

Universiteit Leiden

Surrogate assisted MOEA/D with adaptive weight adjustment

Master Computer Science

Name:	Santiago García Mayayo
Student ID:	s4450361
Date:	19/06/2026
Specialisation:	Foundations of Computing
1st supervisor:	Dr. Hao Wang
2nd supervisor:	Dr. Elena Raponi

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
2333 CA Leiden
The Netherlands

Abstract

Multi-objective evolutionary algorithms based on decomposition, such as MOEA/D, rely on a set of reference vectors to guide the search towards different regions of the Pareto front. On problems whose Pareto front exhibits irregular geometry, extreme curvature, sharp boundaries, or strongly non-uniform density, fixed uniform reference vectors fail to cover the extreme regions, concentrating the population in a particular portion of the front regardless of the number of generations.

This thesis proposes SA-MOEA/D-AWA, a surrogate-assisted extension of MOEA/D that addresses this limitation in the bi-objective case. Instead of modelling the decision-to-objective mapping as in classical surrogate-assisted methods, we propose to learn the Pareto front (a sub-manifold in the objective space) with surrogate models. Since the training data (the current approximation set) is mutually non-dominated, we can consider the Pareto front a real-valued function supported on a scaled version of the standard simplex Δ_{m-1}. We then apply a surrogate model to approximate this Pareto front function. The hypervolume is then maximised analytically on the surrogate to derive a new set of reference vectors, requiring zero additional evaluations of the true objective functions.

The method is evaluated on 18 benchmark problems from the ZDT, DTLZ, and geometrically challenging families. SA-MOEA/D-AWA outperforms standard MOEA/D on all five ZDT problems, and wins on 10 out of 13 problems with non-standard Pareto-front geometries, including the DTLZ set of problems. Against MOEA/D-AWA, the method wins on 5 out of 8 problems in hypervolume. The results confirm that fitting a surrogate directly in the objective space is an effective strategy for guiding weight adjustment towards geometrically difficult regions of the Pareto front.

Acknowledgments

First, I would like to thank my supervisor, Dr. Hao Wang, for being a key supporter throughout the whole project. Thank you for all your feedback, your confidence, and your commitment, with weekly meetings and even conversations throughout the summer to discuss our possible collaboration. Thank you also for the trust you placed in me beyond the thesis, for choosing me, more than once, to take on responsibilities in your courses, and for personally recommending me to lead an external mathematics recap. That confidence meant a lot to me. To my second supervisor, Dr. Elena Raponi, thank you for taking time out of your busy schedule to supervise my thesis, for joining the weekly meetings, and for providing valuable feedback that helped improve this work. But most of all, thank you for your attitude and generosity, which always helped me push myself with a smile even in stressful or hard moments.

I would not be here without the unconditional support of my family, particularly my parents, my brother, and my girlfriend, Antía. She deserves a special mention for always being there during every up and down of this thesis, and for listening to me through countless hours of me talking about my work, only to conclude with "Yes, yes, triangles, triangles".

I would also like to thank my "abroad" family, a key part not only of my master's but of my whole life these last two years. Special thanks to Melvin, my first friend in the Netherlands and my right hand throughout the entire master's. Thank you for helping me to catch up on all the computer science concepts I did not have in exchange for only chocolate and pizzas.

It is impossible to mention everyone who has helped me during my academic journey; however, Dr. Natalia López Barquilla and Dr. Rafael del Vado Vírseda deserve a big thank you from me. Thank you, Natalia, not only for helping me find the best master's program for me despite your extremely busy schedule, but for believing in me even when I did not, for your involvement at every step, and for all our laughs along the way. Thank you, Rafa, for your help when I was deciding on the master's and choosing my path, and for always being willing to lend a hand when I reached out. Your guidance at that point meant a lot to me.

To conclude, thank you to everyone who was part of my life during my whole academic career. I remember every one of you, and I will always be grateful. Thank you all.

Contents

Chapter 1 Introduction **1**

Chapter 2 Background **2**
2.1 Multi-Objective Optimization . . . 2
2.2 Performance Indicators . . . 3
2.2.1 Distance-based Metrics . . . 3
2.2.2 Hypervolume indicator . . . 4
2.3 Multi-Objective Evolutionary Algorithms . . . 5
2.4 Surrogate-Assisted Evolutionary Algorithms . . . 6

Chapter 3 Related Work **6**

Chapter 4 Approach **7**
4.1 Basic Idea . . . 7
4.2 Phase I: MOEA/D . . . 8
4.3 Objective Space Modeling . . . 9
4.4 Dynamic Weights Adjustment for MOEA/D . . . 10
4.5 Evolution with new weights . . . 10
4.6 Computational Complexity of SA-MOEA/D-AWA . . . 11

Chapter 5 Experiments **12**
5.1 Experimental Setup . . . 12
5.1.1 Benchmark Problems . . . 12
5.1.2 Performance metrics . . . 12
5.1.3 Algorithms for Comparison . . . 13
5.1.4 Parameter Settings . . . 13
5.2 Results . . . 13
5.2.1 Internal Configuration Analysis . . . 13
5.2.2 Performance on Standard Benchmarks . . . 15
5.2.3 Geometrically Challenging Benchmarks . . . 17
5.2.4 Convergence Analysis . . . 17
5.2.5 Visual Analysis of Pareto fronts . . . 18

