32, 33, 34], but it is still important to determine if there exists an optimal parameterization for a given 0-1 KP instance.

1.1 Our results

In this paper, we build upon the assumption of $\mathcal{NP} \neq \mathcal{P}$ [5, 6, 7], improve the recent theoretical results of the branch and bound algorithm (B&B) [35, 36], integrating generating function approaches[37, 38], introduce a new reduction method and an improved mutation operator (IMO). Simultaneously, the method is applicable in computing an upper bound of the mutation probability for the 0-1 KP, as well as in constructing a counterexample where the mutation probability does not tend toward 0 with an increasing number of decision variables [25]. The contributions of this paper can be summarized as follows.

1. We propose an improved reduction method for the 0-1 KP, and provide the upper bound of leaves in the search tree (Theorem 3.6).
2. We propose the IMO and provide a general mathematical formulation for the upper bound of the mutation probability in solving the 0-1 KP (Theorem 3.7).
3. We construct examples where the mutation probability of the IMO does not tend towards 0 with the increase in problem size (Theorem 3.9).
4. We demonstrate that for large-scale instances, the performance of the IMO is superior to that of the MO (Theorem 3.11).

1.2 Organization of remainder paper

The remainder of this paper is organized as follows. In Section 2, we introduce the preliminary and relevant concepts of GAs. Section 3 presents the theoretical results, including a novel reduction method, the IMO proposed from this reduction method, and the associated theoretical findings. Additionally, we provide a convergence analysis of both the IMO and the MO. Finally, in the last section, we summarize the work of this paper and discuss future research directions.

2 Preliminary and existing conclusions

2.1 Preliminary

From the objective function (1.1) and inequality (1.2), a primitive greedy algorithm can be easily identified: the likelihood of an item appearing in the optimal solution increases as its profit increases and its weight decreases. Consequently, we introduce the classic concept of profit density[1, 2] to characterize the return on profit per unit weight of the items.

Definition 2. For $j \in N$, the profit density of item j is defined by

$$e_j = p_j/w_j. \quad (2.4)$$

Without loss of generality, we always assume that the items are arranged in non-increasing order, denoted as $e_1 \geq e_2 \geq \dots \geq e_n$. The greedy algorithm selects items according to their profit density in descending order until the first item that cannot be packed in the knapsack. The item is known as the *break item* b and the solution obtained employing this greedy algorithm is the break solution. The characteristics of the break item and the break solution are defined as follows.

Definition 3. The break item b is defined to be the first item that cannot be packed in the knapsack with profit density descending order, that is,

$$\sum_{j=1}^{b-1} w_j \leq C \quad \text{and} \quad \sum_{j=1}^b w_j > C \quad (2.5)$$

Correspondingly, the residual capacity r is defined as

$$r = C - \sum_{j=1}^{b-1} w_j \quad (2.6)$$

Definition 4. The break solution is denoted as $\mathbf{X}^* = (x_1^*, x_2^*, \dots, x_n^*) \in \{0, 1\}^n$, where $x_j^* = 1$ if $j \in \{1, 2, \dots, b-1\}$ and $x_j^* = 0$ otherwise. For the sake of comparison, let \mathbf{Y} be the optimal solution.

To further verify that items with higher profit density are more likely to be selected in the optimal solution, Pisinger[39] conducted a comparison between the optimal solution \mathbf{Y} and the break solution \mathbf{X}^* for 1000 instances. Each instance had a data size of 1000, with the break item being the 500th. Pisinger found that the primary concentrations of differences between \mathbf{Y} and \mathbf{X}^* were around the break item.

Based on empirical evidence derived from the greedy algorithm and a significant number of instances, it can be empirically observed that items with higher profit density tend to be more frequently selected for inclusion in the optimal solution \mathbf{Y} . However, it is regrettable that this observation may be challenged by a classic counterexample, which can be shown as follows.

Example 2.1. [2] Let $n = 2$ and $C = M$. The first item is given by $w_1 = 1$ and $p_1 = 2$ whereas the second item is defined by $w_2 = p_2 = M$, where M is a large constant.

Clearly, the break solution is $\mathbf{X}^* = (1, 0)$, but the optimal solution for the counterexample is $\mathbf{Y} = (0, 1)$. From this counterexample, we can observe that the profit density of an item does not necessarily determine whether it will be selected in the optimal solution, and also implies that in the process of solving through GA, the optimal probability distribution for the MO cannot be determined for a given instance.

Although it is currently impossible to determine all decision variables in polynomial time complexity, with the continuous development of solving algorithms, a reduction algorithm has been discovered that can exactly solve a part of decision variables in polynomial time complexity. Given an instance \mathcal{P} , the reduction algorithm treats the solving process as a search tree, where each branch represents a decision variable in the problem. We employ the notation $(\mathcal{P}|x_j = \alpha)$ to denote the subproblem where the j -th decision variable is fixed at $\alpha \in \{0, 1\}$. To facilitate subsequent discussions, $\bar{v}(\mathcal{P}|x_j = \alpha)$ is introduced as an upper bound for this subproblem. Furthermore, we introduce $\underline{v}(\mathcal{P})$ as a lower bound for the original problem \mathcal{P} , typically computed using the greedy algorithm or a meta-heuristic algorithm.

