Measuring Population Diversity in Variable Dimension Search Spaces

Michal Pluhacek
Tomas Bata University in Zlin
Zlin, Czech Republic
pluhacek@utb.cz

Adam Viktorin
Tomas Bata University in Zlin
Zlin, Czech Republic
aviktorin@utb.cz

Tomas Kadavy
Tomas Bata University in Zlin
Zlin, Czech Republic
kadavy@utb.cz

Jozef Kovac
Tomas Bata University in Zlin
Zlin, Czech Republic
j_kovac@utb.cz

Roman Senkerik
Tomas Bata University in Zlin
Zlin, Czech Republic
senkerik@utb.cz

ABSTRACT

Measuring diversity in evolutionary algorithms presents a complex challenge, especially in optimization tasks with variable dimensionality. Current literature offers limited insights on effectively quantifying diversity under these conditions. This paper addresses this gap by evaluating the effectiveness of conventional diversity measures in variable dimension contexts and identifying their limitations. We introduce a novel diversity measurement approach tailored to these dynamic environments. Our method comprehensively captures both the structural and parametric diversity of populations, providing a more nuanced understanding of diversity changes over time. Through a series of experimental scenarios, we demonstrate that our proposed measure effectively tracks the evolution of diversity in populations with variable dimensions.

CCS CONCEPTS

• **Mathematics of computing → Evolutionary algorithms**.

KEYWORDS

Variable dimension, Diversity, Evolutionary computation

1 INTRODUCTION

In the realm of artificial intelligence (AI), optimization plays a pivotal role, particularly in applications involving network design [4], resource allocation [23], and machine learning [19]. Metaheuristic algorithms have emerged as powerful tools for solving complex optimization problems due to their flexibility and robustness [15].

However, a critical challenge arises when these algorithms are applied to problems with variable problem dimensionality, a scenario frequently encountered in dynamic and uncertain environments. Examples can be found in various domains: navigation and manipulation tasks in robotics, further logistics, scheduling, mission and path planning and also in data preprocessing in machine learning.

Population diversity is a key factor in the success of meta-heuristic algorithms, influencing their ability to explore the search space and avoid premature convergence [26]. Despite its importance, standard methods for measuring population diversity are primarily designed for fixed–dimension problems, leading to inadequacies in scenarios where the dimensionality of the problem varies.

This paper aims to bridge this gap by focusing on the evaluation and comparison of standard methods for measuring population diversity within the context of variable dimension search spaces. By addressing this underexplored area, our research endeavors to enhance the performance and applicability of metaheuristic algorithms, providing practical solutions for a wide range of domains facing the challenge of variable problem dimensionality.

The originality and motivation of the present research can be defined as follows.

The experimental study here focuses on the selected use case of object selection with parameter optimization, specifically the mixed-integer domain. The research questions are simple: How do we measure diversity in such a case, and are there any differences between possible approaches?

Motivation can be viewed from several directions. This research seeks to support the development of efficient algorithms for optimization in a problem environment with variable dimension by implementing not only the principle itself to deal with the problem given by the nature of the dynamic optimization problem, but also by analyzing the diversity of the population. This data would then provide beneficial feedback into the control or adaptation of the algorithm run and could also help to understand the characteristics of the optimization problem. Currently, the standard procedure for the aforesaid optimization problem is to complement missing dimensions to the maximum possible length of an individual, and this inherently increases the search space complexity significantly. The desire to develop new approaches for algorithms and for analyzing their dynamics clearly defines the research goal.

The organization of the paper is as follows: First, a brief overview of the problem of optimization with variable-length (dimension)

of the individual (solution) is given, followed by a review of approaches to computing diversity in the fixed–dimension problem domain. The following experimental section provides a demonstration and comparison of population diversity computation approaches for synthetic data generated on the basis of a real–world optimization problem. The findings and possible future directions are then summarized in the discussion and conclusion sections.

2 VARIABLE DIMENSIONALITY IN OPTIMIZATION

The problem of optimization with variable dimensionality can be identified for several real-world problems. In most cases, the conversion or encoding of the solution for an optimization algorithm is performed so that the dimensionality of the solution, and hence e.g. the number of genes for a genetic algorithm (GA), remains fixed throughout the algorithm's run. The simplest representation is the problem of subset selection, or selecting elements with a fixed number of optimized parameters. Examples include several navigation problems, such as [28], where an underwater glider is navigated by a fixed number of surfacings, adjusting (optimizing) angles and headings with respect to weather conditions and sea current forecasts. Another example would be the navigation of robots [22] in the environment and optimizing the route from the starting point to the destination over obstacles or a variable number of waypoints. From the robotics field, also the industrial robotics arm controlling problem [13] was identified as variable dimension optimization task [12]. The AI field covers variable dimensional problems like feature selection in high dimensional classification [25], as well as neural network architecture design [17]. A selected example that also inspired the data generation was a mixed–integer operational research task where the number and location of municipal waste incinerators were decided, and the routing problem for the waste collection for these newly located incinerators was optimized considering many constraining criteria and the economics of the operation and collection [10].