Chapter 6 Discussion **19**
6.1 Performance characterisation by problem geometry . . . 19
6.2 Neutral scenarios and their causes . . . 19
6.3 The Hypervolume/IGD trade-off . . . 20
6.4 Practical implications . . . 21
6.5 Limitations . . . 22
6.5.1 Dimensionality . . . 22
6.5.2 Hyperparameter tuning . . . 22

Chapter 7 Conclusion **22**

References **25**

Chapter A Convergence plots for the full benchmark suite **26**

Chapter B Final Pareto front approximations for the full benchmark suite 27

Chapter 1 Introduction

Many real-world problems involve more than one objective function to optimise simultaneously. These problems are called multi-objective optimisation problems (MOPs), and can be defined as:

$$\begin{aligned} \min_{x} \quad & f(x) = (f_1(x), f_2(x), \ldots, f_m(x)) \\ \text{s.t.} \quad & x \in \Omega \end{aligned} \tag{1}$$

with m objective functions $f_i(x)\colon R^n \to R$. The feasible space Ω is the subset of R^n, that is, the decision space, that satisfies all the constraints. The image of f is a subset of R^m called the objective space. Since the objectives are typically conflicting, problem (1) does not admit a single optimal solution, but rather a set of trade-offs known as the Efficient Set in the decision space, and its image in the objective space, called the Pareto front. The goal of any solution method is to produce a finite set of points that approximates the Pareto front with good convergence and coverage.

To solve this problem, one classical approach is to transform the MOP into a single objective problem. The Weighted Sum method is a popular choice; however, it often struggles to find solutions in non-convex regions of the Pareto front [7]. Other approaches, such as the ϵ-Constraint method, can overcome this geometric limitation but require prior knowledge from the decision maker and must be solved repeatedly to recover different trade-offs, biasing the search towards the initially preferred regions.

These limitations have motivated a shift towards population-based methods, in particular Evolutionary Algorithms (EAs) [9]. These methods have gained popularity in recent decades due to several advantages: they produce a set of solutions in a single run, handle different types of decision variables, and exhibit global convergence behaviour that prevents them from getting stuck in local optima. Some well-known algorithms, such as SMS-EMOA [1] and MOEA/D [21], have proven effective across a wide range of applications. However, on problems whose Pareto front exhibits irregular geometric characteristics or extreme curvature, sharp boundaries, or strongly non-uniform density, standard decomposition-based methods such as MOEA/D often fail to distribute solutions across the front. The cause is structural, as the uniform reference vectors used to scalarise the problem are not adapted to the actual shape of the front, and the population concentrates in the central regions while leaving the extreme regions uncovered.

Various extensions of MOEA/D have been proposed to adapt the reference vectors during the run. The most prominent is MOEA/D-AWA [18], which redistributes the vectors based on a crowding-based heuristic. Surrogate-assisted methods [5] and Bayesian optimisation [13] offer a different perspective, replacing expensive evaluations of the true objective by predictions from a learned model of the decision-to-objective map. Both lines, however, share a common characteristic: the surrogate model is fitted in the decision space and used to approximate the objective values of unknown candidates.

This thesis proposes a different use of surrogate modelling. Instead of approximating the objective values of decision variables, we fit a surrogate directly in the objective space to model the geometry of the Pareto front itself. The non-dominated solutions of the current population are the training points, and the surrogate represents the front as a continuous curve. To handle the near-vertical regions that destabilise a direct fit, the objective space is rotated by 45° before training the surrogate; this transformation maps the boundary of the Pareto front to a well-defined, single-valued function. Once the surrogate is available,

the set of points that maximises the hypervolume on the surrogate is identified by optimising the HV functional directly on the predicted curve, a computation that requires no additional evaluations of the true objective functions. The resulting points are then used to derive a new set of reference vectors for MOEA/D. The full method, called SA-MOEA/D-AWA, alternates between standard MOEA/D evolution and surrogate-driven weight adjustment, and is designed to reach the extreme regions of the Pareto front that uniform reference vectors do not cover.
The main contributions of this work are the following:

1. An objective-space surrogate model with a 45° rotation that handles near-vertical regions of bi-objective Pareto fronts without instability, guaranteeing that the rotated front is always single-valued.

2. An optimisation procedure to derive hypervolume-optimal reference directions from the surrogate, requiring zero additional evaluations of the true objective functions, the dominant cost in expensive optimisation scenarios.

3. A surrogate-driven weight adjustment with standard MOEA/D, compatible with other decomposition-based variants.

4. A study on 18 benchmark problems, including 11 standard ZDT/DTLZ problems and 7 problems with geometrically challenging Pareto fronts, showing that the proposed method outperforms MOEA/D and MOEA/D-AWA in hypervolume on the problems where Pareto-front geometry is the main source of difficulty.