Despite the subproblem $(\mathcal{P}|x_j = \alpha)$ is also known to be \mathcal{NP} -hard, relaxing the solution space to the real field converts it into a convex problem[40]. Consequently, the optimal solution for the relaxed subproblem, $\bar{v}(\mathcal{P}|x_j = \alpha)$, can be obtained within polynomial time complexity. These findings lead to several significant conclusions.

Theorem 2.2. For the 0-1 KP, if $\bar{v}(\mathcal{P}|x_j = \alpha) \leq \underline{v}(\mathcal{P})$, then x_j is set to $1 - \alpha$ in the optimal solution, where $\alpha \in \{0, 1\}$.

The decision variable x_j is fixed, and the branch $x_j = \alpha$ is pruned. To compute $\bar{v}(\mathcal{P}|x_j = \alpha)$, the most common approach is to relax the solution space from the integer field to the real field. By doing this, we can obtain the well-known Dantzig upper bound[41], which is denoted as U and is given by

$$U = \sum_{j=1}^{b-1} p_j + \frac{rp_b}{w},$$

where r represents the residual capacity. It is clear that the decision variables dominated by the Dantzig upper bound U can be expressed as follows:

Theorem 2.3. [39] *For the j -th item, if $e_j \geq e_b$ (resp. $e_j < e_b$) and $p_j > (w_j + r)p_b/w_b$ (resp. $p_j < (w_j - r)p_b/w_b$), then we have $x_j = 1$ (resp. $x_j = 0$) in the optimal solution.*

The time complexity of computing the Dantzig bound is $O(n)$. It is evident that we can solve a decision variable exactly within polynomial time complexity. A significant amount of research has been conducted to improve the upper bounds obtained within the time complexity of $O(n^2)$ [39, 42, 43, 44, 45], thereby enabling the exact fixing of more decision variables within polynomial time complexity. However, in the worst-case scenario, the method cannot solve all decision variables. Consequently, for a preprocessed \mathcal{NP} -hard problem, to the best of our knowledge, there is currently no relevant research available regarding which decision variables are more likely to be selected in the optimal solution. Moreover, if an \mathcal{NP} -hard problem can be exactly solved within polynomial time complexity, heuristic algorithms, such as GAs, that can only provide locally optimal solutions within polynomial time complexity lose their significance. Based on the current state of research, we make the following assumptions.

Assumption 1. $\mathcal{NP} \neq \mathcal{P}$.

If Assumption 1 holds, then for an unpruned feasible solution, we cannot determine its status as an exact solution within polynomial time complexity. Consequently, each unpruned feasible solution has an equal probability of being the optimal solution.

2.2 Genetic algorithm Framework

With the prosperity of commercial activities, a vast number of business operations can be described as \mathcal{NP} -hard problems. Considering Assumption 1, the development of exact algorithms for solving such problems has been slow. Meanwhile, heuristic algorithms have flourished [50, 51]. As a classic algorithm in the field of heuristics, GAs have been extensively applied to solve \mathcal{NP} -hard problems, particularly the knapsack problem and its variants [46], over the past decades. In GAs, each solution in the search space can be represented by an individual composed of several genes. By simulating natural selection, genetic mutation, and crossover operations, GAs evolve and select individuals within the current population, aiming to find the optimal solution. The flowchart of the GA can be illustrated in Figure 1. It is worth noting that Rudolph pointed out, through Markov chain analysis, that the GA lacks global convergence, whereas the Elitist Genetic

Algorithm (EGA), which selects the current population and the optimal individual from the previous generation, possesses the property of converging to the global optimum as time approaches infinity [22]. The main distinction between EGA and GA lies in the selection operators, while the same mutation operator is employed. Since this paper primarily focuses on the theoretical results of mutation operators, we will still adopt the representation of the GA in this study.

