An EXplainable Approach to Dominance Move (X-DoM) for Multi- and Many-Objective Optimization

Claudio Lucio do Val Lopes , Flávio Vinícius Cruzeiro Martins , Elizabeth Fialho Wanner Kalyanmoy Deb , *Fellow, IEEE,*

COIN Report 2026014

***Abstract*—Quality indicators (QIs) play a central role in assessing the performance of multi- and many-objective optimization algorithms. However, most existing QIs yield only a single scalar value that, while useful for comparison and ranking, offers little insight into why a solution set performs as it does. This lack of interpretability limits their diagnostic power, making it difficult to identify algorithmic weaknesses related to convergence, diversity, uniformity, or cardinality. The Dominance Move (DoM) indicator compares two solution sets, valued for its intuitive and physically interpretable formulation. In this paper, we extend DoM into a novel Explainable Dominance Move (X-DoM) framework. By analyzing the corrective moves required for one solution set to weakly dominate another, X-DoM decomposes the original scalar value into four intuitive and quantifiable facets: Convergence, Spread, Uniformity, and Cardinality. Comprehensive experiments on benchmark suites present that X-DoM effectively diagnoses nuanced performance issues, including poor global spread, internal structural gaps, and cardinality deficits, aspects often misrepresented or overlooked by traditional indicators, such as Hypervolume, Inverted Generational Distance, and Spacing. We also critically analyze the framework's behavior in high-dimensional many-objective scenarios with more than six objectives, identifying an "Explanatory Masking" effect where the dimension-sensitive cardinality penalty can overshadow other structural deficiencies. The X-DOM framework advances DoM from a purely comparative indicator to an explainable diagnostic tool, offering a deeper and more transparent understanding of algorithmic behavior in multi- and many-objective optimization.**

***Index Terms*—Explainability, quality indicators, multi- and many-objective optimization.**

I. Introduction

In the field of multi- and many-objective optimization, comparing and evaluating the solution sets generated by different algorithms is essential for understanding their effectiveness and guiding decision-making. Unlike single-objective problems, where a single, optimal solution can be clearly identified and ranked, multi- and many-objective problems yield a set of Pareto-optimal solutions that represent different trade-offs among conflicting objectives. As a result, evaluating the quality of these solution sets is inherently more complex and demands the use of specialized performance assessment tools known as quality indicators (QIs) [1].

Over the past decades, researchers have proposed a wide range of quality indicators (QIs) to quantify different facets of solution set quality in multi- and many-objective optimization. These typically assess *convergence*, the closeness of solutions to the true Pareto front; *spread*, the extent of front coverage; *uniformity*, the evenness of solution distribution; and *cardinality*, the number of distinct solutions in the set. Among them, the Hypervolume (HV) indicator is one of the most widely adopted due to its solid theoretical foundations, yet it remains computationally demanding and highly sensitive to the choice of reference point [2]. The Inverted Generational Distance (IGD), though effective for evaluating convergence and diversity jointly, relies on a dense and uniformly distributed reference set, something often difficult in real-world applications [3]. Similarly, the ϵ-indicator, while conceptually straightforward, tends to lose information by considering dominance in only one objective at a time [4].

The limitations of traditional QIs reveal a critical gap. Most provide only a single scalar value that, while convenient for ranking, offers little insight into *why* one solution set outperforms another. This lack of **explainability** prevents researchers and practitioners from diagnosing the specific strengths and weaknesses of an algorithm. Beyond knowing that a solution set performs better, it is essential to understand *how* it performs, whether its advantage stems from a superior convergence, a balanced distribution, or a larger coverage, or whether its deficiencies arise from a poor spread, an uneven uniformity, or having a limited cardinality.

The Dominance Move (DoM) indicator offers a promising opportunity for addressing these challenges. As a binary quality indicator, DoM quantifies the minimum total "effort" or distance required to move the solutions of one set so that it weakly dominates another [5], [6]. Its formulation inherently captures the four fundamental aspects of solution quality, convergence, spread, uniformity, and cardinality, without relying on predefined reference sets or suffering from the

Corresponding author: Cláudio Lúcio V. Lopes is founder and CIO at A3Data, Brazil (email: claudiolucio@gmail.com)

Flávio V. C. Martins is with the Computer Department at the Centro Federal de Educação Tecnológica de Minas Gerais, 30510-000 Brazil (e-mail: flaviocruzeiro@cefetmg.br).

Elizabeth F. Wanner is with the School of Computer Science and Digital Technologies, Aston University, Birmingham, UK and the Computer Department at the Centro Federal de Educação Tecnológica de Minas Gerais, 30510-000 Brazil (e-mail: e.wanner1@aston.ac.uk).

K. Deb is with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, US, e-mail: kdeb@egr.msu.edu (see http://www.coin-lab.org).

information loss typical of other indicators. This makes DoM both comprehensive and conceptually elegant.

Nevertheless, despite its conceptual richness and intuitive physical meaning, the original DoM formulation, particularly in its mixed-integer programming (MIP) implementation, still yields only a single scalar score [5]. To overcome this limitation, this paper proposes the Explainable Dominance Move (X-DoM) framework, an extension of DoM that decomposes this scalar outcome into interpretable components, providing deeper insight into the structure and quality of solution sets.

The proposed framework decomposes the overall DoM value into contributions corresponding to the four fundamental quality facets. In particular, convergence, spread, and uniformity are quantified through a local analysis of the specific movements required by individual solutions, whereas cardinality is assessed globally by comparing the total number of solutions in each set. This global score is then distributed proportionally among the individual moves, ensuring that all four aspects are consistently reflected in the final explainable breakdown. Through this decomposition, DoM evolves from a purely comparative ranking measure into a diagnostic instrument that reveals the underlying reasons for performance differences, offering a richer and more actionable understanding of algorithmic behavior than conventional quality indicators can provide.

The remainder of this paper is structured to provide a comprehensive exploration of the X-DoM framework. Section II establishes the necessary background by reviewing traditional quality indicators, discussing the principles of Explainable AI as they apply to optimization, and providing a formal definition of the Dominance Move indicator. Section III introduces our primary contribution: the detailed methodological framework for X-DoM, explaining the rationale and calculation for each of the four facets. Section IV presents a series of empirical studies, including controlled experiments and comparative analyses against established indicators on benchmark problems, to validate the framework's diagnostic accuracy and practical utility. Finally, Section V concludes the paper by summarizing our findings, discussing the implications for the EMO community, and outlining promising directions for future research.

II. Background

Quality indicators are essential tools for quantifying and comparing the quality of Pareto-optimal solution sets generated by different algorithms. In multi-objective optimization, where no single optimal solution exists, evaluation must capture multiple facets of performance rather than rely on a single fitness value [4]. This evaluation typically includes convergence toward the true Pareto front, diversity, encompassing both spread and uniformity of solutions, and cardinality, or the number of distinct solutions in the set [1]. The strengths and weaknesses of several widely used traditional QIs demonstrate a clear trade-off between computational efficiency, data dependency, and diagnostic interpretability, underscoring the need for more informative and balanced evaluation methods.

A. Explainable AI and its application to optimization

The growing demand for transparent and trustworthy algorithmic systems has led to the rise of Explainable Artificial Intelligence (XAI), a field dedicated to making complex and often opaque models comprehensible to humans. Within XAI, a key distinction exists between interpretability and explainability [7], [8]. Interpretability refers to a model's intrinsic clarity - its decision-making process can be directly understood without auxiliary tools, as in the case of shallow decision trees or linear regression models. Explainability, by contrast, refers to the ability to provide imeaningful insights into the decisions of complex "black-box" systems, typically through post-hoc analysis conducted after model training [9], [8]. The present work aligns with the latter approach, extending explainability principles to the domain of optimization by introducing interpretive mechanisms for a complex performance indicator.

The field of explainable AI has produced a variety of techniques for interpreting the behavior of complex machine learning models. One of the most influential is SHapley Additive exPlanations (SHAP), which applies concepts from cooperative game theory to quantify the contribution of each input feature to a model's prediction [10]. Another is Local Interpretable Model-agnostic Explanations (LIME), which approximates the behavior of a black-box model locally around a specific data point with a simpler, more interpretable model, such as a linear regression [11].

The application of explainability principles has expanded beyond traditional machine learning, increasingly influencing the fields of evolutionary computation and optimization. In this context, the challenge of explainability can be viewed from two complementary perspectives [12], each corresponding to the intended audience or purpose of the explanation:

- *Explaining the solution*: This category of methods focuses on providing insights into why a particular optimal or Pareto-optimal solution was generated. The goal is to help a human decision-maker understand the trade-offs involved and to build trust in the recommended solution. Some frameworks and methods based on counterfactual explanations fall into this category [13]. They act as an interface between the algorithm and the human, providing high-level summaries of the relationship between inputs (preferences, problem contexts) and outputs (the final solution).
- *Explaining the Algorithm*: This category of methods aims to reveal the internal workings, dynamics, and iterative behavior of the optimization algorithm itself. These are diagnostic tools designed for algorithm developers and researchers. They answer questions about how the search is progressing, which internal operators are effective, and how the population is evolving.

A key distinction in explainable AI lies between explaining predictions, as commonly done in machine learning, and explaining decisions, which is central to optimization. In classification tasks, for example, the number of possible outcomes (classes) is typically small and manageable. In contrast, optimization problems often involve an enormous, potentially intractable space of feasible solutions, making the explanation

of a specific decision far more challenging. Consequently, frameworks for explainable optimization must either provide a concise rationale for the chosen solution or identify the minimal changes in the problem context that would have led to an alternative, superior decision.

A data-driven framework for explainable solutions is presented in [14], where explainability is treated as a second criterion alongside the objective value in a multi-objective optimization problem. The core concept is that a solution is considered explainable if it is similar to solutions that were previously implemented in similar historical situations. The framework uses feature functions to aggregate information from both past instances and their solutions into feature vectors, and then uses similarity metrics to measure the distance between them. By explicitly modeling explainability in this manner, the framework allows for a trade-off between a solution's optimality and its similarity to trusted past decisions, which the authors argue can be as valuable as achieving strict optimality. The approach is validated on a shortest path problem, showing that enforcing explainability often incurs only a minimal cost, with fully explainable solutions requiring only a slight increase in the objective value.