The remainder of this thesis is organized as follows: Section 2 introduces the necessary background and preliminaries to understand the proposed work. Section 3 discusses the related literature. Section 4 details the proposed methodology, including the hybrid architecture and the geometry-driven surrogate modeling. Section 5 describes the experimental setup and presents the results. Section 6 discusses the performance of the algorithm. Finally, Section 7 concludes the thesis and explores future directions to extend this work.

Chapter 2 Background

2.1 Multi-Objective Optimization

In Multi-Objective Optimization, solutions are compared using a partial order known as Pareto dominance. Let $x_1, x_2 \in \Omega$ be two decision vectors in the feasible space. The solution x_1 dominates x_2 (denoted as $x_1 \prec x_2$) if and only if:

$$\begin{aligned} &f_i(x_1) \leq f_i(x_2) \quad \forall i \in \{1, 2, \ldots, m\}, \text{ and} \\ &f_j(x_1) \neq f_j(x_2) \quad \text{for at least one } j \in \{1, 2, \ldots, m\} \end{aligned} \tag{2}$$

A trivial observation from the definition is that the second condition is equivalent to stating that there has to be at least one j such that $f_j(x_1) < f_j(x_2)$. Therefore, if x_1 dominates x_2, it implies that x_1 is better than or equal to x_2 in all objectives, and strictly better in at least one. We can relax this definition by eliminating the second condition, yielding weak Pareto dominance ($\preceq$). It is important to highlight that it is a partial preorder property in the decision space, as two different decision vectors can have the

same objective values. Consequently, if neither $x_1 \preceq x_2$ nor $x_2 \preceq x_1$, both decision vectors are incomparable, which is denoted as $x_1 \parallel x_2$.
A point $x \in \Omega$ is called efficient if there is no other solution $x' \in \Omega$ that dominates it. The set of all the efficient points in the decision space is known as the Efficient Set (ES), defined as:

$$ES = \{x \in \Omega \mid x' \in \Omega \text{ such that } x' \prec x\} \tag{3}$$

The image of the ES into the objective space is called the Pareto front (PF), or nondominated set:

$$PF = \{f(x) \in R^m \mid x \in ES\} \tag{4}$$

Figure 1 illustrates these concepts in a bi-objective problem. The non-dominated points (blue) lie on the front and form a valid approximation of the PF, while the dominated points (red) are each surpassed by some non-dominated point in both objectives. The reference point (star) is introduced here for later use in the hypervolume computation (Section 2.2.2).

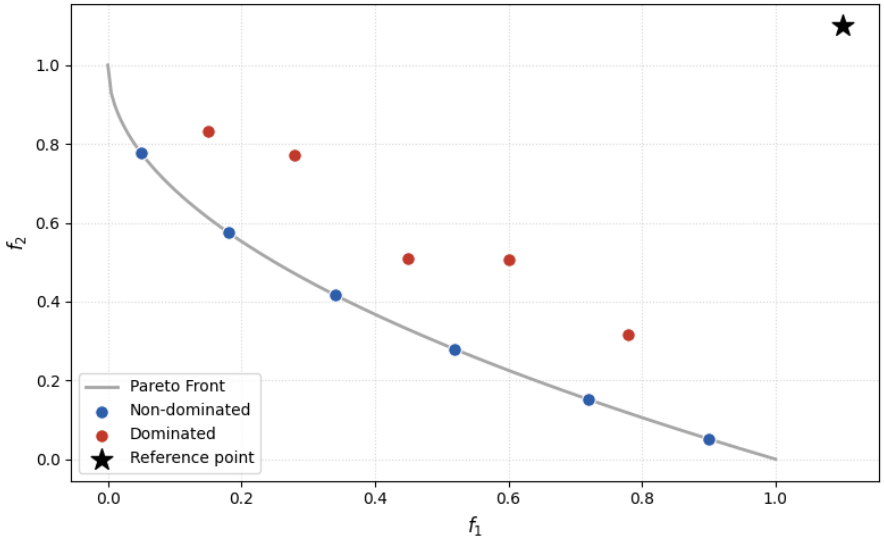


Figure 1: Pareto dominance in a bi-objective problem. Blue points are non-dominated and approximate the Pareto front; red points are dominated, as a non-dominated point is better or equal in both objectives. The reference point (star) bounds the region used for hypervolume.

2.2 Performance Indicators

The goal of any method that solves a Multi-Objective Optimization problem is to find a finite set of solutions that closely approximates the true PF, both in terms of convergence (distance to the true front) and diversity (even distribution along the front). To quantify these properties, various performance indicators have been proposed in the literature.