In the context of metaheuristic algorithms, the most straightforward approach involves utilizing an encoding scheme that is tailored to the specific features of the problem at hand. Nevertheless, when addressing the issue of variable dimensionality and aiming to understand the interactions among genes, a distinct version of a genetic algorithm, known as the messy GA [7], is employed. This specialized algorithm is designed to navigate the complexities associated with varying chromosome lengths. The algorithm was proposed to accelerate convergence on selected multimodal problems, and the core idea was to include the search for relations in black–box optimization problem representations. Thus, the genes do not have a fixed position. The algorithm was also characterized by a specific encoding of genes, two different mechanisms for dealing with so–called overspecified (conflicting) genes or underspecified genes, which are resolved by filling in missing information (genes) based on the population's best–known individual (solution). Other examples from the associated algorithmic domain are Fast Messy Genetic Algorithm [6], A multi–population pattern searching algorithm (MuPPetS) introduced in [16] which presents an answer to situations where long–coded individuals are a must and uses some of the messy GA ideas like coding and operators. Another

algorithm derived from messy GA, which mainly tried to remove the computational drawbacks of the original algorithm design, was Gene Expression Messy GA (GEMGA) [14].

Beyond the group of GA-based algorithms, popular metaheuristics have also been specifically adapted for a particular variable dimension applications. Examples are Differential Evolution (DE) [22] and Particle Swarm Optimization (PSO) [25].

3 POPULATION DIVERSITY STATE OF THE ART

In the field of continuous single–objective optimization, metaheuristic algorithms stand out for their ability to navigate complex search spaces efficiently. A pivotal aspect of their success lies in the adaptive mechanisms these algorithms often employ, particularly in response to population diversity. Population diversity serves as a crucial indicator of the search process's state, guiding the algorithm towards exploration or exploitation as necessary [26].

The evolution of metaheuristic algorithms has seen the incorporation of advanced adaptive strategies, explicitly focusing on the measurement and management of population diversity. This involves the fine–tuning of algorithm–specific parameters like mutation rates in GAs [8], flight parameters in PSO [11, 18, 21], or scaling factors in DE [2, 3, 24], all predicated on the ongoing evaluation of diversity levels. These strategic adjustments facilitate an algorithm's ability to elude local optima, promoting a more effective convergence towards the optimal solution by maintaining a balance between diversity and algorithmic focus.

The significance of these adaptive strategies lies in their role in enhancing the robustness and flexibility of metaheuristic algorithms, enabling them to navigate through the complex landscapes of optimization problems adaptively. By integrating responses to diversity fluctuations, these algorithms demonstrate an exceptional capacity for both exploring new possibilities and exploiting known information to identify optimal solutions efficiently. The dynamic management of population diversity stands as a cornerstone in the algorithm's ability to tackle a broad spectrum of optimization challenges effectively.

Nevertheless, the analysis of population dynamics through diversity measures needs to be viewed critically, as demonstrated in [20]. The paper analyzes the prevalent methods used to quantify population diversity and reveals a significant insight: the common practice of encapsulating the complex state of a metaheuristic population into a single numerical value of diversity is overly simplistic and can be misleading. The presented results demonstrate that a lower diversity measure does not necessarily correlate with diminished prospects for generating new feasible solutions. This finding challenges existing assumptions in the field and underscores the need for a more nuanced understanding of population diversity's role in the effectiveness of metaheuristic algorithms.

Considering the criticisms and the usefulness of the diversity–based adaptations in metaheuristic algorithms for single–objective optimization with consistent dimensionality, the applicability and effectiveness of these approaches in environments characterized by variable dimensionality might be promising. Such extension suggests a potential for improved adaptiveness and solution quality in more complex optimization scenarios, thereby underscoring the

adaptive mechanisms' crucial role in the ongoing development and success of metaheuristic algorithms.

The following section describes common ways of calculating diversity in a population, according to [26], and [27].

3.1 Diversity meassures

(1) **Euclidean Distance–based Measure**: This method calculates the average of the pairwise Euclidean distances between all individuals in the population (1). A higher average distance implies greater diversity, indicating that the solutions are spread out over the search space. the measure is more beneficial in continuous spaces.

$$D_{\text{Euclidean}} = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{\sum_{k=1}^{d}(x_{ik} - x_{jk})^2} \qquad (1)$$

(2) **Variance of Solutions**: This approach measures the average dispersion of solutions across all dimensions (2). A high variance suggests that the population is exploring a wider region of the search space.

$$D_{\text{variance}} = \frac{1}{d} \sum_{k=1}^{d} \left(\frac{1}{N} \sum_{i=1}^{N}(x_{ik} - \bar{x}_k)^2 \right) \qquad (2)$$

(3) **Standard Deviation of Fitness Values**: This method assesses diversity based on the variability of fitness values among the individuals (3). A larger standard deviation indicates that the fitness differences among the explored solutions are higher.

$$D_{\text{fitness}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N}(f_i - \bar{f})^2} \qquad (3)$$

(4) **Search Space Coverage**: This method involves dividing the search space into a grid and assessing the proportion of this grid occupied by the population (4). It helps in understanding how uniformly the population is distributed across the search space.

$$D_{\text{coverage}} = \frac{\text{Number of occupied cells}}{\text{Total number of cells in grid}} \qquad (4)$$

Note: The specifics of grid definition and counting occupied grids depends on the problem and implementation.