The R-XIMO framework is an important development in the field of explainable interactive multi-objective optimization (XIMO) because it directly addresses the problem of supporting a human decision-maker in an iterative, preference-driven process [13]. Unlike methods that aim to explain an algorithm's entire behavior, R-XIMO focuses on providing immediate, actionable insights to the decision-maker. By treating the optimization process as a function, SHAP values can be computed to quantify how each component of the decision-maker reference point contributes to the value of each objective in the final solution, as an aspiration level for a specific objective. The result is a $k \times k$ matrix $\mathbf{\Phi}$ where $\phi_{i,j}$ represents the effect of the *j*-th component of the reference point on the *i*-th component of the solution. A positive $\phi_{i,j}$ indicates an "impairing effect", meaning that the aspiration level for objective *j* negatively impacted the final value of objective *i*, thereby revealing a local trade-off or conflict between these objectives.

The Explainable and Learnable Evolutionary Multiobjective Optimization (XLEMOO) framework [15] is a hybrid approach that integrates the exploratory power of evolutionary algorithms with the descriptive capabilities of interpretable machine learning models. The framework operates in an iterative loop of a "Darwinian mode" where solutions are evolved, and a "learning mode" where an interpretable model, such as rules, is trained on high- and low-performing solutions to formulate a hypothesis about what constitutes a good solution. The central contribution is that the rules extracted by this machine learning model, presented as simple IF-THEN statements, provide concrete and actionable explanations to a human decision-maker, clarifying the relationship between decision variables and objective functions. The authors' experiments, including a vehicle crash worthiness problem, show that this process-integrated explainability can not only guide the decision-maker to better solutions but also positively affect the performance of the algorithm itself.

The EvoMapX framework provides process-level, real-time interpretability for population-based algorithms [9]. Instead of a post-hoc analysis of a black box's inputs and outputs, EvoMapX is an intrinsic framework that provides a real-time, process-level diagnostic of an algorithm's internal dynamics [9]. Its primary purpose is not to assist a non-expert decision-maker, but to provide deep, structural insights for algorithm developers and researchers, helping them to understand how and why an algorithm behaves in a certain way over time. The framework does this by passively logging an algorithm's internal events without interfering with its execution, and then generating a series of interpretable structures and indicators. It consists of three key components: the operator attribution matrix, which quantifies the contribution of operators like crossover and mutation over time; the population evolution graph, which visualizes the lineage of solutions; and the convergence driver score, an indicator that identifies which operators are driving convergence.

The development of explainable frameworks for understanding algorithmic behavior and decision outcomes naturally leads to the need for explainability in quality indicators. While existing indicators can quantify performance, they offer little insight into the reasons behind success or failure, whether an algorithm underperforms due to poor convergence, lack of diversity, or insufficient coverage of the Pareto front. Introducing explainability at the indicator level is therefore a critical step toward transforming performance assessment into diagnostic understanding. The proposed framework advances this direction by focusing not merely on how algorithms operate, but on why their resulting solution sets exhibit particular strengths and weaknesses, providing actionable guidance for algorithm refinement and design.

B. Dominance Move

Dominance move (DoM) is a binary quality indicator used in multi- and many-objective optimization to compare two solution sets obtained by different algorithms [5]. The indicator emerged as a response to the shortcomings of traditional QIs by providing a comprehensive, reference-set-free method for comparing two solution sets. Beyond capturing differences in *convergence*, the DoM indicator is also sensitive to other critical aspects of solution set quality, including *spread*, *uniformity*, and *cardinality*.

Dominance Move (DoM): Consider that $\boldsymbol{P}$ and $\boldsymbol{Q}$ are two sets, with points $\boldsymbol{p}_i$, $i \in \{1, \ldots, |\boldsymbol{P}|\}$ and points $\boldsymbol{q}_j$, $j \in \{1, \ldots, |\boldsymbol{Q}|\}$. The dominance move of $\boldsymbol{P}$ to $\boldsymbol{Q}$, $DoM(\boldsymbol{P}, \boldsymbol{Q})$, is the minimum total distance of moving points of $\boldsymbol{P}$, such that the moved set $\boldsymbol{P'} = \{\boldsymbol{p}'_1, \boldsymbol{p}'_2, \ldots, \boldsymbol{p}'_{|\boldsymbol{P}|}\}$ (with some or all $\boldsymbol{p}'_i$ are allowed to be infeasible) from $\boldsymbol{P} = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_{|\boldsymbol{P}|}\}$ dominates $\boldsymbol{Q}$ and that the total move from $\boldsymbol{P}$ to $\boldsymbol{P'}$ must be minimum [5].

The above definition guarantees that every element in $\boldsymbol{Q}$ is dominated by at least one element in $\boldsymbol{P}'$. Since DoM aims to minimize the total movement required, it is possible that some elements in $\boldsymbol{P}'$ become strictly better in one or more objectives compared to the corresponding dominating elements in $\boldsymbol{Q}$.

The formal expression of DoM can be stated as:

$$DoM(\boldsymbol{P}, \boldsymbol{Q}) = \min_{\boldsymbol{P}' \prec \boldsymbol{Q}} \sum_{i=1}^{|\boldsymbol{P}|} d(\boldsymbol{p}_i, \boldsymbol{p}'_i), \tag{1}$$

in which $d(\boldsymbol{p}_i, \boldsymbol{p}'_i)$ can be the Manhattan distance between $\boldsymbol{p}_i$ to $\boldsymbol{p}'_i$.

Despite its conceptual richness and intuitive foundation, the original formulation of DoM presents significant computational challenges. Its exact calculation is inherently combinatorial, and although a mixed-integer programming (MIP) formulation was later proposed to make it more tractable for problems with three or more objectives [5], the outcome remains a single scalar value. While this value is effective for ranking solution sets, it conceals the wealth of information embedded in the individual movements that constitute the total dominance cost. Consequently, the DoM score summarizes a complex relationship without revealing its underlying causes. A low score might stem from poor convergence, limited spread, or uneven distribution, yet the indicator itself offers no way to distinguish among these possibilities. This limitation underscores the need to evolve from purely comparative scores toward diagnostic tools capable of decomposing and explaining their results, thereby transforming DoM from a ranking measure into a genuinely analytical instrument.

III. The Explainable Dominance Move (X-DoM) Framework

To transform the scalar DoM into a diagnostic tool, we introduce the explainable Dominance Move (X-DoM) framework. The framework operates on a simple yet powerful principle: the necessity of moving a solution $\boldsymbol{p}_i$ to a new position $\boldsymbol{p}'_i$ arises from one or more deficiencies in the original solution set P. Our goal is to attribute the cost of each move, its Manhattan distance, to four fundamental, mutually exclusive facets of quality: convergence, spread, uniformity, and cardinality.

The core of the methodology lies in analyzing each individual move generated by the DoM solver. For every point $\boldsymbol{p}_i \in \boldsymbol{P}$ that is moved, the X-DoM framework calculates four normalized factors that quantify distinct types of deficiencies in the solution set P. All factors are normalized to a $[0, 1]$ scale, where a value of 1 indicates a maximum deficiency in that facet.

The framework also distinguishes between local and global deficiencies. The factors for convergence, spread, and uniformity are calculated locally for each move, based on the properties of the specific point $\boldsymbol{p}_i$ and its relationship to its neighbors within the set P. In contrast, the cardinality factor is assessed globally by comparing the total number of solutions in sets $\boldsymbol{P}$ and $\boldsymbol{Q}$. The global deficiency is then assigned proportionally to each move, reflecting its contribution to the overall effort.

To ensure a fair and consistent comparison, all calculations are performed within a normalized objective space $[0, 1]^M$, where M is the number of objectives. The cost of each move is then proportionally distributed according to these factors. By aggregating these attributed costs across all moves, the framework produces a final, intuitive percentage breakdown that explains the composition of the total DoM value. The following subsections detail the precise rationale and calculation for each of these four factors.

A. Factor 1: Convergence

The first facet addresses a fundamental quality property: how close a solution set is to the true Pareto front. This factor quantifies the convergence deficiency of an individual point $\boldsymbol{p}_i$ by directly measuring the magnitude of its corrective move. The underlying rationale is to exploit the DoM solver's primary output, the new optimal position $\boldsymbol{p}'_i$, as the reference for where $\boldsymbol{p}_i$ should ideally be located to contribute to dominating the target set $\boldsymbol{Q}$. This is a local measure, as it is computed independently for each point that undergoes a movement.

The convergence factor is the Manhattan distance of the move, normalized by the number of objectives, M, to scale the value to the range $[0, 1]$:

$$f_{i,\text{conv}} = \frac{d(\widehat{\boldsymbol{p}}'_i - \widehat{\boldsymbol{p}}_i)}{M}, \tag{2}$$

in which $\widehat{\boldsymbol{p}}_i$ is the original coordinates of point $\boldsymbol{p}_i$, normalized to $[0, 1]^M$ using the combined bounds of sets $\boldsymbol{P}$ and $\boldsymbol{Q}$. The $\widehat{\boldsymbol{p}}'_i$ is the new coordinate of the point after the move, normalized within the same context, and d is the Manhattan distance.

A score approaching 1 indicates that the point required a substantial displacement to achieve dominance, implying that its contribution to the total DoM value was largely driven by a significant convergence gap.

B. Factor 2: Spread

This factor isolates the portion of the DoM cost associated with an insufficient spread, specifically capturing the inability of set $\boldsymbol{P}$ to cover the full extent of the target set $\boldsymbol{Q}$. The movement of an extreme point within $\boldsymbol{P}$ serves a strong indicator of this deficiency. The rationale for using a $\boldsymbol{P}$-centric normalization lies in the fact that a point's "extremeness" is inherently defined relative to the boundaries of its own set. This design enables the factor to discern when a corrective move primarily aims to expand the overall extent of the approximated Pareto front.

The factor is non-zero only if some of the moved points are identified as extreme points of $\boldsymbol{P}$. It is the product of a binary flag and a continuous extremeness score, which quantifies how peripheral a point is:

$$f_{i,\text{spread}} = \text{IsExtreme}(\boldsymbol{p_i}) \times E(\hat{\boldsymbol{p}}_{\boldsymbol{i}}), \tag{3}$$

where $\text{IsExtreme}(\boldsymbol{p_i})$ is a binary flag that is 1 if $\boldsymbol{p_i}$ is an extreme point of $\boldsymbol{P}$ (i.e., it has the minimum or maximum value for at least one objective within $\boldsymbol{P}$), and 0 otherwise. The $E(\hat{\boldsymbol{p}}_{\boldsymbol{i}})$ is the continuous extremeness score, which measures how far a point is from the center of its own set's normalized space:

$$E(\hat{\boldsymbol{p}}_{\boldsymbol{i}}) = \frac{\sum_{m=1}^{M} |\hat{p}_{i,m} - 0.5|}{M/2}, \tag{4}$$

in which $\hat{\boldsymbol{p}}_i$ is the point $\boldsymbol{p_i}$ normalized using only the minimum and maximum values found within set $\boldsymbol{P}$.