2.2.1 Distance-based Metrics

Distance indicators measure how far the approximation set S is from the true Pareto front related to a predefined distance function. Among the most common distance metrics, the

Hausdorff distance is widely used. The Generational Distance (GD) [16] measures the average distance from the solutions in our approximation set S to their closest points in the true Pareto front P:

$$GD(S,P) = \frac{1}{|S|} \sum_{s \in S} \inf_{p \in P} ||s - p||_2 \tag{5}$$

This indicator strongly focuses on the convergence of S and is highly robust to outliers, as it gives a great penalty to them. However, it does not penalise poor diversity. On the other hand, the Inverted Generational Distance (IGD) [6] measures the average distance from the points in the true Pareto front P to their closest solutions in the approximation set S:

$$IGD(S,P) = \frac{1}{|P|} \sum_{p \in P} \inf_{s \in S} ||s - p||_2 \tag{6}$$

This metric penalizes both poor convergence and poor diversity, but can lack in penalizing outliers. To obtain a metric that is more robust against outliers, these can be combined into the Average Hausdorff Distance (Δ), defined as the maximum of both distances:

$$\Delta_2(S,P) = \max\{GD(S,P), IGD(S,P)\} \tag{7}$$

However, all these methods rely on prior knowledge of the Pareto front. In real-world problems, the true PF is unknown, making it impossible to measure GD, IGD, or Δ. Therefore, usually an algorithm requires a performance metric that can be evaluated only with the information of the evaluated points.

2.2.2 Hypervolume indicator

The Hypervolume indicator is one of the most widely used metrics in evolutionary multi-objective optimization, especially in lower dimensions. Given a set of solutions S and a predefined reference point r, the Hypervolume measures the Lebesgue measure of the set dominated by the approximation set, delimited by the reference point, since the objective space is otherwise unbounded:

$$HV(S,r) = \Lambda(\{z \in R^m : \exists\, s \in S,\ s \preceq z \preceq r\}) \tag{8}$$

The selection of this reference point, typically derived from the estimation of the Nadir objective vector [10], is crucial for our study, and it can lead to a different subset of points that the algorithm prioritizes depending on where it is.
An important theoretical property of the HV is that it is Pareto-compliant. This means that if an approximation set A strictly dominates another set B (i.e., every solution in B is strictly dominated by at least one solution in A and is denoted as $A \triangleleft B$), then the hypervolume of A will be strictly greater than that of B. It is important to highlight that the reverse is not true; it is not Pareto-compatible. A higher hypervolume does not necessarily guarantee that set A dominates set B. However, Zitzler [24] proved that it is not possible to define a comparison function that is Pareto-compatible and Pareto-compliant at the same time. Therefore, it is reasonable to use this indicator in practice, as an approximation set that maximizes this metric consists only of Pareto optimal elements. To measure the individual impact of a solution, we use the Hypervolume Contribution (HVC), defined as the exclusive volume of the objective space dominated by a single

solution $x \in S$ and by no other solution in the set. In this work, this indicator plays a crucial role in guiding the surrogate-assisted search.
Among the indicators introduced above, IGD and HV are the two used throughout this work, as they capture two different, important complementary notions. Figure 4 illustrates both on the same bi-objective front: IGD (left) measures how far the true-front sample lies from the approximation set, rewarding proximity and an even spread, but requires the true front to be known; HV (right) measures the dominated area bounded by the reference point and can be computed without it.

2.3 Multi-Objective Evolutionary Algorithms

Multi-Objective Evolutionary Algorithms (MOEAs) are population-based methods specially designed for black-box, inexpensive, multi-objective problems. Apart from not requiring prior information about the problem itself, another great advantage that they have over standard methods is that they can operate with a set of solutions in a single run. Also, they can escape from local optima and allow a point to jump from different clusters of a disconnected Pareto front. However, this family of algorithms can have a slow convergence, requiring a great number of function evaluations before obtaining an acceptable result, which is the reason why they need problems that are not expensive to evaluate.
MOEAs are iterative methods, based on crossover, mutation, and selection. First, an existing population of μ individuals is defined. From there, a new set of λ offspring is generated from the variation of the first group. These variation operators mainly consist of recombining the previous population (crossover) with a small random mutation, which gives these algorithms the global convergence property. Then, a selection process determines which individuals survive to the next generation, which will be the initial population μ to restart the loop. Usually, termination criteria are function evaluations or the number of generations.
MOEAs can be classified into three different groups depending on how they rank the points or the general approach to solving the problem:

- Dominance-based algorithms: These methods rank the points following the Pareto dominance relation. The most important example is NSGA-II [8]. It divides the population into a sequence of fronts using a fast non-dominated sorting approach. Later, it uses a secondary diversity measure known as the crowding distance to rank solutions in the same front. This ranking is used in the selection part of the algorithm.

- Indicator-based algorithms: These approaches rank the points using a performance indicator. A state-of-the-art example is SMS-EMOA [1]. It uses the HVC to evaluate and select individuals that are in the same front, trying to maximize the total hypervolume.

- Decomposition-based algorithms: These algorithms reduce the problem into a set of single-objective problems. The most used algorithm of this family is MOEA/D [21]. It generates a set of reference vectors and utilizes scalarization functions, such as the Tchebycheff or Penalty-based Boundary Intersection scalarizations, to optimize these subproblems simultaneously. This approach forces an excellent diversity across

the entire PF. The decomposition structure enforces a structured diversity across the front, making it particularly effective at recovering the general shape of the Pareto front, including regions that dominance-based methods may miss.