(5) **Radius of Gyration**: The radius of gyration measures the spread of the population around its center of mass (5). A larger radius indicates that the population is more dispersed in the search space. This metric is particularly useful for understanding the spatial distribution of solutions in the search space [1].

$$D_{\text{gyration}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \|\mathbf{x}_i - \mathbf{x}_{\text{cm}}\|^2} \qquad (5)$$

Where N is the number of individuals in the population, $\mathbf{x}_i$ represents the position vector of the i–th individual, and $\mathbf{x}_{\text{cm}}$ is the center of mass of the population.

These methods vary in computational complexity and sensitivity. For instance, Euclidean distance measures can be computationally intensive for large populations and large–scale optimization, while variance–based measures offer a simpler alternative but might not capture all aspects of the diversity.

3.2 Limitations and applicability in variable dimension search spaces

(1) **Euclidean Distance–based Measure**
These measures assume the same number of dimensions for calculating distances between individuals. In variable dimension scenarios, individuals might have different numbers of components, making it difficult to compute a standard Euclidean distance, as there is no straightforward way to calculate distances between such vectors.

(2) **Variance of Solutions**
Both variance and standard deviation require a fixed number of dimensions for meaningful calculations. When the dimensions vary across individuals, these measures might not be applicable as they presuppose uniformity in the data structure. This non–uniformity leads to difficulties in defining and computing variance or standard deviation across the entire population.

(3) **Standard Deviation of Fitness Values**
The method of measuring population diversity through the Standard Deviation of Fitness Values, while useful, has its limitations in the context of evolutionary algorithms. This approach assesses diversity based on the variability of fitness values among individuals, with a larger standard deviation suggesting a wider exploration of the solution space. However, this method may not always accurately reflect the true diversity of solutions. In many cases, different regions of the search space can yield similar or identical fitness values, while adjacent regions might exhibit significant fitness disparities. This discrepancy arises because fitness landscapes in optimization problems can be highly irregular and multi–modal. Therefore, relying solely on fitness value variability can sometimes provide an incomplete or misleading picture of the population's diversity and the extent of search space exploration.

(4) **Search Space Coverage**
This approach is typically defined for a fixed and predetermined search space. In scenarios where individuals have varying numbers of dimensions, the search space becomes dynamic and ill–defined. This variability complicates the process of assessing how well the population covers the search space since the boundaries and dimensions of the space are constantly changing.

(5) **Radius of Gyration**
The radius of gyration measures the spread of the population around a central point, assuming uniform dimensionality. When dimensions vary among individuals, calculating a common center of mass and subsequently the radius of gyration becomes problematic. This method is not designed to handle the disparities in dimensionality that occur in such scenarios.

In each case, the primary challenge is the inherent assumption of uniform dimensionality in the population, which is violated in variable dimension scenarios, leading to complications in applying these traditional diversity measures.

3.3 Solutions for variable dimensions

There are also proposals for alternative approaches, such as the phylogenetic metric [9], for which the authors claim the advantages of providing different values as opposed to traditional computational approaches (genotype or phenotype) used in metaheuristics and improved prediction of population status. The reasoning the authors provide is that phylogenetic diversity metrics take into account the evolutionary history of a population. Another approach is represented by calculating diversity based on (binary) chromosome length. It can be defined as the availability of chromosome lengths and their ratio to the population size. In [5], chromosome–length–based diversity for genetic algorithms and variable chromosome length shortest path problem is defined as a metric that measures diversity at the strata level and at the sub–population level.

Traditional approaches, typically reliant on fixed–dimensional structures, falter when faced with the variable ordering of components across individuals. This variability not only complicates the process of comparing like–for–like parameters across the population but also risks misrepresenting the true diversity within the search space.

A viable solution to this quandary lies in the implementation of parameter identification mechanisms. By assigning unique identifiers (IDs) to each parameter, one can maintain a consistent frame of reference across the population, ensuring that comparisons and diversity calculations are made between corresponding parameters, even in the absence of some. This approach necessitates a reorientation of diversity measures, wherein they are applied selectively, considering only those individuals that share common parameters for each specific measurement.

While the issue of variable length of the individual often concerns the absence of parameters towards the end of an individual (e.g., in route planning), a more complex problem emerges when parameters are missing in the middle. This irregularity disrupts the conventional ordering of parameters, complicating the task of aligning and comparing corresponding parameters across the population. Such disparities in parameter arrangement not only hinder the process of conducting like–for–like comparisons but also pose the risk of misrepresenting the actual diversity present in the search space. This work looks at such a scenario, described further in the next section. Due to the limited scope of this study, we will focus in the following sections only on diversity measures utilizing Euclidian distance and Variance of solutions, as these are by far the most frequently used among published works.

4 EXPERIMENT

4.1 Problem definition

Our experimental design draws inspiration from a real–world scenario detailed in [10], which involves determining the locations and quantities of waste–processing plants. A unique aspect of this scenario is the additional parameters associated with each plant, such as capacity, which are included in the decision–making process only when there is an intention to build the particular plant. This leads to a scenario where the number of parameters and the length of each solution (individual) vary significantly.

To tailor this scenario to our research needs, we have simplified it in the following manner:

- We represent each solution as a collection of triplets, where each triplet consists of a facility ID and two facility-specific parameters p_1 and p_2 (representing attributes like capacity and cost).
- The facility IDs are integers ranging from 1 to 40, representing different potential plant locations.
- The two additional parameters p_1 and p_2 associated with each facility can take any real number value between 0 and 1.
- Each individual in our population can contain a minimum of 1 triplet and a maximum of 40, resulting in a total length variation from 3 to 120 parameters per individual.