A high score for this factor provides a clear diagnostic insight: the corrective move was required because a key boundary point was mispositioned, revealing that the solution set did not adequately span the full extent of the Pareto front.

C. Factor 3: Uniformity

The idea is to quantify a score that arises from a poor distribution of solutions, such as gaps and clusters. Non-uniformity is an intrinsic structural flaw that reduces a set's efficiency in dominating a well-distributed target set [16]. Inspired by the established Spacing (SP) indicator[17], this factor is $\boldsymbol{P}$-centric and measures how much a point's local spacing deviates from the set's average spacing. This approach identifies internally flawed distributions without relying on external reference points or centroids.

The uniformity factor is the normalized absolute difference between a point's nearest-neighbor distance and the average nearest-neighbor distance of the entire set $\boldsymbol{P}$:

$$f_{i,\text{unif}} = \frac{|d_{nn}(\boldsymbol{p_i}) - \bar{d}_{nn}(\boldsymbol{P})|}{d_{\max}}, \tag{5}$$

in which $d_{nn}(\boldsymbol{p_i})$ is the Manhattan distance from point $\boldsymbol{p_i}$ to its single nearest neighbor within set $\boldsymbol{P}$, calculated in the normalized space. The $\bar{d}_{nn}(\boldsymbol{P})$ is the average of all nearest-neighbor distances (d_{nn}) for the entire set $\boldsymbol{P}$, and $d_{\max}$ is the maximum nearest-neighbor distance found across all points in set $\boldsymbol{P}$, used for normalization.

A high score indicates that $\boldsymbol{p_i}$ contributes to poor uniformity, being positioned either within an overly dense cluster (small d_{nn}) or an excessively sparse region (large d_{nn}). Such structural irregularities increase the overall effort required for the set to achieve dominance.

D. Factor 4: Cardinality

This global factor quantifies the deficiency in the number of solutions in set $\boldsymbol{P}$ relative to set $\boldsymbol{Q}$. Its rationale lies in the observation that the severity of a solution shortage depends strongly on the number of objectives M. As dimensionality increases, a substantially larger number of solutions is required to adequately approximate the Pareto front [18]. Consequently, a deficit in cardinality ($|\boldsymbol{P}| < |\boldsymbol{Q}|$) becomes a far more critical weakness in many-objective contexts. This factor is therefore designed to be dimension-sensitive, amplifying the penalty for insufficient solution density as M grows.

The calculation uses a power function to scale the deficit penalty based on the problem's dimensionality. The process starts with the base deficit ratio (r), that measures the fractional shortage of solutions in $\boldsymbol{P}$ relative to the total number of solutions:

$$r = \frac{||\boldsymbol{P}| - |\boldsymbol{Q}||}{|\boldsymbol{P}| + |\boldsymbol{Q}|}. \tag{6}$$

A comparability exponent (η) is derived from the probability that two random vectors are comparable in an M-dimensional space, providing a strong theoretical basis for scaling the penalty:

$$\eta = \frac{1}{2^{M-1}}. \tag{7}$$

Finally, the cardinality factor (f_{card}) is computed as the base deficit raised to a comparability exponent, expressed as $F_{\text{card}} = r^{\eta}$. When the number of objectives M is large, the exponent η becomes small, causing even moderate deficits $r \in [0, 1]$ to yield scores close to 1, thereby substantially amplifying the penalty. This global score is then uniformly propagated to each individual move, such that $f_{\text{card},i} = F_{\text{card}}$, ensuring that the dimensionality-driven cardinality penalty is consistently reflected across all local computations.

The score captures both the magnitude of the solution deficit and the problem's dimensionality. A score of 0 indicates no cardinality issue ($|\boldsymbol{P}| \geq |\boldsymbol{Q}|$), while for a fixed deficit, the score increases sharply with M. For example, consider a 20% solution deficit ($|\boldsymbol{P}| = 80, |\boldsymbol{Q}| = 100$, so $r \approx 0.11$). In a problem with $M = 3$, the exponent $\eta = 0.25$, and $F_{\text{card}} = (0.11)^{0.25} \approx 0.576$. For a problem with $M = 10$, where the exponent $\eta \approx 0.002$, the resulting is $F_{\text{card}} = (0.11)^{0.002} \approx 0.996$. In contrast, for a 10-objective problem ($M = 10$), where $\eta \approx 0.002$, the score becomes $F_{\text{card}} = (0.11)^{0.002} \approx 0.996$. This illustrates how the same 20% deficit is appropriately recognized as a critical shortfall, score ≈ 0.996, in a 10-objective scenario, but only a moderate one, score ≈ 0.576, in a lower-dimensional, 3-objective case.

E. Final attribution and calculation

The total cost of each individual move, C_i (defined as the Manhattan distance between $\boldsymbol{p_i}$ and its corrected position $\boldsymbol{p_i'}$), is distributed across the four quality facets in proportion to their normalized deficiency scores. This proportional allocation ensures that the sum of the facet-specific contributions exactly equals the total cost of the move:

$$C_{i,\text{facet}} = C_i \times \frac{f_{i,\text{facet}}}{\sum_{k \in \{\text{conv, spread, unif, card}\}} f_{i,k}}. \tag{8}$$

By aggregating these attributed costs over all moves, the X-DoM framework produces two key diagnostic outputs:

1) Overall facet contribution: A summary indicating the percentage of the total DoM score attributed to each of the four quality facets. This provides a high-level diagnosis of the primary weaknesses in the solution set.
2) Detailed move analysis: A detailed record for each move, reporting its total cost (C_i) and the proportional contribution of each facet. This enables a fine-grained analysis of individual solutions that drive the overall DoM score.

IV. Experiments

To evaluate the effectiveness and practical utility of the X-DoM framework, we conducted a series of computational experiments designed to demonstrate its diagnostic capabilities across a range of scenarios. First, we illustrate that X-DoM can accurately identify and quantify induced weaknesses in solution sets, attributing the associated cost to the intended quality facet. Next, we analyze common structural challenges, including poor spread in high-convergence sets, the presence of internal gaps within fronts, and deficiencies in cardinality for many-objective problems. We then explore the framework's potential as a tool for algorithm development, highlighting how

it can inform parameter tuning by revealing the underlying trade-offs in algorithmic behavior. Finally, we examine the framework's ability to explain performance differences on problems with geometrically complex Pareto fronts, demonstrating its value in understanding how algorithm design interacts with problem characteristics.

A. DoM Facets: some insightful visual explainability

To empirically validate the diagnostic capabilities of the DoM explainability framework, we designed a series of controlled experiments. For each facet, convergence, cardinality, uniformity, and spread, we use solution sets $\boldsymbol{P}$ with specific deficiencies relative to a reference set $\boldsymbol{Q}$. The goal is to show that the DoM explainability framework can identify and quantify these induced weaknesses, attributing the majority of the total DoM cost to the intended facet. The results, visualized in Figure 1, confirm the framework's ability to provide a nuanced, explainable breakdown of the quality of the solution sets.

The convergence tests highlight cases where the primary deficiency of set $\boldsymbol{P}$ lies in its distance from the Pareto front. In the first row, left plot, the set $\boldsymbol{P}$ (red circles) closely replicates the shape of $\boldsymbol{Q}$ (black squares) but remains consistently displaced, clearly indicating a lack of convergence. The DoM solver's moves (blue 'X's) are predominantly horizontal, drawing $\boldsymbol{P}$ closer to $\boldsymbol{Q}$. The inset box shows that 50.10% of the total DoM cost is attributed to convergence, the largest contribution among all facets in this scenario. This shows that the framework correctly identifies poor proximity to the front as the main issue. While other facets contribute, their effects are limited, as $\boldsymbol{P}$ maintains its relative spread, uniformity, and number of points. The second plot (right) illustrates a more severe convergence deficiency. Here, $\boldsymbol{P}$ lies farther from $\boldsymbol{Q}$ than in the previous case, particularly in certain regions, resulting in longer DoM moves. The X-DoM analysis reflects this by assigning an even higher proportion of the total cost to convergence (90.79%), underscoring the sensitivity of the indicator to the magnitude of convergence errors. This increased percentage validates the factor's sensitivity to the magnitude of convergence errors. The dominance of this factor confirms that the primary weakness of $\boldsymbol{P}$ is its failure to approach the Pareto-optimal front effectively, with uniformity contributing only marginally by comparison.

The cardinality tests examine the effect of an insufficient number of solutions in set $\boldsymbol{P}$ relative to $\boldsymbol{Q}$. In the left plot, $\boldsymbol{P}$ has one point less than $\boldsymbol{Q}$. Although some of these points appear reasonably converged, the overall sparsity of $\boldsymbol{P}$ is clear. The X-DoM analysis identifies this deficiency, assigning 63.56% of the total DoM cost to cardinality. This high proportion highlights the framework's ability to detect and penalize solution sets that lack sufficient representation to cover the Pareto front adequately. The remaining contributions, 9.06% from convergence, 5.29% from spread, and 22.10% from uniformity, indicate that while some points might have convergence or distribution issues, the limited number of solutions remains the dominant problem identified by the analysis. In the right plot, a similar cardinality deficiency is observed, with $\boldsymbol{P}$ again showing a clear deficit in the number of solutions compared to $\boldsymbol{Q}$. The X-DoM analysis reinforces this finding, assigning an even higher 79.92% of the cost to cardinality. The minimal contributions from convergence (5.28%) and spread (0.00%), along with a moderate contribution from uniformity (14.80%), confirm that the existing points are generally well positioned relative to the front and reasonably spaced, with no issues at the extremes. The critical problem overwhelmingly lies in the overall scarcity of solutions.