The proposed SA-MOEA/D-AWA is an extension of MOEA/D, with adaptive weights adjustments following the directions that maximize the total hypervolume of the set. Then, this method can be considered to be a hybrid indicator-based and decomposition-based algorithm.

2.4 Surrogate-Assisted Evolutionary Algorithms

When function evaluations are computationally expensive, the direct application of MOEAs becomes impractical due to the large number of evaluations required. Surrogate-Assisted Evolutionary Algorithms [5], denoted as SAEAs, address this limitation by replacing some evaluations of the true objective function with predictions from a surrogate model. The surrogate is trained on previously evaluated solutions and used to guide the search towards promising regions.
A particular approach of this family is Bayesian Optimisation [13], denoted as BO, which constructs a probabilistic surrogate model, typically a Gaussian process, and uses an acquisition function to balance exploration and exploitation in selecting the next candidate for evaluation. Classical BO and most SAEAs operate by mapping from the decision space to the objective space, modelling the objective values of unevaluated candidates. While effective for expensive problems, this approach does not explicitly model the geometry of the Pareto front itself. The proposed method takes a different perspective, fitting the surrogate directly in the objective space to reconstruct the shape of the Pareto front, as described in Section 4.

Chapter 3 Related Work

The proposed SA-MOEA/D-AWA is at the intersection of two active research areas: decomposition-based multi-objective optimisation with adaptive weight vectors, and surrogate-assisted evolutionary algorithms for expensive problems.
MOEA/D [21] decomposes a multi-objective problem into a set of scalar subproblems defined by a fixed set of reference vectors, and optimises them simultaneously using neighbourhood-based variation. Its effectiveness depends critically on the alignment between the reference vectors and the actual shape of the Pareto front. When the front is irregular or strongly non-uniform, uniform reference vectors leave extreme regions uncovered and concentrate the population in the central portion of the front. Several extensions address this limitation by adapting the reference vectors during execution. MOEA/D-AWA [18] redistributes vectors based on a crowding-based heuristic, adding vectors in sparse regions and removing them from dense ones. MOEA/D-AV [17] adapts the number of vectors dynamically. RVEA [3] guides the search using angle-penalised distances, with periodic vector adaptation. Decomposition with Pareto adaptive scalarizing methods (PaS) [19] adapts the scalarising function to the estimated front shape. A common characteristic of all these methods is that the vectors are adjusted based on the current population distribution, without any predictive model of where the front lies.

A different line of work studies the case where function evaluations are expensive by replacing some true evaluations with surrogate predictions. ParEGO [15] scalarises the problem and applies Efficient Global Optimisation [13] to each scalarisation. MOEA/D-EGO [22] builds a Gaussian process per MOEA/D subproblem and uses expected improvement to select candidates for true evaluation, combining the decomposition structure of MOEA/D with surrogate-assisted search. K-RVEA [4] combines Kriging surrogates with reference-vector-guided selection to balance convergence and diversity under a limited evaluation budget. All of these methods share a common characteristic that the surrogate is trained from the decision space, modelling the mapping from decision variables to objective values, and is used to identify promising candidates for evaluation rather than to derive reference directions.
SA-MOEA/D-AWA differs from both families along two dimensions. Unlike the weight-adaptation methods above, it derives new reference directions from a predictive model of the Pareto front geometry rather than from the current population distribution. Unlike existing surrogate-assisted methods, the surrogate is fitted directly in the objective space, modelling the shape of the Pareto front itself instead of mapping from the decision to the objective space. The hypervolume is then maximised analytically on this surrogate to produce HV-optimal reference directions at zero additional evaluation cost.

Chapter 4 Approach

This section presents SA-MOEA/D-AWA, a surrogate-assisted extension of MOEA/D that models the Pareto front geometry directly in the objective space to derive hypervolume-optimal reference directions. We describe the four-phase framework of the algorithm, detail each component, and analyse its computational complexity.

4.1 Basic Idea

The proposed method operates through an iterative, four-phase framework designed to bridge global evolutionary exploration with the surrogate-driven exploitation. The algorithm can be divided into four phases:
In the first phase, a standard Multi-Objective Evolutionary Algorithm can be used to drive the population close to the true Pareto front. This preliminary convergence phase is essential to discover the general geometry and boundaries of the objective space before applying any mathematical modeling.
In the second phase, we predict the PF with the non-dominated points of the current population. Unlike traditional surrogate-assisted methods that create regression models mapping the decision space to the objective space, our approach directly models the geometry in the objective space itself, finding a curve that can closely interpolate our points. The surrogate model is a configurable component of the framework; this work evaluates Radial Basis Functions (RBF) and B-Splines as candidates, with the best-performing variant selected through the internal configuration study described in Section 5.2.1.
In the third phase, we compute the set of points in the predicted PF that maximizes the hypervolume. Once Z_{max} is identified, an ideal point z^* is computed from the minimum observed values of each objective. The new reference vectors for MOEA/D are then derived by normalising the displacement from z^* to each point in Z_{max}, as detailed in Section 4.4.