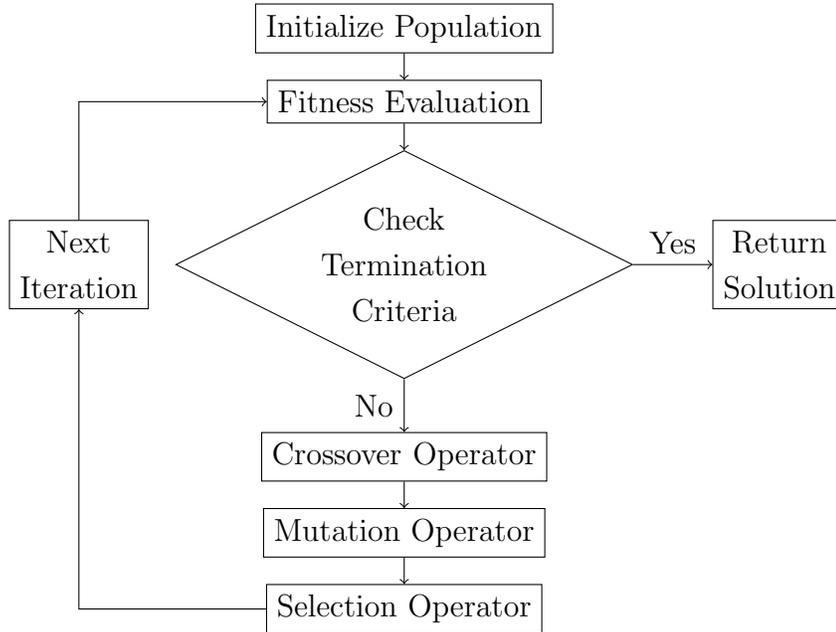


Figure 1: Flowchart of a basic genetic algorithm

The GA begins by initializing the population and generating initial solutions, and the fitness value of each individual is calculated by the objective function. The algorithm then checks if the termination condition is satisfied based on the specified number of iterations. If the termination condition is met, the results are output; otherwise, the next generation population is obtained by applying the crossover, mutation, and selection operators sequentially. Common crossover operators include single-point crossover, k-point crossover, and uniform crossover. Frequent mutation operators include flip bit mutation, swap mutation, and inversion mutation. Selection operators mainly consist of roulette wheel selection, stochastic universal sampling, and rank-based selection. Various works on these operators are available in the references [47, 48]. In this study, we utilize the widely employed single-point crossover, flip bit mutation, and roulette wheel selection operators, along with presenting pseudo-code for the GA.

Let $O_t^k(j)$ represent the value of the j -th decision variable in the k -th individual of the t -th ($1 \leq k \leq \text{pop}$) generation where pop represents the population size, and let $\min f(O_t)$

(resp. $\max f(O_t)$) denote the minimum (resp. maximum) fitness value of individuals in the t -th generation population. Additionally, LOS refers to the local optimal solution, while I and t represent the maximum number of iterations and the current iteration number, respectively. Moreover, p_m and p_c represent the mutation probability and the crossover probability, respectively.

Algorithm 1 Genetic algorithm

INPUT: $n, W, P, C, \text{pop}, I, p_c, p_m$

OUTPUT: LOS

Generating initial population O_1 and let $t = 1$.

while $t \leq I$ **do**

 //Crossover Operator

for $k = 1 : 2 : \text{pop}$ **do**

if $\text{rand} \leq p_c$ **then**

$r_t = \text{rand} * (n - 1) + 1, T = O_t^k,$

$O_t^k(r_t + 1 : n) = O_t^{k+1}(r_t + 1 : n),$

$O_t^{k+1}(r_t + 1 : n) = T(r_t + 1 : n).$

end if

end for

 //Mutation Operator

$O_t = |O_t - [\text{rand}(\text{pop}, n) \leq p_m]|.$

 //Selection Operator

 The selection probability of the k -th individual is determined as

$\frac{f(O_t^k) - \min f(O_t) + 1}{\sum_{k'=1}^{\text{pop}} f(O_t^{k'}) - \text{pop} \min f(O_t) + \text{pop}},$ and resulting in $O_{t+1}.$

$t = t + 1.$

end while

LOS = $\max f(O_I)$

The GA starts by generating initial solutions and then enters a loop. Next, the algorithm performs the crossover and mutation operators on the individuals determined by random numbers. These operations involve randomly generating crossover points, denoted as t_r , crossing two individuals from position t_r to n , and flipping their bits. Afterward, the GA calculates the probability of each individual being selected using the roulette wheel selection operator. Particularly, as problem size increases, the difference between the objective function values of sub-optimal and local optimal solutions decreases, increasing the likelihood of sub-optimal solutions being selected. To address this, we assign a value of 1 to the objective function value of the worst solution in the current generation, then subtract this value from the objective function values of the other solutions, and

add 1. Finally, the algorithm terminates the loop after a limited number of iterations and returns the local optimal solution (LOS).

It is worth noting that GAs were initially perceived as a robust search algorithm, leading to a lack of extensive attention to parameter control in the early stages of its development [22]. As the research on algorithm theory progressed, however, numerous researchers acknowledged the impact of parameter control on search outcomes, and it has since become one of the most crucial issues in the field of evolutionary algorithms [49].

3 Mainly theoretical result

3.1 A novel reduction method

Recently, a new reduction method was proposed by Dey et al.[36] for analyzing the performance of the B&B in solving random integer linear programming problems. This method divides the decision variables into different regions and constrains the branching generated by decision variables within each region. In this section, we combine traditional reduction methods [39] with the approach by Dey et al. [36] to solve the 0-1 KP. We introduce a colored regions partitioning method for 0-1 KP, as demonstrated in Theorem 3.1. This method determines the minimum or maximum number of items that should be included in the optimal solution for each colored region.

Theorem 3.1. *Given a set of items N' where the j -th item satisfies the inequality*

$$\frac{p_j}{w_j + r/i} > \frac{p_b}{w_b} \tag{3.7}$$

for $i \in \mathbb{N}$, we can conclude that $\sum_{j \in N'} |x_j - y_j| \leq i - 1$, where \mathbf{Y} represents the optimal solution.

It is evident that for the 0-1 KP, the number of leaves in the search tree for items in N' is reduced from $2^{|N'|}$ to $\sum_{k=1}^{i-1} \binom{|N'|}{k}$. Finding ways to effectively reduce the algorithmic complexity and improve the solving speed, particularly by exploring more efficient prune strategies between different colored regions, remains an intriguing question. Regrettably, existing research has not investigated the relationships between items within different colored regions. Therefore, we present the following definition based on Theorem 3.1.

Definition 5. *The item set $N_1 = \{1, 2, \dots, b - 1\}$, consisting of items with a higher profit density than the break item b , can be divided into m disjoint subsets, denoted as*

$N_1 = \sqcup_{i=1}^m N_{1i}$. Each N_{1i} subset is defined as follows:

$$N_{1i} = \{j \mid \lfloor \frac{rp_b}{p_j w_b - p_b w_j} \rfloor + 1 = i, e_j > e_b\}. \quad (3.8)$$

If no item satisfies the formula (3.8) for any integer $i(1 \leq i \leq m)$, then $N_{1i} = \emptyset$.