The uniformity tests explore scenarios in which the set $\boldsymbol{P}$ exhibits an irregular distribution, such as clusters or gaps. In the first plot, $\boldsymbol{P}$, shows clear non-uniformity: points are densely packed in some regions and sparse in others, despite generally covering the range of $\boldsymbol{Q}$. DoM indicates adjustments that target these inconsistencies. The X-DoM analysis correctly identifies uniformity as the main deficiency, attributing 52.67% of the total DoM cost to this factor. These results indicate the framework's ability to detect internal structural flaws in the solution set's distribution, weakness that aggregated indicators might otherwise conceal. The spread facet contributes 26.76%, reflecting the need for $\boldsymbol{P}$ to extend further to dominate $\boldsymbol{Q}$, while convergence accounts for 20.77%, indicating that although distribution is the dominant issue, some points also require movement towards the Pareto front. The right plot amplifies the uniformity problem: $\boldsymbol{P}$ displays even stronger clustering and larger gaps than in the previous case. Accordingly, the X-DoM analysis assigns an even higher 65.74% of the total cost to uniformity. This increase underscores the framework's sensitivity to severe distribution irregularities and confirms its effectiveness in diagnosing when a solution set fails to provide a well-distributed representation of the Pareto front.

The spread tests examine situations in which set $\boldsymbol{P}$ fails to cover the full extent of the Pareto front. In the left plot, $\boldsymbol{P}$ is heavily concentrated in the central region of the objective space, missing the extreme points reached by $\boldsymbol{Q}$. The DoM analysis is predominantly outward, extending $\boldsymbol{P}$ towards the boundaries to improve coverage. The X-DoM analysis attributes 37.25% of the total DoM cost to convergence and 62.25% to spread, revealing a dual deficiency: $\boldsymbol{P}$ not only lacks adequate range but also contains points that are not fully converged. This result demonstrates the framework's capability to detect and quantify overlapping sources of error. In the right plot, the spread deficiency becomes more pronounced. Set $\boldsymbol{P}$ spans only a small portion of the objective space covered by $\boldsymbol{Q}$, leaving large regions of the Pareto front unexplored. As a result, the DoM solver performs substantial outward moves to extend coverage. The X-DoM framework identifies this issue clearly, assigning 80.00% of the total cost to spread. This attribution confirms the framework's precision and sensitivity in diagnosing situations where the principal shortcoming of a solution set is its inability to represent the full range of the Pareto front.

Collectively, these controlled experiments demonstrate the diagnostic strength of the DoM explainability framework. In each case, the facet attributions adaptively and intuitively capture the specific weakness embedded in the solution set $\boldsymbol{P}$. Whereas traditional indicators provide only a single, opaque

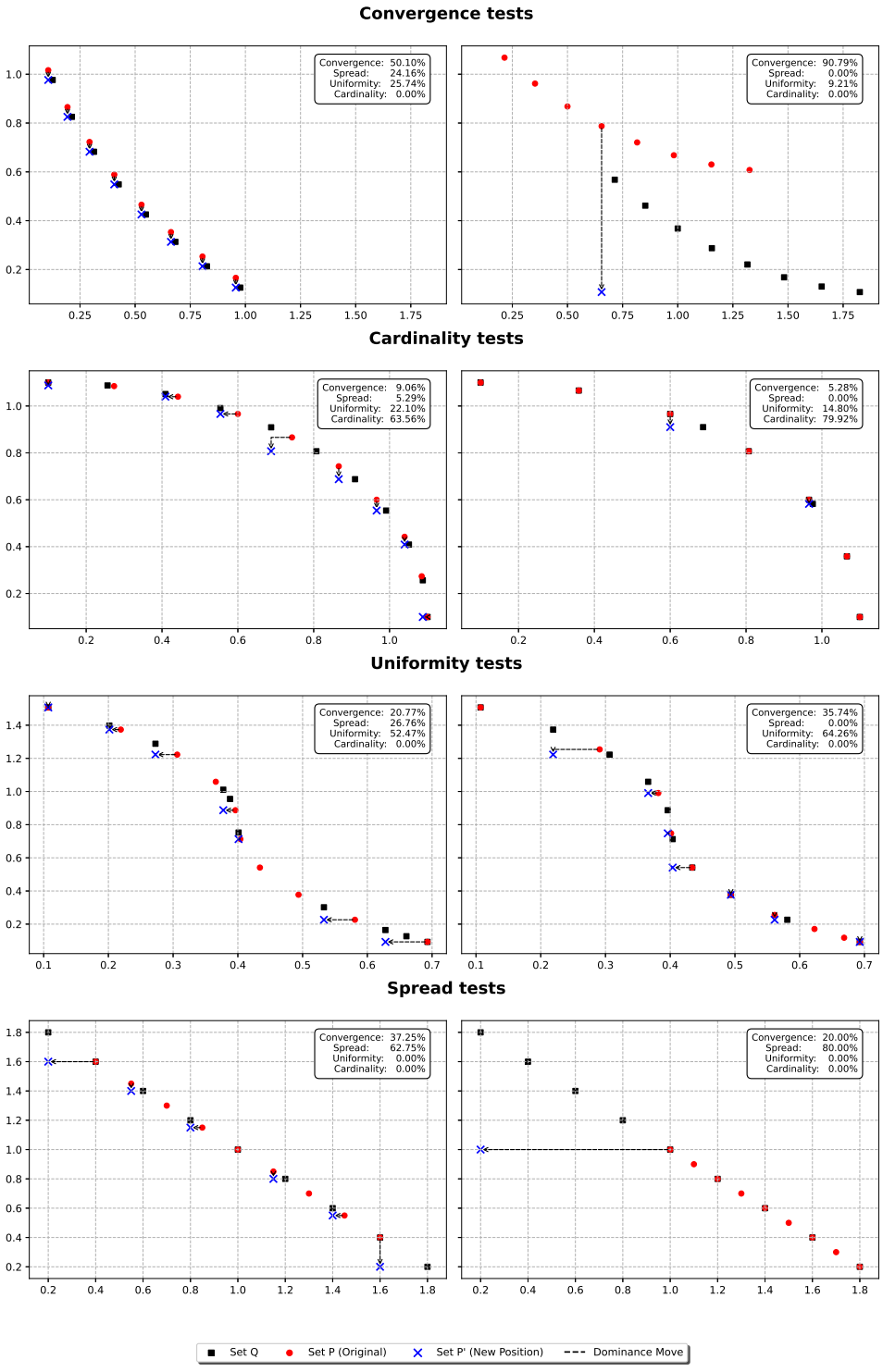


Figure 1. Experimental results showcasing the DoM explainability framework's ability to diagnose specific deficiencies in solution sets. Each row illustrates tests for a distinct facet: Convergence, Cardinality, Uniformity, and Spread. Black squares represent the reference set $\boldsymbol{Q}$, red circles represent the original solution set $\boldsymbol{P}$, and blue 'X' marks denote the new positions ($\boldsymbol{P'}$) for the moved points of $\boldsymbol{P}$ generated by the DoM solver. Dashed gray lines indicate the dominance move. The inset box in each plot displays the percentage contribution of each X-DoM facet to the total DoM score for that scenario.

scalar value, our framework delivers a detailed decomposition that reveals not only which set performs better, but more importantly, why a particular set is deficient. This level of granularity enables the identification of whether the primary shortcoming lies in convergence, cardinality, uniformity, or spread. In doing so, the framework transforms the Dominance Move from a purely comparative performance metric into an actionable diagnostic tool, offering researchers deeper

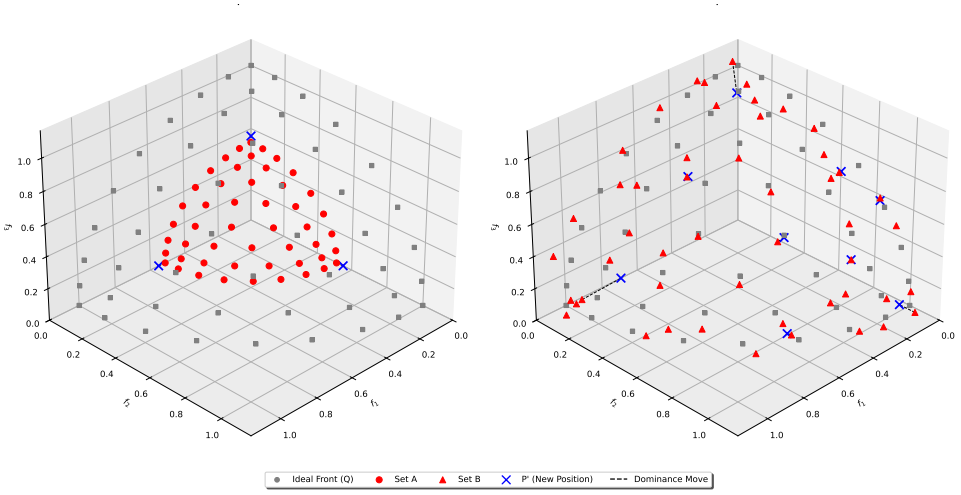


Figure 2. Visualization of the solution sets used in the Deceptive Convergence Trap experiment. The reference front $\boldsymbol{Q}$ (gray squares) is compared against set $\boldsymbol{A}$ (left, red circles) and set $\boldsymbol{B}$ (right, red triangles). The blue 'X' markers ($\boldsymbol{P'}$) and dashed lines show the corrective moves determined by the DoM solver.

and more interpretable insights into algorithmic behavior and performance.

B. When close is not enough: Evaluating spread in high-convergence sets

This experiment investigates how different quality indicators manage the trade-off between strong local convergence and poor global spread. Indicators that aggregate proximity measures, such as the Inverted Generational Distance Plus (IGD+) [3], can be biased toward solution sets that are highly converged but cover only a small region of the true Pareto front. Consequently, IGD+ tends to favor spatially compressed yet locally optimal sets, while penalizing those that offer broader coverage but slightly lower convergence. In contrast, the X-DoM framework provides a more balanced evaluation by explicitly identifying and quantifying the lack of spread in compressed sets, thereby exposing the structural advantages of more expansive, well-distributed solution sets.

The experiment was conducted using the DTLZ2 benchmark problem with $M = 3$ objectives, which features a continuous, concave Pareto front well-suited for visualizing spread [19]. As illustrated in Figure 2, we constructed two distinct solution sets, $\boldsymbol{A}$ and $\boldsymbol{B}$, for comparison against a reference set $\boldsymbol{Q}$ of 50 uniformly distributed points on the true Pareto front. Set $\boldsymbol{A}$, the compressed front, consists of 50 points clustered in the central region of the Pareto front. By design, these points exhibit excellent local convergence but suffer from limited spread. In contrast, set $\boldsymbol{B}$, the expanded front, also contains 50 points but is distributed across the full extent of the Pareto front. These points achieve complete coverage but are less uniform and have a greater average distance to the true front, reflecting a trade-off between convergence, uniformity, and overall spread.