We conclude by allowing MOEA/D to evolve the population from our last iteration along these new directions for several iterations. We repeat the process from the second phase, continuously refining the surrogate model and allowing it to discover new promising directions. [H] SA-MOEA/D-AWA **Input:** CA: Convergence Algorithm, MOEA/D variables,

PM: Preferred Surrogate Model, R: Reference Point, MaxUpdates: Number of weight updates, IR: Number of iterations given to MOEA/D with the new weights to converge to a better solution

Output: P - Final set approximating the Pareto front [1] **Phase 1: Global Exploration**

Initial convergence criteria not met $P \leftarrow$ EvolveAlgorithm(CA, P) **Main Loop (Phases**

2 & 3 & 4) $update = 1$ **to** $MaxUpdates$ **Phase 2: Objective Space Modeling** $A \leftarrow$

GetNonDominated(P) $SM \leftarrow$ TrainSurrogateModel(PM, A) Fits the preferred surrogate model **Phase 3: Dynamic Weights Adjustment** $Z_{max} \leftarrow \arg\max_{Z \in SM}$ Hypervolume(Z, R)

Set of points that maximises the HV $z^* \leftarrow$ GetIdealPoint(A) Minimum observed values $D \leftarrow$ CalculateWeights(z^*, Z_{max}) Dynamic injection into MOEA/D **Phase 4: Evolution with new directions** $i = 1$ **to** IR $P \leftarrow$ EvolveMOEAD(P, D) **return** P

4.2 Phase I: MOEA/D

The global efficacy of SA-MOEA/D-AWA heavily relies on the quality of the first Pareto front approximation. If the initial predicted model is far from the true PF, or if it does not mimic the real PF geometry, the predicted weights for MOEA/D can be misleading. Such inaccuracies can propagate through the iterative loop, causing subsequent surrogate models to misrepresent the Pareto front geometry and producing increasingly unreliable weight directions.

To avoid this and obtain a good prediction of the PF geometry, a Multi-Objective Evolutionary Algorithm, denoted in Algorithm 4.1 as the Convergence Algorithm (CA), is used. While the choice of the CA is flexible, standard algorithms such as MOEA/D, SMS-EMOA, or NSGA-II are highly suitable for this task. During the experiments, MOEA/D is adopted as the CA. The CA evolves an initial randomized population P of a predefined size using its conventional variation operators and selection mechanisms. Figure 7 displays this process: the initial population, Figure 5, is dispersed across the objective space, and by the end of the phase, Figure 6, it has converged close to the Pareto front, yielding the non-dominated set used in the subsequent surrogate modelling.

This first phase runs until an initial convergence criterion is satisfied. Currently, this criterion is a percentage of the total evaluation budget depending on how difficult the PF is to approximate. For unimodal or geometrically regular problems, a shorter Phase 1 converges faster, allowing more of the evaluation budget to be spent on the surrogate-driven refinement. For multimodal or difficult-to-approximate problems, a longer Phase 1 is necessary to ensure that the initial population covers the main features of the Pareto front before the surrogate modelling begins.

4.3 Objective Space Modeling

Once the initial convergence phase is done, the algorithm transitions to the predictive modeling phase. Unlike traditional surrogate-assisted methods that create regression models mapping the decision space to the objective space (i.e. $f\colon R^n \to R^m$), in our approach, we model the geometry directly in the objective space using the current non-dominated population, avoiding the computational burden associated with high-dimensional decision spaces. To create this model, the algorithm first computes the non-dominated solutions from the current population P. The surrogate model is a configurable component of the method. This work evaluates Radial Basis Functions (RBF) and B-Splines as candidates, with the best-performing variant selected through the internal configuration study described in Section 5.2.1.

However, a fundamental challenge in modeling some boundaries is their geometric nature. In extreme regions of the PF, the boundary often becomes nearly vertical, creating a non-Lipschitz continuous area. If we try to train a standard regression model on a nearly vertical boundary, the mapping becomes unstable. Figure 10 illustrates this, as several non-dominated points share almost the same f_1 value along the near-vertical wall (Figure 8), so an RBF fitted directly in (f_1, f_2) oscillates far away from the front (Figure 9). To resolve this, the algorithm applies a 45° rotation to the objective space before fitting the surrogate.

The guarantee that the rotated front is always single-valued follows directly from the non-dominance filter. For a vertical region to exist in the rotated (u, v) space, a point would have to share the same u coordinate as another point while having a strictly greater v value. In the original objective space, this would imply that the point with the lower v value strictly dominates the other — a contradiction, since only mutually non-dominated solutions are retained before the transformation.

Given a bi-objective MOP, a point in the objective space $y = (f_1, f_2)^T$, can be transformed to the new coordinates via the following rotation matrix:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \tag{9}$$

This matrix is the result of $R(\frac{\pi}{4})$ (see Eq. 10), which is the rotation matrix in two dimensions.