Theorem 3.2. Let D_{1i} and s_{1i} denote the item set and the number of items that are not packed into the optimal solution in N_{1i} , respectively, where $1 \leq i \leq m$. Furthermore, we have $s_{1i} = |D_{1i}| = \sum_{j \in N_{1i}} |y_j - 1|$. All s_{1i} form the vector $S = (s_{1i})_{i=1}^m$. Therefore, we can conclude that:

$$\sum_{i=1}^m \frac{s_{1i}}{i} \leq 1.$$

Proof. Given a vector \mathbf{Y}' , we can determine the corresponding D'_{1i} and $S' = (s'_{1i})_{i=1}^m$ of \mathbf{Y}' for $i = 1, 2, \dots, m$ using the definitions of D_{1i} and S .

Assume, for the sake of contradiction, that \mathbf{Y}' is the optimal solution and satisfies

$$\sum_{i=1}^m s'_{1i}/i > 1.$$

It is clear that the optimal solution \mathbf{Y}' must not exceed the Dantzig bound. Therefore, we have

$$\sum_{i=1}^m \sum_{j \in D'_{1i}} p_j \leq \left(\sum_{i=1}^m \sum_{j \in D'_{1i}} w_j + r \right) p_b / w_b. \quad (3.9)$$

Using the formulas (3.8) and Theorem 3.1, we can deduce the following inequality:

$$\frac{\sum_{i=1}^m \sum_{j \in D'_{1i}} p_j}{\sum_{i=1}^m \sum_{j \in D'_{1i}} (w_j + r/i)} = \frac{\sum_{i=1}^m \sum_{j \in D'_{1i}} p_j}{\sum_{i=1}^m \sum_{j \in D'_{1i}} w_j + r \sum_{i=1}^m s'_{1i}/i} > \frac{p_b}{w_b}$$

and

$$\begin{aligned} \sum_{i=1}^m \sum_{j \in D'_{1i}} p_j &> \left(\sum_{i=1}^m \sum_{j \in D'_{1i}} w_j + r \sum_{i=1}^m s'_{1i}/i \right) p_b / w_b \\ &> \left(\sum_{i=1}^m \sum_{j \in D'_{1i}} w_j + r \right) p_b / w_b. \end{aligned} \quad (3.10)$$

Inequality (3.9) and inequality (3.10) contradict each other, proving the proposition. Hence, Theorem 3.2 is established. \square

It is worth noting that s_{1i} represents the items in N_{1i} that are not selected by the optimal solution, so s_{1i} belongs to the set of natural numbers, i.e., $s_{1i} \in \mathbb{N}$. Even when the integer constraints of the original problem are relaxed, i.e., $x_j \in [0, 1]$ and $s_{1i} \in \mathbb{R}$, the conclusion still holds. This means that

Corollary 3.3. *Given a solution $\mathbf{X} \in [0, 1]^n$ and $s_{1i} = \sum_{j \in D_{1i}} |x_j - 1|$. If*

$$\sum_{i=1}^m s_{1i}/i > 1, \quad (3.11)$$

then \mathbf{X} is not the optimal solution.

Furthermore, we establish the following equivalent corollary to facilitate algorithmic computation, considering the presence of numerous empty sets N_{1i} in Theorem 3.2.

Corollary 3.4. *Consider a vector $H = \{h_1, h_2, \dots, h_n\}$, where each element is defined as:*

$$h_j = \begin{cases} \lfloor \frac{rp_b}{p_j w_b - p_b w_j} \rfloor + 1, & \text{if } e_j > e_b, \\ \infty, & \text{otherwise.} \end{cases}$$

If \mathbf{X} is the optimal solution, then the following inequality holds:

$$\sum_{j=1}^n \frac{1 - x_j}{h_j} \leq 1.$$

Based on Corollary 3.4, we can derive the following corollary for items with a profit density lower than the break item b .

Corollary 3.5. *Consider a vector $L = \{l_1, l_2, \dots, l_n\}$, where each element is defined as:*

$$l_j = \begin{cases} \lfloor \frac{rp_b}{p_b w_j - p_j w_b} \rfloor + 1, & \text{if } e_j < e_b, \\ \infty, & \text{otherwise.} \end{cases}$$

If \mathbf{X} is the optimal solution, then the following inequality holds:

$$\sum_{j=1}^n \frac{x_j}{l_j} \leq 1.$$

The number of leaves pruned in the search tree of the 0-1 KP by Corollaries 3.4 and 3.5 can be expressed using generating functions[37, 38] as follows.

Theorem 3.6. Consider the number of leaves in the search tree, denoted as ω , and the number of items in N_{1i} , denoted as n_{1i} , respectively. Additionally, let $\mathcal{X}(\lambda)$ be the polynomial representing a potential optimal solution, given by:

$$\mathcal{X}(\lambda) := \prod_{i=1}^m \left(\sum_{j=0}^{\min\{n_{1i}, i\}} \binom{n_{1i}}{j} \lambda^{j/i} \right). \quad (3.12)$$

Then, the number of leaves ω is equal to the sum of the coefficients of terms with an exponent no greater than 1.