The quantitative results of the comparison are summarized in Table I. The IGD+ indicator reported scores of 0.0218 for set $\boldsymbol{A}$ and 0.0220 for set $\boldsymbol{B}$. Based on these nearly identical values, a researcher might conclude that the two sets are of virtually equivalent quality. This illustrates the so-called "deceptive trap": the aggregated nature of IGD+ masks the substantial structural differences between the sets.

By contrast, the total DoM score provides a clearer distinction. Set $\boldsymbol{A}$ achieved a DoM score of 0.249, whereas set $\boldsymbol{B}$ received 0.766, indicating that $\boldsymbol{A}$ is superior according to this metric. The X-DoM facet breakdown further clarifies this result, revealing the specific contributions to the total cost and explaining why each set performs differently.

Table I
PERFORMANCE COMPARISON FOR THE DECEPTIVE CONVERGENCE TRAP EXPERIMENT.

Solution set (Type)	**IGD+** ↓	**DoM** ↓	**Conv.** (%)	**Spread** (%)	**Unif.** (%)	**Card.** (%)
Set $\boldsymbol{A}$ (Compressed)	**0.0218**	**0.249**	4.33	52.19	43.49	0.00
Set $\boldsymbol{B}$ (Expanded)	0.0220	0.766	33.87	0.00	66.13	0.00

The facet decomposition presented in Table I provides a clear diagnostic explanation of each set's performance. For set $\boldsymbol{A}$, the DoM cost is primarily attributed to structural deficiencies, with spread accounting for 52.19%. This indicates that, although the points in $\boldsymbol{A}$ are very well converged, the set is heavily penalized for its limited coverage of the Pareto front and uneven internal distribution. This observation aligns with the visualization in Figure 2 (left), where corrective moves extend the set's boundaries outward. In contrast, set $\boldsymbol{B}$ exhibits a different diagnostic profile. Its cost is dominated by uniformity (66.13%), while the spread contribution is zero, confirming that the set effectively covers the entire front. The high DoM score arises from poor internal spacing and the collective distance of the points from the true front, as illustrated by the inward corrective moves in Figure 2 (right).

While the IGD+ indicator can be ambiguous, the X-DoM framework provides a clear and explainable assessment. It reveals that sets $\boldsymbol{A}$ and $\boldsymbol{B}$ exhibit distinct, quantifiable weaknesses. Set $\boldsymbol{A}$, despite its excellent local convergence, is structurally deficient due to limited coverage and uneven distribution. In contrast, set $\boldsymbol{B}$ achieves full coverage of the

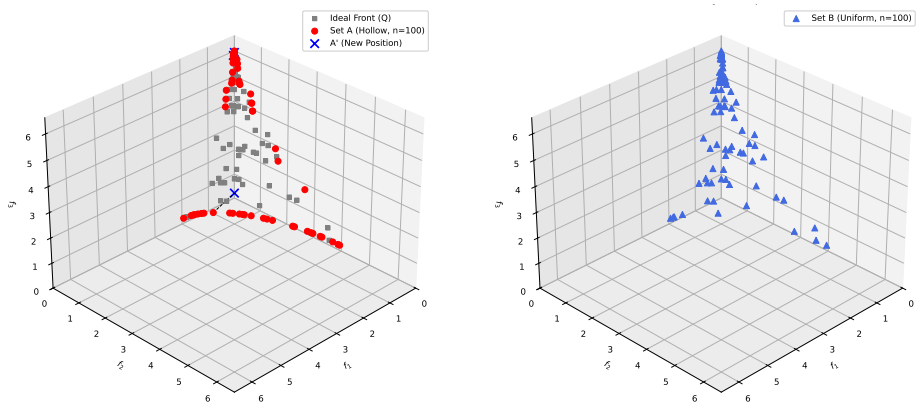


Figure 3. Visualization of the solution sets for the Hollow Front Illusion experiment on WFG2. The reference front $\boldsymbol{Q}$ (gray squares) is shown on the left. Set $\boldsymbol{A}$ (left, red circles) covers the boundaries but has a large central gap. The blue 'X' marks indicate the new positions ($\boldsymbol{A}'$) after the DoM calculation. set $\boldsymbol{B}$ (right, blue triangles) represents a uniformly sampled front.

front but is penalized for suboptimal placement and poor internal spacing of its points.

C. The hollow front illusion: Detecting structural flaws beyond local uniformity

This experiment examines the ability of quality indicators to differentiate between local uniformity and global structural integrity. We hypothesize that indicators focused on extensity and local spacing, such as Maximum Spread (MS) and Spacing (SP), may overlook significant internal gaps in a solution set that otherwise covers the front's boundaries. In contrast, the DoM explainability framework is expected to accurately detect such large-scale structural deficiencies through its uniformity facet, which evaluates the global distribution of solutions.

The WFG2 test problem with $M = 3$ objectives was chosen for this experiment due to its convex Pareto front, which makes the interior region both distinct and critical for a complete representation [19]. Three solution sets, each containing 100 points, were constructed and can be visualized in Figure 3: the reference set $\boldsymbol{Q}$, uniformly sampled from the true WFG2 Pareto front to represent an ideal distribution; set $\boldsymbol{A}$, the "hollow front", with points located exclusively along the front's boundaries, intentionally leaving a large empty region at the center; and set $\boldsymbol{B}$, "the uniform front", sampled evenly across the entire front, mirroring the structure of $\boldsymbol{Q}$.

The results of traditional diversity indicators were examined first. The Maximum Spread (MS) values for sets $\boldsymbol{A}$ and $\boldsymbol{B}$ were identical, since both sets were constructed to include the same extreme points as the reference front $\boldsymbol{Q}$. Consequently, MS alone fails to differentiate between the structurally deficient hollow set and the uniformly distributed set.

The results of the Spacing (SP) indicator were even more revealing. Contrary to expectation, the flawed set $\boldsymbol{A}$ achieved a lower (better) SP score of approximately 0.194, compared to 0.219 for the uniform set $\boldsymbol{B}$. Figure 4 illustrates the reason: the nearest-neighbor distances in $\boldsymbol{A}$ form a bimodal distribution, with a large cluster of very small distances along the boundary points and a few very large distances that span the central gap. Because SP measures the standard deviation of these distances, this bimodal pattern produces a smaller value than the more uniformly distributed nearest-neighbor distances in $\boldsymbol{B}$. This outcome highlights how an indicator focused solely on local spacing variance can produce a misleading assessment of global uniformity. Consequently, relying on MS and SP alone would erroneously favor the hollow front over the truly uniform set.

The analysis of DoM($\boldsymbol{A}$, $\boldsymbol{Q}$) using the X-DoM framework provides a more precise and informative diagnosis. The total DoM score was 1.507, with the following facet breakdown: convergence 40.47%, spread 0.00%, uniformity 59.53%, and cardinality 0.00%.

The X-DoM results accurately highlight the primary issues with set $\boldsymbol{A}$. The zero contribution from spread confirms that the set fully covers the front's boundaries. In contrast, the uniformity facet accounts for the majority of the cost (59.53%), explicitly diagnosing the structural deficiency created by the hollow interior. A secondary contribution of 40.47% from convergence indicates that the points are not perfectly aligned with the true front. This analysis demonstrates that a diagnostic framework like X-DoM provides deeper insight than traditional quality indicators, capturing global structural flaws rather than focusing solely on local spacing.

D. The role of cardinality in high-dimensional fronts

This experiment aims to highlight the distinctive, M-sensitive cardinality facet of the X-DoM framework, demonstrating its effectiveness in many-objective scenarios where maintaining a sufficient number of solutions is critical to adequately represent the Pareto front [20]. It is expected that in high-dimensional spaces, a substantial shortfall in the number of solutions represents a major failure. While scalar indicators such as hypervolume may yield ambiguous assessments in such cases, the X-DoM framework is expected to explicitly and accurately diagnose deficiencies arising from inadequate cardinality.

The experiment was conducted using the DTLZ1 benchmark problem with $M = 10$ objectives. Three distinct solution sets were generated for analysis. The reference set, $\boldsymbol{Q}$, comprises

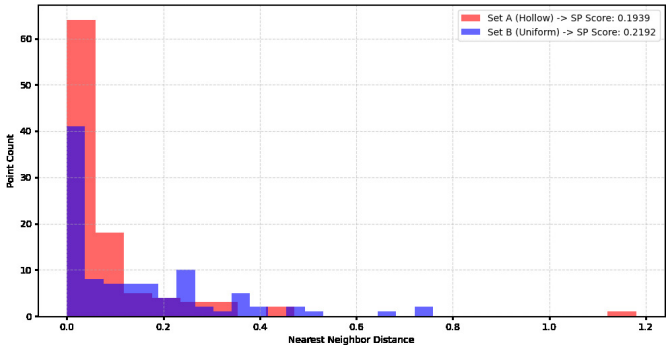


Figure 4. Histogram of nearest-neighbor distances for set $\boldsymbol{A}$ (Hollow) and set $\boldsymbol{B}$ (Uniform). set $\boldsymbol{A}$'s distribution is highly skewed towards very small distances, resulting in a deceptively better Spacing (SP) score despite its large central gap.

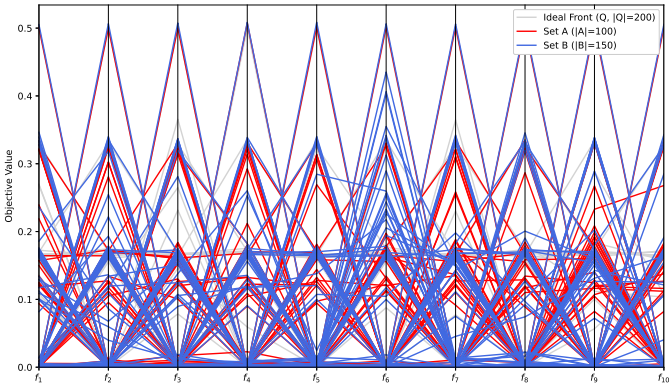


Figure 5. Parallel coordinate plot for the cardinality issue experiment on DTLZ1 ($M = 10$). The ideal front $\boldsymbol{Q}$ ($|\boldsymbol{Q}| = 200$) is shown in light gray. set $\boldsymbol{A}$ ($|\boldsymbol{A}| = 100$, red) and set $\boldsymbol{B}$ ($|\boldsymbol{B}| = 150$, blue) are plotted against it. The plot visually suggests that while the set $\boldsymbol{B}$ is denser, the solutions in the set $\boldsymbol{A}$ may exhibit better overall distribution.