This formulation of the problem introduces a mixed–integer and variable–dimension aspect, making it more representative of the complexities typically encountered in real–world optimization scenarios.

The data were generated synthetically emulating the evolutionary optimization process with gradual convergence of the population. Such data are sufficient for investigating approaches to diversity computation, as the goal in this paper is not to directly link the analyzed/processed diversity data back to any metaheuristic algorithm and evaluate its performance. Another argument for synthetic data is the speed and simplicity of possible whole optimization instance (simulated progress) generation.

The initial population of 50 individuals is generated randomly with each individual having a unique set (set size from 1 to 40) of IDs (ranging from 1 to 40) and associated parameters (2 for each ID). Through iterative cycles, a new population is formed by first directly transferring half of the current population. The remainder of the new population is filled by randomly selecting individuals from the current population, which are then subjected to potential mutations regarding the number of IDs, the IDs themselves, and their parameters with decreasing probability (starting at initial probability and lowering by factor of 0.95 with each following iteration). This loop continues until a termination condition is met. Which, in this case, was 100 iterations. For better illustration, the process is depicted in Algorithm 1.

4.2 Algorithms

We consider a number of approaches to compute the population diversity in variable dimensional search spaces. Firstly, we discuss three ways of tailoring existing methods (Euclidian distance and Variance) to the specifics of variable dimensionality of solutions. Afterward, we propose a new diversity calculation method specifically designed for this scenario.

4.2.1 Naive approach. This approach involves three key steps: We independently compute the diversity (either by Mean Euclid distance or variance) of the three distinct elements (ID, p_1, p_2) across the entire population:

Each of the calculated diversity values is then normalized. This normalization is crucial to ensure that the diversity measures, which may vary in scale and range, are brought to a common platform for accurate comparison and aggregation.

The final step involves averaging these normalized diversity values to yield a single comprehensive diversity metric. This aggregated measure aims to encapsulate the overall diversity of the

---

Algorithm 1 Testing data generation

Initialize $populationSize \leftarrow 50$
Initialize $ID_{min} \leftarrow 1, ID_{max} \leftarrow 40$
Initialize $ID_{countMin} \leftarrow 1, ID_{countMax} \leftarrow 40$
Initialize $parameterRange \leftarrow [0, 1]$
Initialize $mutationProbabilities \leftarrow [0.2, 0.2, 0.2]$ ▷ For ID count, ID, and parameters
$population \leftarrow$ GenerateInitialPopulation($populationSize, ID_{min}, ID_{max}, ID_{countMin}, ID_{countMax}, parameterRange$)
while not termination condition **do**
    $nextPopulation \leftarrow$ SelectRandomIndividuals($population, populationSize/2$)
    while SizeOf($nextPopulation$) < $populationSize$ **do**
        $individual \leftarrow$ SelectRandomIndividual($population$)
        With probability $mutationProbabilities[0]$, mutate ID count of $individual$
        With probability $mutationProbabilities[1]$, mutate each ID in $individual$
        With probability $mutationProbabilities[2]$, mutate each parameter in $individual$
        Add $individual$ to $nextPopulation$
    end while
    $population \leftarrow nextPopulation$
    $mutationProbabilities = mutationProbabilities * 0.95$
end while

---

population, considering the variability in IDs and their corresponding parameters should decrease over time.

These variants are hereinafter noted as Naive - Euclid (for Euclidian distance diversity) and Naive - Variance (for variance-based diversity).

Naive - Euclid:

$$D_{ID} = \text{MeanEuclideanDistance}(ID)$$
$$D_{p1} = \text{MeanEuclideanDistance}(p_1)$$
$$D_{p2} = \text{MeanEuclideanDistance}(p_2)$$

Naive - Variance:

$$D_{ID} = \text{var}(ID)$$
$$D_{p1} = \text{var}(p_1)$$
$$D_{p2} = \text{var}(p_2)$$

In both cases, each calculated diversity value is then normalized:

$$D_{ID_{norm}} = \frac{D_{ID}}{\max(D_{ID})}$$
$$D_{p1_{norm}} = \frac{D_{p1}}{\max(D_{p1})}$$
$$D_{p2_{norm}} = \frac{D_{p2}}{\max(D_{p2})}$$

The final step involves averaging these normalized diversity values to obtain a single comprehensive diversity metric:

$$Diversity_{Score} = \frac{1}{3}(D_{ID_{norm}} + D_{p1_{norm}} + D_{p2_{norm}})$$

4.2.2 Grouping by ID. In this approach, the parameter values associated with a particular ID are grouped together, and the diversity is computed for each group separately. The results are then averaged. By averaging these individual diversity scores, it provides a succinct yet comprehensive metric of overall population diversity. This method benefits from maintaining parameter–specific relevance.

These variants are hereinafter noted as Grouping - Euclid (for Euclidian distance diversity) and Grouping - Variance (for variance–based diversity).