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{10}$$

Then, the surrogate model is fitted to learn the mapping $v = \hat{f}(u)$. Once we have the pair (u, v), to transform it to the usual coordinates again, we use the $-45°$ rotation matrix:

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{11}$$

Figure 13 shows the result of this procedure. In the rotated (u, v) space the front becomes single-valued and the same RBF fits it cleanly (Figure 11); mapping the surrogate back to the original coordinates then reconstructs the full front, including the near-vertical wall (Figure 12).

4.4 Dynamic Weights Adjustment for MOEA/D

Given the surrogate model SM fitted in Phase 2, we formulate an internal optimisation problem to find Z_{max}, the set of $|P|$ points on the surrogate curve that maximises the hypervolume with respect to the reference point R:

$$Z_{max} = \arg\max_{z \in SM} \text{Hypervolume}(Z, R) \tag{12}$$

To solve this problem, several optimisers can be used. Gradient-based methods such as BFGS converge rapidly but are susceptible to local optima. Global optimisers such as Dual Annealing explore the search space more thoroughly at the cost of slower convergence. Differential Evolution offers a population-based alternative with intermediate characteristics. This work evaluates all three as candidate optimisers, with the best-performing variant selected through the internal configuration study described in Section 5.2.1. A critical advantage of this approach, regardless of the chosen optimiser, is that the hypervolume is evaluated entirely on the surrogate model, requiring zero evaluations of the true objective functions.

This study is made in the rotated space, so once the optimal set is identified by the optimizer, the inverse rotation matrix is applied to map these points back to the original objective space. It is important to note that Z_{max} only exists in the objective space. The corresponding decision variables are unknown, and no assumption is made about whether a feasible decision vector maps to each predicted point. The role of Z_{max} is to define the target reference directions for MOEA/D, not to produce evaluable candidates.

Once the set Z_{max} is found and mapped back to the original objective space, we need to use it to generate new weights for MOEA/D. To do this, we first identify the ideal point z^* using the minimum observed values in the current approximation.

For each point $z^i \in Z_{max}$, we compute Δ_i, which is the difference between the value of the predicted point and the minimum of that objective. To prevent division by zero in an extreme case where we only have one point, a lower bound $\epsilon = 10^{-4}$ is enforced:

$$\Delta_j^i = \max(z_j^i - z_j^*, \epsilon) \tag{13}$$

The new weight vector D^i from the ith-point is then obtained by normalizing the components of Δ_i so that their sum equals 1:

$$D_j^i = \frac{\Delta_j^i}{\sum_{k=1}^{M} \Delta_k^i} \quad \forall j \in \{1, \ldots, M\} \tag{14}$$

To ensure that we push the PF to the boundaries, we fix the weights [1,0] and [0,1], replacing them with the weights that yield the highest values for f_1 and f_2, respectively. Figure 16 contrasts the resulting directions with the uniform reference vectors of standard MOEA/D. While uniform vectors (Figure 14) concentrate the search in the central region of the front, the HV-optimal directions derived from the surrogate (Figure 15) spread towards the extreme regions, reaching the corners that uniform vectors leave uncovered.

4.5 Evolution with new weights

In the final phase of the iterative loop, after deriving the new set of weight vectors, the new weight vectors are integrated into the evolutionary process. In contrast to existing

approaches that only perform partial weight updates, such as the previously discussed MOEA/D-AWA [18] or MOEA/D-AV [17], our method changes the entire weights of MOEA/D with the new vectors.

At the start of each Phase 4 iteration, the population is initialised with the solutions from the end of the previous phase, preserving the evolutionary progress made so far. The current population P is sorted based on the primary objective (f_1), and we align these points with each corresponding new weight, as we also sort them. It is important to highlight that if the new weights were assigned to the existing population randomly, the mismatch between a solution's spatial location and its assigned search direction would cause a severe drop in the algorithm's convergence performance.

However, even if the weights are correctly paired with the population, the algorithm will likely suffer a temporary drop in terms of hypervolume convergence, as the new population must adapt to the new weights. To allow the population to recover from this temporary drop, Phase 4 runs for a fixed number of iterations before the next weight update is triggered. In this work, this budget is set to a fixed percentage of the total evaluation budget, with a larger allocation given to multimodal or difficult-to-approximate problems, where the population requires more time to adapt to the new directions.

4.6 Computational Complexity of SA-MOEA/D-AWA

We analyse the computational complexity of SA-MOEA/D-AWA by considering each phase separately. The total cost per iteration of the main loop comprises: the cost of the Initial MOEA in Phase 1, the surrogate fitting in Phase 2, the hypervolume maximisation in Phase 3, and the MOEA/D evolution in Phase 4.

Since MOEA/D is used as the Initial MOEA in all experiments, we assume it also serves as the CA for the complexity analysis. The time complexity of MOEA/D has been deeply analysed in multiple papers. In the original work [21], it is proved that its complexity is $\mathcal{O}(NTm)$, where N is the population size, T is the neighborhood size, and m is the dimension of the objective space. Currently, SA-MOEA/D-AWA is a bi-objective method, so the complexity of this phase can be simplified to $\mathcal{O}(NT)$, since constants are ignored in Big-O notation.