$|\boldsymbol{Q}| = 200$ points uniformly sampled from the true Pareto front. Two test sets, $\boldsymbol{A}$ and $\boldsymbol{B}$, were created using the MuSt-EMaO framework [21], which allows precise control over the number of solutions and ensures a high degree of uniformity within each set:

- Set $\boldsymbol{A}$ generated with 100 solutions (a 50% cardinality deficit relative to $\boldsymbol{Q}$) using a budget of 100,000 function evaluations to promote high-quality convergence and spread.
- Set $\boldsymbol{B}$ generated with 150 solutions (a 25% cardinality deficit) using a smaller budget of 80,000 function evaluations.

The resulting solution sets are visualized in Figure 5 using a Parallel Coordinate Plot (PCP), where each polyline represents a single solution and each vertical axis corresponds to an objective. The hypervolume (HV) indicator was first used to assess performance. The obtained values were nearly identical: $\text{HV}(\boldsymbol{A}) \approx 0.999993$ and $\text{HV}(\boldsymbol{B}) \approx 0.999992$. As expected, this outcome is inconclusive and provides little diagnostic insight. The superior convergence and placement of the fewer solutions in set $\boldsymbol{A}$ almost entirely offset the reduction in covered volume due to its lower cardinality, making it difficult to draw any meaningful distinction between the sets using HV alone.

The DoM scores and their corresponding X-DoM facet decompositions, summarized in Table II, offer a far more transparent and informative analysis. Although the total DoM values for both sets are coincidentally similar, 0.502 for set $\boldsymbol{A}$ and 0.506 for set $\boldsymbol{B}$, the underlying sources of these costs differ markedly.

Table II
PERFORMANCE COMPARISON FOR THE CARDINALITY ISSUE EXPERIMENT.

Solution set (Cardinality)	**DoM** ↓ (Total)	**Conv.** (%)	**Spread** (%)	**Unif.** (%)	**Card.** (%)
Set $\boldsymbol{A}$ ($\lvert\boldsymbol{A}\rvert = 100$)	0.502	8.99	0.00	1.20	**89.81**
Set $\boldsymbol{B}$ ($\lvert\boldsymbol{B}\rvert = 150$)	0.506	4.73	**47.27**	0.91	47.09

For set $\boldsymbol{A}$, the X-DoM framework attributes 89.81% of the total cost to the cardinality facet. This provides a clear diagnosis: the set's primary weakness lies in its insufficient number of solutions. The minimal contributions from convergence (8.99%) and uniformity (1.20%) confirm that the existing 100 points are well positioned and evenly distributed. However, they are simply too few to adequately represent the complexity of a 10-dimensional Pareto front.

In contrast, the diagnosis for set $\boldsymbol{B}$ reveals a markedly different profile. Its total cost is almost evenly divided between cardinality (47.09%) and spread (47.27%). This indicates that,

although $\boldsymbol{B}$ contains more solutions than $\boldsymbol{A}$, it fails to fully capture the extents of the reference front. This interpretation aligns with the visualization in Figure 5, where the red polylines representing set $\boldsymbol{A}$ span the objective ranges more completely than the blue polylines of set $\boldsymbol{B}$, despite being fewer in number.

The X-DoM framework provided a clear and explainable diagnosis for each set. It accurately identified that set $\boldsymbol{A}$'s primary weakness was its severe shortage of solutions, a critical limitation in many-objective optimization, whereas set $\boldsymbol{B}$ exhibited a combination of a moderate cardinality deficit and a pronounced spread deficiency. This level of granular insight offers a powerful means of interpreting algorithmic performance, revealing not just which set performs better, but precisely why each set falls short.

E. Diagnosing parameter tuning trade-offs

This experiment demonstrates the utility of the X-DoM framework in algorithm development, particularly for uncovering the nuanced effects of parameter tuning. While scalar quality indicators such as hypervolume can identify an apparently optimal parameter configuration, they offer no insight into the underlying performance trade-offs, for instance, improving convergence at the expense of spread. In contrast, the X-DoM framework reveals how these trade-offs evolve across the parameter space, delivering interpretable and actionable feedback to guide researchers in fine-tuning their algorithms.

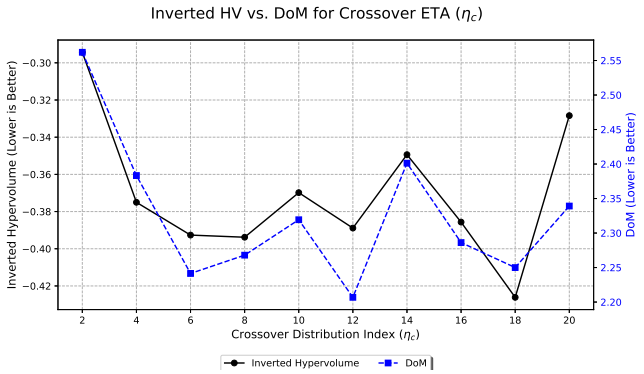


Figure 6. Performance of NSGA-II on DTLZ7 across different η_c values, measured by Inverted Hypervolume and Total DoM. Both indicators suggest that mid-range η_c values yield the best overall performance, but they do not explain the underlying reasons for performance degradation at the extremes.

We employed the NSGA-II algorithm [22] on the DTLZ7 benchmark problem with $M = 3$ objectives. DTLZ7 is particularly well-suited for this analysis, as its Pareto front comprises four disconnected regions, posing a challenge that requires balancing exploration to locate all regions and exploitation to refine solutions within them [23]. The primary parameter under investigation was the crossover distribution index (η_c) of the Simulated Binary Crossover (SBX) operator, which directly influences this balance. Lower values of η_c promote exploration by generating offspring farther from their parents, whereas higher values enhance exploitation by producing offspring in closer proximity to them.

The algorithm was executed 21 times for each η_c value in the range [2, 20], using a population size of 50 over 50 generations, implemented in the Pymoo framework [24]. For each run, the resulting population $\boldsymbol{P}$ was compared against the true Pareto front $\boldsymbol{Q}$. We computed the inverted hypervolume, the total DoM, and performed the X-DoM decomposition. The aggregated findings are discussed in the following analysis.

Figure 6 presents the inverted hypervolume and total DoM scores as functions of η_c. Both scalar indicators exhibit a similar pattern: performance degrades at very low and very high η_c values, while optimal or near-optimal results occur in the mid-range (approximately $\eta_c \in [6, 12]$). Although these scalar results can guide the selection of a suitable parameter value, they fail to reveal the underlying algorithmic dynamics responsible for the observed performance. For example, it remains unclear whether the degradation at $\eta_c = 2$ and $\eta_c = 20$ arises from the same type of deficiency or from distinct failure mechanisms.

The X-DoM facet breakdown, shown in Figure 7, provides the explanatory insight missing from scalar indicators. The analysis reveals a systematic trade-off between convergence and spread as η_c is varied:

- Low η_c values (e.g., 2 and 4): The DoM score is dominated by the convergence facet, accounting for 42.1% and 47.1% of the total cost, respectively. This indicates that the highly explorative search succeeds in locating the disconnected regions of the front but fails to converge adequately within them.
- High η_c values (e.g., 16, 18, and 20): The primary contributors shift to the spread and uniformity facets. At $\eta_c = 20$, these facets together represent over 52% of the DoM cost, demonstrating that the highly exploitative search converges rapidly within a limited set of regions but cannot maintain sufficient diversity to cover all disconnected sections of the front
- The optimal range for DoM and Hypervolume, respectively, $\eta_c = 12$ and $\eta_c = 18$, aligns with the values where a balance is achieved between the convergence and spread/uniformity costs, highlighting the trade-off captured by the X-DoM decomposition.

Visual inspection of the final solution sets for selected η_c values, shown in Figure 8, confirms the X-DoM diagnosis. At $\eta_c = 6$, the solutions are distributed across all four regions of the front but remain visibly distant from the true Pareto front ($\boldsymbol{Q}$), reflecting the high convergence cost. In contrast, at $\eta_c = 18$, the solutions are tightly converged but occupy only two of the four regions, providing direct evidence of the spread deficiency identified by X-DoM. The solution set for $\eta_c = 12$ demonstrates a balanced outcome, with solutions covering all regions and positioned relatively close to the true front.

This experiment illustrates that the X-DoM framework functions as a powerful diagnostic tool for parameter tuning. It goes beyond identifying apparently optimal settings by explaining why they are optimal, revealing the underlying trade-offs in algorithmic behavior.

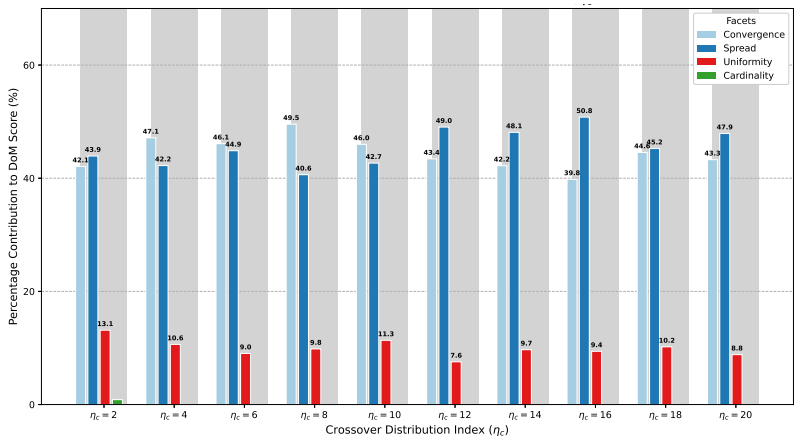


Figure 7. X-DoM facet contributions for NSGA-II on DTLZ7 across different η_c values. The plot shows a clear trade-off: at low η_c, the cost is driven by poor convergence, while at high η_c, the cost is driven by poor spread and uniformity.