Grouping - Euclid: For each ID, compute the mean Euclidean distance among parameters and then average these distances.

$$D_{ID_1} = \text{MeanEuclideanDistance}(Parameters_{ID_1})$$
$$D_{ID_2} = \text{MeanEuclideanDistance}(Parameters_{ID_2})$$
$$\vdots$$
$$D_{ID_n} = \text{MeanEuclideanDistance}(Parameters_{ID_n})$$
$$Diversity_{Score} = \frac{1}{n}\sum_{i=1}^{n} D_{ID_i}$$

Grouping - Variance: For each ID, compute the variance among parameters and then average these variances.

$$V_{ID_1} = \text{var}(Parameters_{ID_1})$$
$$V_{ID_2} = \text{var}(Parameters_{ID_2})$$
$$\vdots$$
$$V_{ID_n} = \text{var}(Parameters_{ID_n})$$
$$Diversity_{Score} = \frac{1}{n}\sum_{i=1}^{n} V_{ID_i}$$

4.2.3 Weighted grouping by ID. The method operates on two levels: it first quantifies the diversity of solution identifiers, reflecting the variety of options being explored within the population. Concurrently, it assesses the diversity of associated solution parameters, capturing the spread of values that these options take. Each diversity metric is independently normalized to account for differences in scale. The identifier diversity is normalized against the total number of possible IDs, while parameter diversity is inherently scaled by its definition as an average pairwise Euclidean distance. The final diversity score is obtained by averaging these normalized

metrics, providing a singular value that represents overall population diversity. This dual–level approach ensures that the composite metric encapsulates a holistic view of diversity, considering both the presence of various solution options and the extent of their exploration in the search space.

These variants are hereinafter noted as Weighted - Euclid (for Euclidian distance diversity) and Weighted - Variance (for variance–based diversity).

Weighted - Euclid:

$$D_{ID} = \text{MeanEuclideanDistance}(ID)$$

$$D_p = \text{MeanEuclideanDistance}(Parameters)$$

Weighted - Variance:

$$D_{ID} = \text{var}(ID)$$

$$D_p = \text{var}(Parameters)$$

Normalization of each metric:

$$D_{ID_{norm}} = \frac{D_{ID}}{\max(D_{ID})}$$

$$D_{p_{norm}} = \frac{D_p}{\max(D_p)}$$

The final diversity score is computed as an average of these normalized metrics:

$$Diversity_{Score} = \frac{1}{2}(D_{ID_{norm}} + D_{p_{norm}})$$

4.2.4 Proposed approach. In the face of the challenges of adapting classical diversity measures to the variable dimension search spaces, we propose a new measure tailored to this scenario. In this measure, the average inverted ID frequency in the population, the average inverted frequency of the number of IDs in individuals, and the mean–variance among parameters (grouped by ID) are normalized and averaged into a single measure. This approach should lead to better capturing the changes in population coverage of the search space.

Let P be the population, where each individual $I \in P$ is composed of a varying number of triplets (ID, p_1, p_2).

(1) **Mean inverted ID frequency** (MF_ID^{-1}): Calculate the mean inverted ID frequency across the population.

$$MF\_ID^{-1} = \text{avg}(\{f_{id}\}_{id \in IDs})^{-1}$$

where f_{id} is the frequency of an ID across the population.

(2) **Mean inverted frequency of the number of IDs in individuals** (MF_nID^{-1}): Calculate the mean inverted frequency of the number of IDs across the population.

$$MF\_nID^{-1} = \text{avg}(\{f_{nID}\} \text{where } nID = n_{min}..n_{max})^{-1}$$

where n_{min} is the lowest possible number of unique IDs in individual I, n_{max} is the highest possible number of unique IDs in individual I and f_{nID} is the frequency of nID across the population.

(3) **Mean Variance Among Parameters Grouped by ID** (MV_p): For each ID, calculate the variance among the corresponding parameters across the population, then take the mean of these variances.

$$MV\_p = \frac{1}{|IDs|} \sum_{id \in IDs} \text{var}(\{p_{id}\}_{I \in P})$$

where p_{id} is the set of parameters associated with ID id in individual I, and $|IDs|$ is the total number of unique IDs.

(4) **Normalization and averaging**: Normalize each measure by dividing it by the maximum possible value for that measure and then average the normalized values to get the final diversity score.

$$Diversity_{score} = \frac{1}{3}\left(\frac{MF\_ID^{-1}}{MF\_ID_{max}^{-1}} + \frac{MF\_nID^{-1}}{MF\_nID_{max}^{-1}} + \frac{MV\_p}{MV\_p_{max}}\right)$$

where $MF\_ID_{max}^{-1}$, $MF\_nID_{max}^{-1}$, and MV_p_{max} are the theoretical maximum values for mean inverted ID frequency, the mean inverted frequency of number of IDs, and mean variance among parameters grouped by ID, respectively.

4.3 Test scenarios:

Using the test data generator described earlier, we have evaluated the above–described diversity measures for several datasets designed in the following way:

Test scenario 1:
- ID number mutation probability = 0.2
- ID value mutation probability = 0.2
- Parameter mutation = 0.2

Test scenario 2:
- ID number mutation probability = 0.1
- ID value mutation probability = 0.1
- Parameter mutation = 0.1

Test scenario 3:
- ID number mutation probability = 0.5
- ID value mutation probability = 0.3
- Parameter mutation = 0.1

For each scenario, a distinct diversity course should be observable.