It is evident from Theorem 3.6 that Corollary 3.4 and Corollary 3.5 significantly enhance the efficiency of B&B algorithms in solving problems such as the 0-1 KP.

3.2 An improvement mutation operator and the upper bound of the mutation probability

As one of the most commonly used tools to enhance population diversity in various heuristic algorithms, the flip bit mutation operator has been widely applied due to its simplicity. Specifically, for $O_t^k(j)$, the flip bit mutation compares the mutation probability p_m (typically set as $p_m \in [0.001, 0.01]$) with a randomly generated number. If the generated random number is less than p_m , the value of $O_t^k(j)$ is flipped, changing 1 to 0 and 0 to 1. This procedure effectively increases the diversity of the population, enabling better exploration of the search space. To facilitate subsequent discussions, we first provide the mathematical representation of the flip bit mutation.

$$O_t^k(j) = \begin{cases} 1 - O_t^k(j), & \text{if rand} < p_m, \\ O_t^k(j), & \text{otherwise,} \end{cases} \quad (3.13)$$

where rand is a randomly generated number within the interval $[0, 1]$.

From Theorem 3.6, it becomes evident that items positioned farther from the line formed by the break item and the origin are more likely to be selected in the optimal solution. Building upon this observation and combining it with a greedy algorithm, we propose the IMO. Given a mutation probability p_m , the IMO initially employs the break item to partition the set of items into two parts. Mutation operations are then performed on items belonging to different parts based on the value of $O_t^k(j)$. The mathematical representation of the IMO can be described using formula (3.14).

For a given mutation probability p_m , the IMO yields an expectation of $p_m n_{1i}$ for the number of unselected items in each set N_{1i} . However, due to the constraint imposed by

$$O_t^k(j) = \begin{cases} 1 - O_t^k(j), & \text{if } e_j > e_b \text{ and } \text{rand} < O_t^k(j)p_m + (1 - O_t^k(j))(1 - p_m), \\ 1 - O_t^k(j), & \text{if } e_j \leq e_b \text{ and } \text{rand} < (1 - O_t^k(j))p_m + O_t^k(j)(1 - p_m), \\ O_t^k(j), & \text{otherwise.} \end{cases} \quad (3.14)$$

Corollary 3.3, there is a maximum limit on the number of unselected items among all sets. Solutions that do not meet this constraint cannot be considered optimal. Consequently, we can further reduce the ineffective search space of the mutation operator and determine an upper bound for the mutation probability p_m .

Theorem 3.7. *Let \bar{p}_m denote the upper bound of the mutation probability, which can be expressed as:*

$$\bar{p}_m = \min\left\{\frac{1}{\sum_{j=1}^n 1/h_j}, \frac{1}{\sum_{j=1}^n 1/l_j}\right\}. \quad (3.15)$$

Proof. Without loss of generality, we assume that $0 < \frac{1}{\sum_{j=1}^n 1/h_j} \leq \frac{1}{\sum_{j=1}^n 1/l_j}$. It is evident that, according to Corollary 3.3 and Corollary 3.4, if $\bar{p}_m > \frac{1}{\sum_{j=1}^n 1/h_j}$, the individual processed by the IMO cannot be an optimal solution in terms of mathematical expectation. The proof of Theorem 3.7 is thereby established. \square

In the past decades, numerous researchers have conducted extensive research on the optimal mutation probability for convex functions or \mathcal{P} problems such as OneMax and LeadingOnes. The commonly obtained value for this probability is $\frac{1}{n}$ [18, 19, 20]. These studies have led to the following conclusions.

Proposition 1. [25] *With the increase in problem size, the mutation probability gradually tends to 0, i.e.,*

$$\lim_{n \rightarrow \infty} p_m = 0.$$

As the problem size increases, it becomes apparent that for a general instance of the 0-1 KP, the number of items with profit density greater than the break item also increases. This raises an intriguing question of whether the upper bound \bar{p}_m for the mutation probability in the IMO gradually tends to 0.

Theorem 3.8. *Let R be a constant such that $R \in \mathbb{N}^+$ and consider an instance \mathcal{P} of the 0-1 KP with $p_j, w_j \in \{1, 2, \dots, R\}$. We can conclude that*

$$\lim_{n \rightarrow \infty} \bar{p}_m = 0.$$

Proof. For an instance of the 0-1 KP, without loss of generality, let the profit density of the break item be constant at p_b/w_b , and let the residual capacity r also be constant. The knapsack capacity C can then be expressed as:

$$C = \sum_{j=1}^{b-1} w_j + r.$$

Consider the item set N_1 consisting of items with profit density greater than the break item b , and let R be any constant. It is evident that there exists a constant $M \in \mathbb{N}^+$ such that $m < M$ and N_1 can be expressed as the disjoint union of sets N_{1i} , as defined in Definition 5. In other words, for any $j \in N_1$, we have:

$$\frac{p_j}{w_j + r/M} > \frac{p_b}{w_b}.$$

Furthermore, by leveraging Corollary 3.4, we are able to acquire a vector H that is associated with the given instance. Consequently, the following conclusion can be derived:

$$\lim_{n \rightarrow \infty} \bar{p}_m \leq \lim_{n \rightarrow \infty} \frac{1}{\sum_{j=1}^n 1/h_j} \leq \lim_{n \rightarrow \infty} \frac{M}{n} = 0.$$

The proof of Theorem 3.8 is complete. □

It is notable that when the profit and weight values of items in a 0-1 KP instance are limited, as exemplified in Theorem 3.8, the instance can be solved within $O(Rn^2)$ time complexity by Dynamic Programming[8, 9]. Since R is a constant, the instance belongs to the \mathcal{P} complexity class. Consequently, it becomes an interesting question of whether the upper bound of the mutation operator in the IMO still tends towards 0 when the instance of the 0-1 KP no longer limits the profit and weight values of items.