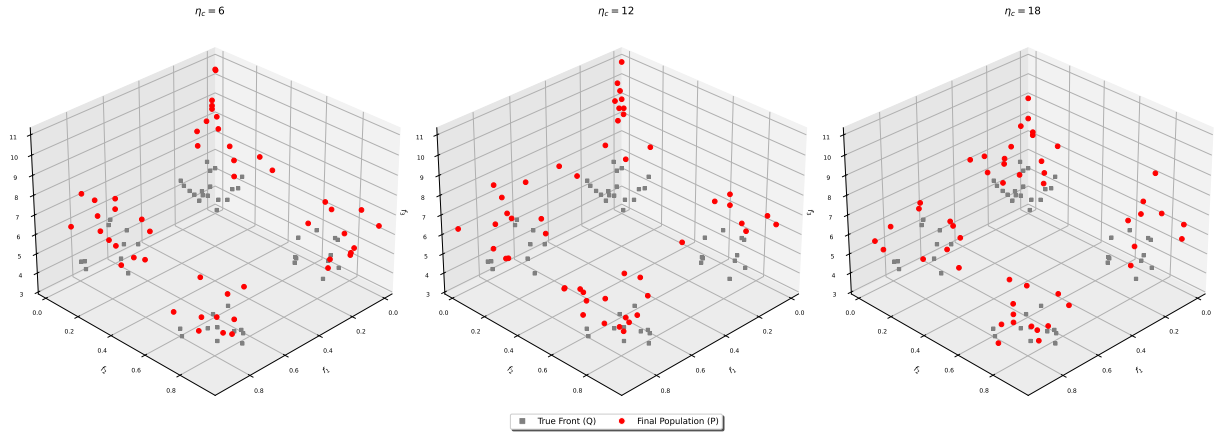


Figure 8. Final populations ($\boldsymbol{P}$, red circles) for selected η_c values on DTLZ7, plotted against the true front ($\boldsymbol{Q}$, black dots). The plots visually confirm the X-DoM diagnosis: low η_c (left) achieves good spread but poor convergence, while high η_c (right) achieves good local convergence but fails to find all regions of the front, resulting in poor spread.

F. Diagnosing algorithm problem mismatch on a degenerate front

In this experiment, we demonstrate how the X-DoM framework can diagnose the underlying reasons for differing performance patterns among algorithm classes on problems with distinctive geometric features. When even high-performing algorithms struggle with unusual Pareto front geometries, such as degeneracy, the X-DoM facet breakdown identifies the root cause of failure, linking it to each algorithm's inherent search biases and tendencies.

The experiment was conducted on the WFG3 test problem with $M = 3$ objectives, which is known to be challenging because its Pareto front is a degenerate, one-dimensional linear segment within a three-dimensional objective space [23]. We compared two prominent algorithmic paradigms: the MOEA/D based on decomposition and the NSGA-III based on reference vector [25]. Both algorithms use the MuSt-EMaO framework [16] to ensure a fair comparison, generating a final population of $|\boldsymbol{P}| = 50$ solutions within a budget of 10,000 function evaluations. The results were aggregated over 21 independent runs, and all calculations used the true linear Pareto front as the reference set $\boldsymbol{Q}$. Figure 9 provides a visual example of the final solution sets from a single run for each algorithm.

An initial assessment using standard scalar indicators suggests that NSGA-III outperforms MOEA/D on this problem. Specifically, NSGA-III achieved a lower DoM value of 3.1340 compared to MOEA/D's 3.2411, and a higher hypervolume of 5.5950 versus MOEA/D's 2.6073. While both metrics agree on the ranking, they provide no insight into why NSGA-III performs better. These scalar values capture the magnitude of the performance difference but fail to reveal the underlying strengths and weaknesses of each algorithm in addressing the degenerate Pareto front. For researchers, knowing which algorithm is superior is often insufficient; understanding the source of that advantage is crucial for improving and refining algorithms.

This limitation is precisely what the X-DoM framework is designed to overcome. Decomposing the DoM value enables a transition from simple performance ranking to a detailed, explanatory analysis of each algorithm's behavior.

The X-DoM framework provides a comprehensive diag-

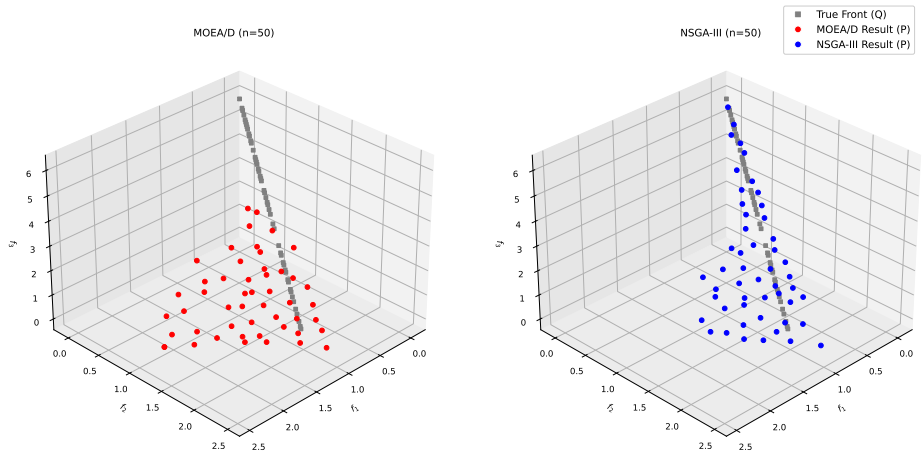


Figure 9. Example of final solution sets on the degenerate WFG3 problem. The true linear Pareto front ($\boldsymbol{Q}$) is shown as gray squares. The left plot shows the result for MOEA/D ($\boldsymbol{P}$, red circles), and the right plot shows the result for NSGA-III ($\boldsymbol{P}$, blue circles). Both algorithms struggle to keep solutions on the narrow front.

nostic breakdown, as summarized in Table III. The analysis reveals that, for both algorithms, the DoM cost is overwhelmingly dominated by the convergence facet (73.81% for NSGA-III and 68.35% for MOEA/D). This clearly identifies the primary challenge of the WFG3 problem: forcing solutions onto the narrow, degenerate front. Both algorithms' search mechanisms result in solutions drifting into the surrounding dominated space, a behavior that is visually confirmed in Figure 9.

Table III
AGGREGATED PERFORMANCE AND X-DOM BREAKDOWN ON WFG3 ($M = 3$) OVER 21 RUNS.

Algorithm	**DoM** ↓ (Mean)	**Conv.** (%)	**Spread** (%)	**Unif.** (%)	**Card.** (%)
MOEA/D	3.241	68.35	**19.31**	12.34	0.00
NSGA-III	3.134	**73.81**	9.32	16.87	0.00

The X-DoM analysis also uncovers subtle but meaningful differences in the algorithms' secondary failure modes. MOEA/D exhibits a notably higher spread cost (19.31%) compared to NSGA-III (9.32%), indicating that NSGA-III's reference-vector-based mechanism is more effective at covering the full extent of the linear front. In contrast, NSGA-III shows a slightly higher uniformity cost (16.87% vs. 12.34%), suggesting that while it spans the front more completely, its solutions are less evenly distributed.

For both algorithms, the dominant challenge on WFG3 remains convergence, driven by the problem's degenerate geometry. Beyond this primary difficulty, X-DoM provides mechanistic insight into their secondary weaknesses, linking NSGA-III's limitations to uniformity and MOEA/D's to spread. This demonstrates the framework's ability to move past simple ranking, offering a clear and actionable diagnosis of how each algorithm's structure interacts with the specific geometry of a problem.

G. Quantifying the "explanatory masking" effect of the cardinality facet

While the X-DoM framework provides deep diagnostic insights, some side effects must be pointed out here and possibly must be the basis for future work. A critical design feature of the framework is the dimension-sensitive scaling of the cardinality facet, which can introduce a side effect, and we term "explanatory masking" in high-dimensional spaces. This experiment aims to quantify this phenomenon and highlight a potential interpretation pitfall.

The X-DoM Cardinality factor (F_{card}) is designed to penalize solution deficits exponentially as the number of objectives M increases, reflecting the criticality of population size in many-objective optimization. We hypothesize that in high-dimensional scenarios (e.g., $M \geq 6$), this penalty becomes so dominant that even a minor cardinality deficit, such as 2%, can overshadow other severe, actionable flaws, such as poor convergence. This masking effect could lead practitioners to misdiagnose the primary algorithmic weakness.

We utilized the DTLZ2 benchmark problem, varying the number of objectives $M \in \{2, 3, 4, 6, 8\}$. For each M, a reference set $\boldsymbol{Q}$ was generated on the true Pareto front with a base cardinality N_{base} appropriate for the dimension (ranging from 50 for $M = 2$ to 150 for $M = 8$). We then constructed a flawed test set $\boldsymbol{P}$ by applying two simultaneous deficiencies.

The primary flaw, the convergence error, was induced by shifting all points away from the front by a fixed distance $\delta = 0.1$. Secondary Flaw, we introduced variable cardinality deficits by randomly removing $\{0\%, 2\%, 5\%, 10\%, 20\%\}$ of the points from $\boldsymbol{P}$. The 0% deficit case serves as a control where only the convergence flaw exists. We analyzed how the percentage contribution of the X-DoM facets shifts as M increases.

Table IV summarizes the X-DoM facet breakdowns from the experiment, obtained with 21 trials. In low dimensions ($M = 2$), the framework behaves as expected. At 0% deficit, the cost is primarily driven by the spread facet at 57.78%, with convergence playing a secondary role. As the cardinality

Table IV
QUANTIFYING THE EXPLANATORY MASKING EFFECT ON DTLZ2. THE TABLE SHOWS THE MEAN DOM SCORE AND THE X-DOM FACET BREAKDOWN AS THE CARDINALITY DEFICIT INCREASES. IN HIGH DIMENSIONS ($M \geq 6$), THE CARDINALITY FACET DISPROPORTIONATELY DOMINATES THE EXPLANATION (BOLDED), MASKING THE UNDERLYING CONVERGENCE AND SPREAD ISSUES.