Test scenario 4: Finally, we devised a different test scenario to evaluate the effectiveness of our proposed diversity measures in dynamic search spaces. This scenario involves a series of evolutionary algorithm simulations across 50 iterations, each comprising a population of 50 individuals. Uniquely, individuals in these populations are characterized by an increasing number of parameters as iterations progress, growing by one in each successive iteration. The key feature of this setup is the dual nature of evolution within the population: while the variance in the number of parameters and IDs (identifiers) per individual decreases over time, the parameter values themselves are generated randomly at each iteration, ensuring no convergence in their distribution. The aim of this experimental setup is to rigorously test whether our diversity measures can accurately detect and quantify the diminishing variance in individual lengths and the diversity of IDs amidst the backdrop of consistently high randomness in parameter values. For better illustration, the process is depicted in Algorithm 2.

5 RESULTS AND DISCUSSION

The results of each above–described experiment scenario are depicted in this section in Figures 1–4 respectively.

The experimental results, as illustrated in the figures, provide valuable insights into the effectiveness of various diversity measurement approaches in dynamic dimensionality search spaces. It

Algorithm 2 Linear population growth testing data generation

1: Initialize $populationSize \leftarrow 50$
2: Initialize $maxIterations \leftarrow 50$
3: Initialize $parameterRange \leftarrow [0, 1]$
4: **for** $iteration \leftarrow 0$ **to** $maxIterations - 1$ **do**
5:     Initialize $population \leftarrow$ empty list
6:     **for** $individualIndex \leftarrow 1$ **to** $populationSize$ **do**
7:         $individualSize \leftarrow \text{Clip}(individualIndex + iteration, 50)$
8:         **for** $id \leftarrow 1$ **to** $individualSize$ **do**
9:             $param1 \leftarrow \text{RandomReal}(0, 1)$
10:           $param2 \leftarrow \text{RandomReal}(0, 1)$
11:           Append $(id, param1, param2)$ to $individual$
12:         **end for**
13:         Append $individual$ to $population$
14:     **end for**
15: **end for**

is evident that certain traditional diversity measures when directly applied to these evolving environments, fail to accurately reflect the true diversity within the population. This underlines the challenge of adapting standard measures to contexts where individual dimensions are not static.

Interestingly, the 'Grouping by ID' approach demonstrates considerable effectiveness in the initial three test scenarios. This method, which focuses on assessing diversity within groups of identical IDs, seems adept at capturing variations in the population's composition. However, it's important to note that while this approach is effective in certain aspects, it may not comprehensively capture all dimensions of diversity, particularly in rapidly evolving populations.

In contrast, our proposed diversity measure shows a distinct ability to capture the typically swift loss of diversity observed in the first few iterations of the evolutionary process. This phenomenon is crucial in evolutionary algorithms, where initial high diversity often quickly converges as the algorithm progresses. Significantly, our measure is uniquely capable of detecting a decrease in the variability of the population's structure, even in the presence of increasing variance in parameter values. This capability is particularly noteworthy, as it indicates the measure's sensitivity not only to the diversity of parameter values but also to the structural diversity of the individuals themselves.

These observations suggest that our proposed measure provides a more nuanced and comprehensive understanding of diversity in variable dimension search spaces. It effectively balances the assessment of both parametric and structural diversity, making it a valuable tool for monitoring and guiding evolutionary processes in complex optimization scenarios.

6 CONCLUSION

The task of implementing classical diversity measures in variable dimension search spaces presents a unique set of challenges, as these environments lack the structural regularity upon which many traditional metrics depend. The synthetic data emulating evolutionary optimization progress generated on the basis of a real–world optimization problem have been used in this study. We have evaluated
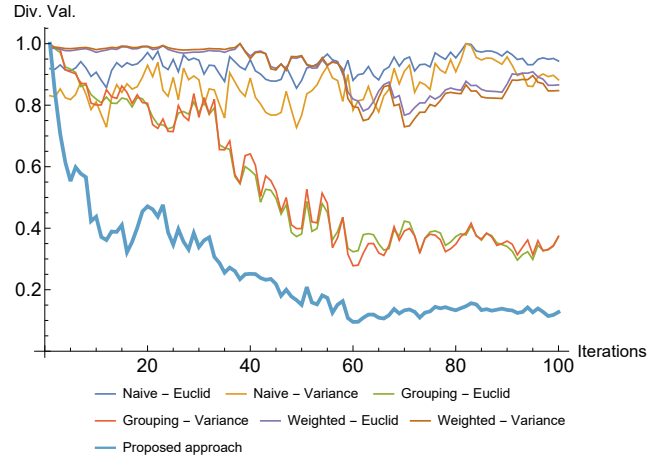
Figure 1: Normalized values for compared diversity measures - test scenario 1
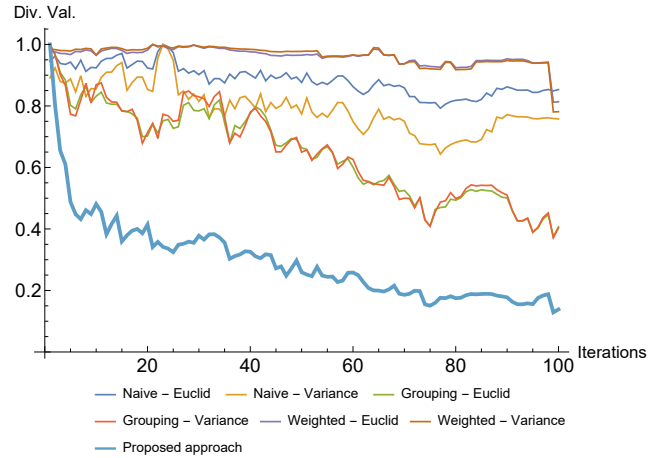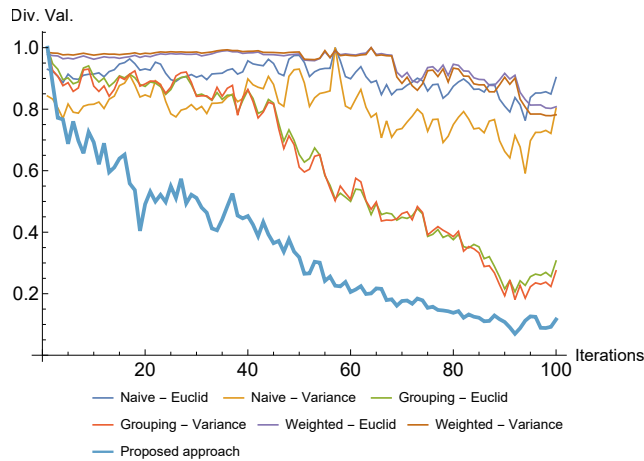