Theorem 3.9. *For any given constant θ within the interval $(0, 1)$, it is feasible to construct an instance of the 0-1 KP with $p_j, w_j \in \mathbb{N}^+$, and we have*

$$\lim_{n \rightarrow \infty} \bar{p}_m = \theta.$$

Proof. Following the same approach as the proof of Theorem 3.8, we assume that the profit density of the break item remains constant at p_b/w_b , while maintaining a constant residual capacity r . Furthermore, the knapsack capacity C must satisfy

$$C = \sum_{j=1}^{b-1} w_j + r.$$

Without loss of generality, we set $\theta = \frac{1}{2}$. Next, we construct the profit and weight of items in the instance, ensuring that their profit densities exceed that of the break item. Let p_j and w_j represent the profit and weight, respectively, of the j -th item, satisfying

$$\lfloor \frac{rp_b}{p_j w_b - p_b w_j} \rfloor + 1 = 2^{j-1}. \quad (3.16)$$

In other words, we have $h_j = 2^{j-1}$. Since $p_j, w_j \in \mathbb{N}^+$, it is clear that there exist p_j and w_j that satisfy equality (3.16). Consequently, we can conclude that

$$\lim_{n \rightarrow \infty} \bar{p}_m = \lim_{n \rightarrow \infty} \frac{1}{\sum_{j=1}^n 1/h_j} = \frac{1}{2} = \theta.$$

The proof of Theorem 3.9 is complete. \square

3.3 Comparison analysis

In \mathcal{NP} -hard problems, given two solutions \mathbf{X}' and \mathbf{X}'' , even if $\|\mathbf{X}' - \mathbf{Y}\| < \|\mathbf{X}'' - \mathbf{Y}\|$, we cannot conclude that $f(\mathbf{X}') \geq f(\mathbf{X}'')$, where \mathbf{Y} is the optimal solution. In other words, solutions with a smaller Hamming distance to the optimal solution \mathbf{Y} do not necessarily dominate solutions with a larger Hamming distance to \mathbf{Y} . Therefore, the existing convergence analysis of GAs is limited to specific scenarios. For a general instance of \mathcal{NP} -hard problems, the expected runtimes of the GA often exceed $O(2^n)$, which lacks strong theoretical guarantees for the performance.

To demonstrate the performance of the IMO, we evaluate the algorithm by considering the probability $\tau(\text{Algorithm})$ of hitting the optimal solution \mathbf{Y} from the zero vector $\mathbf{0}$ within a single iteration for a given instance \mathcal{P} . This evaluation is conducted to assess the performance of the IMO and the MO on general large-scale instances.

Example 3.10. *Given an instance \mathcal{P} of the 0-1 KP with n items, let λ_1 and λ_2 denote the number of selected and unselected items within the first $b-1$ items, respectively. Similarly, λ_3 and λ_4 represent the number of selected and unselected items after the $(b-1)$ -th item, respectively. We have $\lambda_1 = \sum_{j=1}^{b-1} |y_j - 1|$, $\lambda_2 = \sum_{j=1}^{b-1} |y_j|$, $\lambda_3 = \sum_{j=b}^n |y_j|$, and $\lambda_4 = \sum_{j=b}^n |y_j - 1|$.*

Theorem 3.11. *If $\lambda_1 < \lambda_2$ and $p_m < 0.5$, then $\tau(\text{IMO}) > \tau(\text{MO})$.*

Proof.

$$\begin{aligned} \frac{\tau(\text{IMO})}{\tau(\text{MO})} &= \frac{(1-p_m)^{\lambda_1} p_m^{\lambda_2} p_m^{\lambda_3} (1-p_m)^{\lambda_4}}{p_m^{\lambda_1+\lambda_3} (1-p_m)^{\lambda_2+\lambda_4}} \\ &= \frac{p_m^{\lambda_2+\lambda_3} (1-p_m)^{\lambda_1+\lambda_4}}{p_m^{\lambda_1+\lambda_3} (1-p_m)^{\lambda_2+\lambda_4}} \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{1-p_m}{p_m}\right)^{\lambda_2-\lambda_1} \\
&> 1.
\end{aligned}$$

This completes the proof of Theorem 3.11. □

Theorem 3.11 demonstrates that when the number of items in the optimal solution exceeds the number of items not in the optimal solution before the break item, the probability of the IMO hitting the optimal solution in a single iteration is higher, which aligns with empirical experimental data. Moreover, since GAs are primarily employed for solving large-scale \mathcal{NP} -hard problems, which rarely exhibit the scenario of Example 3.10 with $\lambda_1 \geq \lambda_2$, and the mutation probability p_m does not exceed 0.5, the IMO generally outperforms the MO in large-scale instances.