M	Deficit (%)	Mean DoM ↓	Conv. (%)	Spread (%)	Unif. (%)	Card. (%)
2	0 (Control)	0.945	13.43	**57.78**	28.79	0.00
	2	0.961	12.81	**54.31**	27.42	5.46
	5	0.960	12.73	**53.96**	25.61	7.71
	10	0.965	12.07	**50.98**	25.24	11.70
	20	1.003	12.20	**49.02**	22.45	16.34
4	0 (Control)	1.428	38.10	22.50	**39.40**	0.00
	2	1.428	21.60	18.92	22.65	**36.83**
	5	1.428	20.91	16.43	22.60	**40.07**
	10	1.443	20.65	16.17	20.62	**42.57**
	20	1.472	20.41	15.37	19.11	**45.10**
6	0 (Control)	1.641	39.42	**50.67**	9.91	0.00
	2	1.641	14.90	30.48	4.32	**50.29**
	5	1.641	14.61	29.97	4.39	**51.03**
	10	1.644	14.44	29.83	4.36	**51.37**
	20	1.651	15.38	25.63	3.29	**55.70**
8	0 (Control)	1.863	35.85	**53.78**	10.37	0.00
	2	1.871	12.61	28.96	4.41	**54.02**
	5	1.863	12.14	30.92	4.29	**52.65**
	10	1.871	12.16	31.02	4.18	**52.64**
	20	1.867	12.09	30.56	4.62	**52.74**

deficit increases to 20%, the Cardinality facet contribution rises moderately to 16.34%, but it does not overshadow the structural flaws; Spread remains the dominant diagnosable issue, 49.02%.

However, the transition to higher dimensions reveals the masking effect. At $M = 4$, the shift begins: a mere 2% deficit causes cardinality to become the leading factor, 36.83%, overtaking uniformity, which drops from 39.40% to 22.65%.

In many-objective scenarios, $M \geq 6$, this effect becomes pronounced. At $M = 6$, introducing a minimal 2% cardinality deficit causes the cardinality facet contribution to surge from 0% to 50.29%, instantly becoming the dominant explanation. This drastically suppresses the diagnosis of other flaws: the spread contribution drops from 50.67% to 30.48%, and the convergence contribution is slashed from 39.42% to 14.90%. This effectively masks the fact that the solution set still suffers from significant spread and convergence deficiencies. This sensitivity saturates at $M = 8$, where the cardinality contribution stays consistently high ($\approx 53 - 54\%$) regardless of whether the deficit is 2% or 20%.

The results at $M = 8$ reveal a specific limitation inherent to the high-dimensional formulation: penalty saturation. As observed in Table IV, the cardinality contribution exhibits a sharp, immediate jump to 54.02% at a mere 2% deficit. However, distinct from lower dimensions, this contribution does not scale linearly with the deficit; instead, it plateaus and fluctuates slightly around 52.7% even as the deficit increases to 20%. This behavior arises because the penalty exponent $\eta = 1/2^{M-1}$ becomes minimal in high dimensions (at $M = 8$, $\eta \approx 0.0078$). Consequently, the cardinality factor F_{card} saturates near its maximum value, 1.0, even for negligible solution deficits. Once saturated, the relative percentage contribution becomes sensitive to the specific movement costs of the removed points. For instance, removing a point that required a large corrective move reduces the total DoM score, used in the denominator, which can artificially inflate or deflate the percentage shares of the remaining facets, leading to the observed non-monotonic fluctuations.

While X-DoM correctly signals that a deficit of solutions is a critical failure mode in high-dimensional spaces, this high sensitivity results in "explanatory masking". Researchers utilizing the framework for problems with $M \geq 6$ must be aware that a dominant cardinality score may be obscuring significant, concurrent flaws in convergence or diversity. In such high-dimensional scenarios, we recommend a multi-layered diagnostic approach: acknowledge the cardinality warning, but also inspect the raw move costs or re-evaluate the set with the cardinality facet temporarily suppressed to expose underlying structural deficiencies. This saturation phenomenon and masking effect were consistently observed in the DTLZ7 problem as well, the details of which are provided in the Supplementary Material.

V. CONCLUSION AND FUTURE WORKS

In this paper, we addressed a key limitation of traditional quality indicators in multi- and many-objective optimization: their inability to explain why one solution set outperforms another. While scalar indicators are useful for ranking, they provide little insight into the specific strengths and weaknesses of an algorithm. To overcome this, we introduced the Explainable Dominance Move (X-DoM) framework, which decomposes the comprehensive DoM score into four intuitive and quantifiable facets: Convergence, Spread, Uniformity, and Cardinality.

Through a series of targeted experiments, we demonstrated that X-DoM delivers diagnostic clarity beyond conventional indicators. It successfully identifies and quantifies deficiencies such as poor global spread masked by local convergence, internal structural gaps overlooked by spacing measures, and dimensionality-sensitive cardinality deficits in many-objective scenarios. Moreover, X-DoM provides actionable feedback for tasks like algorithm parameter tuning by attributing the DoM cost to its underlying causes, transforming the indicator from a simple ranking tool into a powerful analytical instrument.

The X-DoM framework opens several avenues for future research. A critical area for investigation is the "Explanatory Masking" effect observed in high-dimensional spaces ($M \geq 6$). Our experiments revealed that the dimension-sensitive cardinality penalty can saturate, potentially overshadowing other structural deficiencies like poor convergence or spread. Future work must explore adaptive scaling mechanisms or alternative formulations to mitigate this effect and ensure balanced diagnostics in many-objective scenarios. Additionally, future work could also explore interactions between facets to detect compound failure modes, or introduce additional facets to capture geometric properties such as knee regions or disconnected fronts. Finally, formal user studies could also assess how X-DoM explanations enhance researchers' ability to understand, compare, and improve algorithms, further solidifying its role as a practical tool for the EMO community.

REFERENCES

[1] M. Li and X. Yao, "Quality evaluation of solution sets in multiobjective optimisation: A survey," *ACM Comput. Surv.*, vol. 52, no. 2, Mar. 2019. [Online]. Available: https://doi.org/10.1145/3300148

[2] A. P. Guerreiro, C. M. Fonseca, and L. Paquete, "The hypervolume indicator: Computational problems and algorithms," *ACM Comput. Surv.*, vol. 54, no. 6, Jul. 2021. [Online]. Available: https://doi.org/10.1145/3453474

[3] H. Ishibuchi, H. Masuda, Y. Tanigaki, and Y. Nojima, "Modified distance calculation in generational distance and inverted generational distance," in *Evolutionary Multi-Criterion Optimization*, A. Gaspar-Cunha, C. Henggeler Antunes, and C. C. Coello, Eds. Cham: Springer International Publishing, 2015, pp. 110–125.

[4] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, "Performance assessment of multiobjective optimizers: an analysis and review," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 117–132, April 2003.

[5] C. L. d. V. Lopes, F. V. C. Martins, E. F. Wanner, and K. Deb, "Analyzing dominance move (mip-dom) indicator for multiobjective and many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 3, pp. 476–489, 2022.

[6] ——, "A fast dominance move calculation using mixed-integer programming for many-objective optimization," *IEEE Transactions on Evolutionary Computation*, pp. 1–1, 2026.

[7] M. Garouani, J. Mothe, A. Barhrhouj, and J. Aligon, " Investigating the Duality of Interpretability and Explainability in Machine Learning ," in *2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI)*. Los Alamitos, CA, USA: IEEE Computer Society, Oct. 2024, pp. 861–867. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/ICTAI62512.2024.00125

[8] R. Zhou, J. Bacardit, A. E. I. Brownlee, S. Cagnoni, M. Fyvie, G. Iacca, J. McCall, N. van Stein, D. J. Walker, and T. Hu, "Evolutionary computation and explainable ai: A roadmap to understandable intelligent systems," *IEEE Transactions on Evolutionary Computation*, pp. 1–1, 2024.

[9] B. H. Abed-Alguni, "Evomapx: An explainable framework for metaheuristic optimization algorithms," *Expert Systems with Applications*, vol. 298, p. 129514, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095741742503129X

[10] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 4768–4777.

[11] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1135–1144. [Online]. Available: https://doi.org/10.1145/2939672.2939778

[12] I. Ibs, C. Ott, F. Jäkel, and C. A. Rothkopf, "From human explanations to explainable ai: Insights from constrained optimization," *Cognitive Systems Research*, vol. 88, p. 101297, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389041724000913

[13] G. Misitano, B. Afsar, G. Lárraga, and K. Miettinen, "Towards explainable interactive multiobjective optimization: R-ximo," *Autonomous Agents and Multi-Agent Systems*, vol. 36, no. 2, Oct. 2022. [Online]. Available: https://doi.org/10.1007/s10458-022-09577-3

[14] K.-M. Aigner, M. Goerigk, M. Hartisch, F. Liers, and A. Miehlich, "A framework for data-driven explainability in mathematical optimization," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 19, pp. 20 912–20 920, Mar. 2024. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/30081

[15] G. Misitano, "Exploring the explainable aspects and performance of a learnable evolutionary multiobjective optimization method," *ACM Trans. Evol. Learn. Optim.*, vol. 4, no. 1, Feb. 2024. [Online]. Available: https://doi.org/10.1145/3626104

[16] K. Deb, C. L. d. V. Lopes, F. V. C. Martins, and E. F. Wanner, "Identifying pareto fronts reliably using a multistage reference-vector-based framework," *IEEE Transactions on Evolutionary Computation*, vol. 28, no. 1, pp. 252–266, 2024.

[17] J. R. Schott, "Fault tolerant design using single and multicriteria genetic algorithm optimization." Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Tech. Rep., 1995.

[18] K. Deb and D. Saxena, "On finding pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems," *IEEE Congress on Evolutionary Computation*, 01 2005.

[19] S. Huband, P. Hingston, and L. Barone, "A Review of Multi-objective Test Problems and a Scalable Test Problem Toolkit A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 2006, pp. 477–506, 2011. [Online]. Available: http://ro.ecu.edu.au/ecuworks/2022

[20] K. Deb and D. K. Saxena, "On finding pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems," Tech. Rep., 2005.

[21] K. Deb, C. L. d. V. Lopes, F. V. C. Martins, and E. F. Wanner, "Identifying pareto fronts reliably using a multistage reference-vector-based framework," *IEEE Transactions on Evolutionary Computation*, vol. 28, no. 1, pp. 252–266, 2024.

[22] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[23] I. R. Meneghini, M. A. Alves, A. Gaspar-Cunha, and F. G. Guimarães, "Scalable and customizable benchmark problems for many-objective optimization," *Applied Soft Computing*, vol. 90, p. 106139, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S156849462030079X

[24] J. Blank and K. Deb, "Pymoo: Multi-objective optimization in python," *IEEE Access*, vol. 8, pp. 89 497–89 509, 2020.

[25] H. Li, K. Deb, Q. Zhang, P. Suganthan, and L. Chen, "Comparison between [moea/d] and [nsga-iii] on a set of novel many and multi-objective benchmark problems with challenging difficulties," *Swarm and Evolutionary Computation*, vol. 46, pp. 104 – 117, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2210650218307016