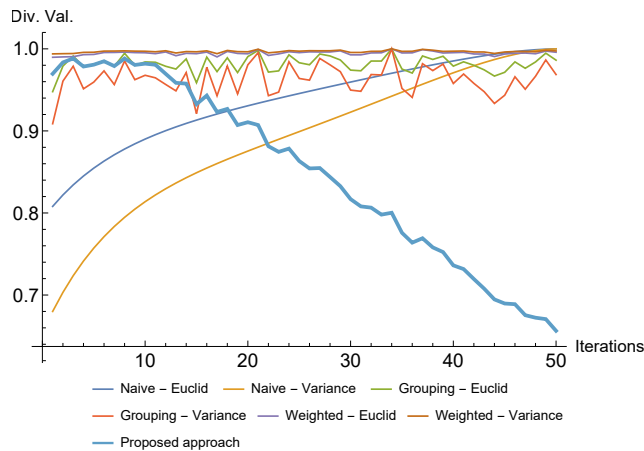**Figure 2: Normalized values for compared diversity measures - test scenario 2**

several adaptations of classical diversity measures and recognized that although they can be applied, their effectiveness varies, and they come with inherent limitations. Some adaptations may not fully capture the nuanced differences between individuals when dimensions are not consistent, while others may inadvertently mask the diversity present by overemphasizing certain aspects of the search space.

In response to these challenges, we have proposed a novel diversity measure tailored to the context of variable dimension search space. Preliminary evaluations indicate that this measure provides a promising approach for effectively quantifying diversity in this scenario, offering a more accurate reflection of the population's exploratory breadth and depth. This measure's potential lies in its ability to maintain the integrity of diversity assessment in the face of varying dimensions while being robust to the irregularities inherent in such search spaces.

diversity, catering to the unique challenges of variable dimensionality in complex optimization problems.

Figure 3: Normalized values for compared diversity measures - test scenario 3

Figure 4: Normalized values for compared diversity measures - test scenario 4

Our contributions lay the groundwork for further research, where the effectiveness of this measure can be empirically evaluated across diverse optimization problems. The promise shown by our proposed measure encourages its integration into evolutionary algorithms, which may lead to enhanced search capabilities and improved optimization outcomes.

For future research, a promising direction involves developing adaptive methods for measuring diversity in evolutionary algorithms with variable dimensions. This entails creating metrics that dynamically adapt to the individual dimensionality within a population. Key aspects would include dimensional weighting based on prevalence, aggregation with normalization to counteract biases towards individuals with more dimensions, and incorporating techniques to estimate the impact of missing dimensions. Such an approach aims to offer a more nuanced and accurate assessment of

ACKNOWLEDGMENTS

During the preparation of this work, the authors used OpenAI ChatGPT 4.0 to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