4 Conclusion and future work

Heuristic algorithms with parameters exhibit significant performance differences depending on the value of the parameters. Due to limitations in mathematical tools, the problem of finding the optimal parameters for heuristic algorithms has received extensive attention in recent decades. Based on the assumption that $\mathcal{NP} \neq \mathcal{P}$, we propose a novel reduction method and apply it to compute the upper bound of the mutation probability.

Furthermore, our method not only proves that in the 0-1 KP, when the weight and profit of items are limited, the mutation probability tends toward 0, but also demonstrates that the mutation probability can tend toward a constant within the open interval $(0, 1)$ when the weight and profit values of items are unrestricted.

For future work, we can approach it from two perspectives. Firstly, the upper bound of the method deserves further improvement. Our research only utilized the linear relaxation technique, and it would be worthwhile to explore better upper bound computation methods, such as the Lagrangian relaxation technique. Secondly, Dey et al. have demonstrated that such reduction methods can apply to not only one-dimensional cases but also multidimensional problems. Therefore, the application scope of these reduction methods can be further expanded, such as multidimensional knapsack problems.

References

- [1] S. Martello and P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*, John Wiley & Sons, Chichester, 1990.

- [2] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*, Springer-Verlag, Berlin, Heidelberg, 2004.
- [3] V. Cacchiani, M. Iori, A. Locatelli, and S. Martello, “Knapsack problems- An overview of recent advances. Part I: Single knapsack problems,” *Computers & Operations Research*, vol. 143, pp. 105692, 2022.
- [4] V. Cacchiani, M. Iori, A. Locatelli, and S. Martello, “Knapsack problems-An overview of recent advances. Part II: Multiple, multidimensional, and quadratic knapsack problems,” *Computers & Operations Research*, vol. 143, pp. 105693, 2022.
- [5] S. A. Cook, “The complexity of theorem-proving procedures,” in *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing(STOC)*, ACM, New York, NY, USA, 1971, pp. 151-158.
- [6] R. M. Karp, “Reducibility among Combinatorial Problems,” in *Complexity of Computer Computations*, R. E. Miller, J. W. Thatcher, and J. D. Bohlinger, Eds. Boston, USA: Springer, 1972, pp. 85-103.
- [7] M. R. Garey and D. S. Johnson, *Computer and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, CA, 1979.
- [8] R. Bellman, *Dynamic programming*, Princeton University Press, Princeton, 1957.
- [9] P. Toth, “Dynamic programming algorithms for the Zero-One Knapsack Problem,” *Computing*, vol. 25, pp. 29-45, 1980.
- [10] E. Horowitz and S. Sahni, “Computing Partitions with Applications to the Knapsack Problem,” *Journal of the ACM*, vol. 21, no. 2, pp. 277-292, 1974.
- [11] R. Schroepel and A. Shamir, “A $T = O(2^{n/2})$, $S = O(2^{n/4})$ algorithm for certain NP-complete problems,” *SIAM Journal on Computing*, vol. 10, no. 3, pp. 456-464, 1981.
- [12] M. V. Sapir, J.-C. Birget, and E. Rips, “Isoperimetric and isodiametric functions of groups,” *Annals of Mathematics*, vol. 156, no. 2, pp. 345-466, 2002.
- [13] I. Dinur and S. Safra, “On the hardness of approximating vertex cover,” *Annals of Mathematics*, vol. 162, no. 1, pp. 439-485, 2005.
- [14] J. H. Holland, *Adaptation in Natural and Artificial Systems*, MIT Press, 1975.
- [15] H. Zhang, J. Xie, J. Ge, Z. Zhang, and B. Zong, “A hybrid adaptively genetic algorithm for task scheduling problem in the phased array radar,” *European Journal of Operational Research*, vol. 272, no. 3, pp. 868-878, 2019.
- [16] Y. Nagata and S. Kobayashi, “A Powerful Genetic Algorithm Using Edge Assembly Crossover for the Traveling Salesman Problem,” *INFORMS Journal on Computing*, vol. 25, no. 2, pp. 346-363, 2012.
- [17] T. Vidal, T. G. Crainic, M. Gendreau, N. Lahrichi, and W. Rei, “A Hybrid Genetic Algorithm for Multidepot and Periodic Vehicle Routing Problems,” *Operations Research*, vol. 60, no. 3, pp. 611-624, 2012.

- [18] Y. Zhou and J. He, “A Runtime Analysis of Evolutionary Algorithms for Constrained Optimization Problems,” *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 5, pp. 608-619, 2007.
- [19] C. Qian, C. Bian, W. Jiang, and K. Tang, “Running Time Analysis of the (1 + 1)-EA for OneMax and LeadingOnes Under Bit-Wise Noise,” *Algorithmica*, vol. 81, pp. 749-795, 2019.
- [20] W. Zheng, H. Chen, and X. Yao, “Analysis of Evolutionary Algorithms on Fitness Function with Time-linkage Property,” *IEEE Transactions on Evolutionary Computation*, vol. 25, no. 4, pp. 696-709, 2021.
- [21] A. E. Eiben, E. H. L. Aarts, and K. M. Van Hee, “Global convergence of genetic algorithms: A markov chain analysis,” in *Parallel Problem Solving from Nature. PPSN 1990*, H.-P. Schwefel and R. Manner, Eds. Springer, Berlin, Heidelberg, 1991, pp. 507-516.
- [22] G. Rudolph, “Convergence analysis of canonical genetic algorithms,” *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 96-101, 1994.
- [23] G. Rudolph, “Finite Markov Chain Results in Evolutionary Computation: A Tour d’Horizon,” *Fundamenta Informaticae*, vol. 35, no. 1-4, pp. 67-89, 1998.
- [24] G. Karafotias, M. Hoogendoorn, and A. E. Eiben, “Parameter Control in Evolutionary Algorithms: Trends and Challenges,” *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 2, pp. 167-187, 2015.
- [25] J. Hesser and R. Manner, “Towards an optimal mutation probability for genetic algorithms,” in *Proceedings of the 1st Conference on Parallel Problem Solving from Nature*, Springer, Dortmund, Germany, 1991, pp. 23-32.
- [26] F. G. Lobo, C. F. Lima, and Z. Michalewicz, *Parameter Setting in Evolutionary Algorithms*, Springer-Verlag Berlin Heidelberg, 2007.
- [27] O. Kramer, *Genetic Algorithm Essentials*, Springer, 2017.
- [28] Z. Michalewicz, “Ubiquity symposium: Evolutionary computation and the processes of life: the emperor is naked: evolutionary algorithms for real-world applications,” *Ubiquity*, vol. 2012, no. November, pp. 1-13.
- [29] Z. Michalewicz, “Quo Vadis, Evolutionary Computation?” in *Advances in Computational Intelligence*, J. Liu, C. Alippi, B. Bouchon-Meunier, G. W. Greenwood, and H. A. Abbass, Eds., vol. 7311, Springer, Berlin, Heidelberg, 2012.
- [30] K. Sorensen, “Metaheuristics-the metaphor exposed,” *International Transactions in Operational Research*, vol. 22, pp. 3-18, 2015.
- [31] D. H. Wolpert and W. G. Macready, “No free lunch theorems for optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67-82, Apr. 1997.
- [32] J. C. Culberson, “On the futility of blind search: An algorithmic view of ‘No Free Lunch’,” *Evolutionary Computation*, vol. 6, no. 2, pp. 109-127, Jun. 1998.

- [33] Y. C. Ho, Q. C. Zhao, and D. L. Pepyne, “The No Free Lunch Theorems: Complexity and Security,” *IEEE Transactions on Automatic Control*, vol. 48, no. 5, pp. 783-795, May 2003.
- [34] G. J. Koehler, “Conditions that Obviate the No-Free-Lunch Theorems for Optimization,” *INFORMS Journal on Computing*, vol. 19, no. 2, pp. 273-279, May 2007.
- [35] A. H. Land and A. G. Doig, “An automatic method of solving discrete programming problem,” *Econometrica*, vol. 28, no. 3, pp. 497-520, Jul. 1960.
- [36] S. S. Dey, Y. Dubey, and M. Molinaro, “Branch-and-bound solves random binary IPs in poly(n)-time,” *Mathematical Programming*, vol. 200, pp. 569-587, 2023. Accepted for publication in SODA 2021.
- [37] R. A. Brualdi, *Introductory Combinatorics*, Pearson, Fifth Edition, 2017.
- [38] E. Y. Jin, “Symmetric generating functions and Euler-Stirling statistics on permutations,” *Journal of Combinatorial Theory, Series A*, vol. 197, p. 105752, Mar. 2023.
- [39] D. Pisinger, “An expanding-core algorithm for the exact 0-1 Knapsack Problem,” *European Journal of Operational Research*, vol. 87, no. 1, pp. 175-187, Nov. 1995.
- [40] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [41] G. B. Dantzig, “Discrete-Variable Extremum Problems,” *Operations Research*, vol. 5, no. 2, pp. 266-288, Apr. 1957.
- [42] E. Balas and E. Zemel, “An Algorithm for the Large Zero-One Knapsack Problems,” *Operations Research*, vol. 28, no. 5, pp. 1130-1154, Sep. 1980.
- [43] D. Pisinger, “A Minimal Algorithm for the 0-1 Knapsack Problem,” *Operations Research*, vol. 46, no. 5, pp. 758-767, Oct. 1997.
- [44] S. Martello, D. Pisinger, and P. Toth, “Dynamic programming and strong bounds for the 0-1 knapsack problem,” *Management Science*, vol. 45, no. 3, pp. 414-424, Mar. 1999.
- [45] S. Martello, D. Pisinger, and P. Toth, “New trends in exact algorithms for the 0-1 knapsack problem,” *European Journal of Operational Research*, vol. 123, no. 2, pp. 325-332, Jun. 2000.
- [46] X. Wang and Y. He, “Evolutionary Algorithms for Knapsack Problems,” *Journal of Software*, vol. 28, no. 1, pp. 1-16, Jan. 2017. (in Chinese)
- [47] L. M. Schmitt, “Theory of genetic algorithms,” *Theoretical Computer Science*, vol. 259, no. 1-2, pp. 1-61, Nov. 2001.
- [48] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, 1996.
- [49] A. E. Eiben, R. Hinterding, and Z. Michalewicz, “Parameter Control in Evolutionary Algorithms,” *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 2, pp. 124-141, Jul. 1999.
- [50] B. Chopard and M. Tomassini, *An Introduction to Metaheuristics for Optimization*, Springer, 2018.
- [51] M. Gendreau and J. Y. Potvin, *Handbook of Metaheuristics*, Springer, 2nd Edition, 2010.