REFERENCES

[1] Hans-Georg Beyer and Hans-Paul Schwefel. 2002. Evolution strategies–a comprehensive introduction. *Natural computing* 1 (2002), 3–52.
[2] Jianchao Cheng, Zhibin Pan, Hao Liang, Zhaoqi Gao, and Jinghuai Gao. 2021. Differential evolution algorithm with fitness and diversity ranking-based mutation operator. *Swarm and Evolutionary Computation* 61 (2021), 100816.
[3] Swagatam Das, Sankha Subhra Mullick, and Ponnuthurai N Suganthan. 2016. Recent advances in differential evolution–an updated survey. *Swarm and evolutionary computation* 27 (2016), 1–30.
[4] Stephanie Alvarez Fernandez, Angel A Juan, Jésica de Armas Adrián, Daniel Guerreiro e Silva, and Daniel Riera Terrén. 2018. Metaheuristics in telecommunication systems: network design, routing, and allocation problems. *IEEE Systems Journal* 12, 4 (2018), 3948–3957.
[5] Aiman Ghannami, Jing Li, Ammar Hawbani, and Naji Alhusaini. 2021. Diversity metrics for direct-coded variable-length chromosome shortest path problem evolutionary algorithms. *Computing* 103 (2021), 313–332.
[6] David E Goldberg, Kalyanmoy Deb, Hillol Kargupta, and Georges Harik. 1993. Rapid, accurate optimization of difficult problems using messy genetic algorithms. In *Proceedings of the Fifth International Conference on Genetic Algorithms,(Urbana, USA)*. Proceedings of the Fifth International Conference on Genetic Algorithms ..., 59–64.
[7] David E Goldberg, Bradley Korb, and Kalyanmoy Deb. 1989. Messy genetic algorithms: Motivation, analysis, and first results. *Complex systems* 3, 5 (1989), 493–530.
[8] Deepti Gupta and Shabina Ghafir. 2012. An overview of methods maintaining diversity in genetic algorithms. *International journal of emerging technology and advanced engineering* 2, 5 (2012), 56–60.
[9] Jose Guadalupe Hernandez, Alexander Lalejini, and Emily Dolson. 2022. What can phylogenetic metrics tell us about useful diversity in evolutionary algorithms? In *Genetic programming theory and practice XVIII*. Springer, 63–82.
[10] Dušan Hrabec, Radovan Šomplák, Vlastimír Nevrlỳ, Adam Viktorin, Michal Pluháček, and Pavel Popela. 2020. Sustainable waste-to-energy facility location: Influence of demand on energy sales. *Energy* 207 (2020), 118257.
[11] Jing Jie, Jianchao Zeng, and Chongzhao Han. 2006. Adaptive particle swarm optimization with feedback control of diversity. In *Computational Intelligence and Bioinformatics: International Conference on Intelligent Computing, ICIC 2006, Kunming, China, August 16-19, 2006. Proceedings, Part III 2*. Springer, 81–92.
[12] Martin Juricekek, Roman Parak, and Jakub Kudela. 2023. Evolutionary computation techniques for path planning problems in industrial robotics: A state-of-the-art review. *Computation* 11, 12 (2023), 245.
[13] Ganesan Kanagaraj, SAR Sheik Masthan, and F Yu Vincent. 2022. Meta-heuristics based inverse kinematics of robot manipulator's path tracking capability under joint limits. In *Mendel*, Vol. 28. 41–54.
[14] Hillol Kargupta. 1996. Messy genetic algorithms: Recent developments. (1996).
[15] Neha Khanduja and Bharat Bhushan. 2021. Recent advances and application of metaheuristic algorithms: A survey (2014–2020). *Metaheuristic and evolutionary computation: algorithms and applications* (2021), 207–228.
[16] Halina Kwasnicka and Michal Przewozniczek. 2011. Multi population pattern searching algorithm: a new evolutionary method based on the idea of messy genetic algorithm. *IEEE Transactions on evolutionary computation* 15, 5 (2011), 715–734.
[17] Aiqin Liu, Yuezhong Zhang, Honghua Zhao, Shi Wang, and Dianmin Sun. 2021. Neural network control system of cooperative robot based on genetic algorithms. *Neural Computing and Applications* 33 (2021), 8217–8226.

[18] Christopher K Monson and Kevin D Seppi. 2006. Adaptive diversity in PSO. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation*. 59–66.

[19] Diego Oliva, Essam H Houssein, and Salvador Hinojosa. 2021. *Metaheuristics in machine learning: theory and applications*. Springer.

[20] Michal Pluhacek, Adam Viktorin, Tomas Kadavy, and Anezka Kazikova. 2021. On the common population diversity measures in metaheuristics and their limitations. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 1–7.

[21] Jacques Riget and Jakob S Vesterstrøm. 2002. A diversity-guided particle swarm optimizer-the ARPSO. *Dept. Comput. Sci., Univ. of Aarhus, Aarhus, Denmark, Tech. Rep* 2 (2002), 2002.

[22] Alejandro Rodríguez-Molina, José Solís-Romero, Miguel Gabriel Villarreal-Cervantes, Omar Serrano-Pérez, and Geovanni Flores-Caballero. 2021. Path-planning for mobile robots using a novel variable-length differential evolution variant. *Mathematics* 9, 4 (2021), 357.

[23] Arun Kumar Sangaiah, Ali Asghar Rahmani Hosseinabadi, Morteza Babazadeh Shareh, Seyed Yaser Bozorgi Rad, Atekeh Zolfagharian, and Naveen Chilamkurti. 2020. IoT resource allocation and optimization based on heuristic algorithm. *Sensors* 20, 2 (2020), 539.

[24] Jason Teo. 2006. Exploring dynamic self-adaptive populations in differential evolution. *Soft computing* 10 (2006), 673–686.

[25] Binh Tran, Bing Xue, and Mengjie Zhang. 2018. Variable-length particle swarm optimization for feature selection on high-dimensional classification. *IEEE Transactions on Evolutionary Computation* 23, 3 (2018), 473–487.

[26] Rasmus K Ursem. 2002. Diversity-guided evolutionary algorithms. In *International Conference on Parallel Problem Solving from Nature*. Springer, 462–471.

[27] Mark Wineberg and Franz Oppacher. 2003. The underlying similarity of diversity measures used in evolutionary computation. In *Genetic and Evolutionary Computation—GECCO 2003: Genetic and Evolutionary Computation Conference Chicago, IL, USA, July 12–16, 2003 Proceedings, Part II*. Springer, 1493–1504.

[28] Aleš Zamuda and José Daniel Hernández Sosa. 2019. Success history applied to expert system for underwater glider path planning using differential evolution. *Expert Systems with Applications* 119 (2019), 155–170.