The true innovation of this algorithm lies in the 2nd and 3rd phases of the algorithm. The second phase starts by extracting the non-dominated points of the population. The algorithm presented by Kung et al [2] presents a simple algorithm for locating the nondominated set of a collection of points in 2 and 3 dimensions that runs in time $\mathcal{O}(Nlog(N))$. To train the surrogate model, first, the rotation is applied to each point to change the support axis to the rotated axis by 45°, adding a linear cost of $\Theta(N)$. Then, fitting a model requires solving a system of linear equations. In the worst-case scenario, this has a cost of $\Theta(N^3)$, as we need to invert a matrix to solve it. On the other hand, the cost of getting the set of points that maximises the surrogate is $\mathcal{O}(INlog(N))$, where I is the number of generations given to the algorithm to find the best combination of points, currently a constant, and the $\mathcal{O}(Nlog(N))$ is the cost of computing the HV for a set of points in a two dimension space [12]. The cost of getting the ideal point and calculating the weights that represent it is $\mathcal{O}(N)$.

The total complexity per weight update is therefore $\mathcal{O}(N^3 + N \log N)$, and over the full run:

$$\mathcal{O}\big(MaxUpdates \cdot (N^3 + N \log N)\big) \tag{15}$$

The bottleneck is the surrogate fitting. The surrogate fitting cost of $\Theta(N^3)$ is lower than that of typical surrogate-assisted methods that model the mapping from the decision space to the objective space, where the input dimensionality n often dominates the cost. By fitting directly in the two-dimensional objective space, the cost is independent of the decision space dimension n.
The overall complexity is higher than that of standard MOEAs such as MOEA/D, which run in $\mathcal{O}(NTm)$ per generation. However, the surrogate fitting and hypervolume maximisation are performed only $MaxUpdates$ times throughout the entire run, keeping the practical overhead small relative to the total evaluation budget.

Chapter 5 Experiments

5.1 Experimental Setup

5.1.1 Benchmark Problems

To evaluate the performance of SA-MOEA/D-AWA, the selected problems include a diverse suite of continuous multi-objective benchmark problems. The baseline tests include the 5 non-discrete problems from the ZDT family [23] (ZDT1 to ZDT4, and ZDT6), and the problems from the DTLZ family [11], except the problem DTLZ4. While these popular benchmark problems are excellent for evaluating general convergence issues, they may lack the extreme geometric irregularities found in real-world problems. Therefore, to test certain characteristics, seven distinct problems following the structure defined in [20] are included in the study. Particularly, these problems test PF with extreme concavity, extreme convexity, severe multimodality, disconnected regions, Difficult-to-Approximate (DtA) boundaries, and concavity and convexity simultaneously. Figure 17 shows the resulting front geometries. Each problem was executed for 10 independent runs, with a seed in each run to ensure reproducibility in the results.

5.1.2 Performance metrics

The metrics used to evaluate the performance of the different algorithms are the Hypervolume (HV) and the Inverted Generational Distance (IGD). In our experiments, as we know the objective functions' optimal values, specific reference points are defined for each problem to ensure a consistent HV scaling. A higher HV usually indicates a better approximation to the PF, showing better convergence, more diversity, or both. For the baseline problems, the true Pareto fronts are known. The set of points that maximised the HV in each problem is provided by Eckart Zitzler and his research group at ETH Zurich[1]. Therefore, the convergence plot shows the difference from that value, providing a clear picture of both convergence speed and final approximation quality. Finally, to compare against the benchmark methods, the IGD metric will be used to provide a more comprehensive performance of the different algorithms.

[1] `https://www.sop.tik.ee.ethz.ch/download/supplementary/testproblems/index.html`

5.1.3 Algorithms for Comparison

To test the proposed SA-MOEA/D-AWA, its performance is compared against the standard MOEA/D. The reason why we use this algorithm is to compare whether the dynamic adaptive weights can improve its general performance, achieving better convergence and diversity. Additionally, we compare the algorithm against MOEA/D-AWA [18], an adaptive weight adjustment method that represents the standard alternative within the decomposition-based family.

5.1.4 Parameter Settings

To ensure a fair comparison, we use the same parameter settings for all the compared algorithms. The population size is set to $N = 50$ for all instance problems. The methods used for the crossover and mutation operations are Simulated Binary Crossover (SBX) and Polynomial Mutation (PM), respectively, with their default hyperparameters. All the compared algorithms stop when they reach the limit of function evaluations of 50,000.
The proposed SA-MOEA/D-AWA requires three additional hyperparameters, which were tuned empirically on a representative subset of the benchmark suite and held fixed across all problems. Initially, MOEA/D runs for 400 iterations to generate a good initial approximation of the Pareto front. Once the new directions are found, the new MOEA/D evolves for 75 iterations. In order to meet the 50,000 function evaluations, we compute these new directions 8 times. In Section 7, we discuss how this work can be extended by substituting these hyperparameters with some deterministic policy. Different Surrogate Model settings or different configurations to maximize the HV in that model will be analysed in Section 5.2.1.

5.2 Results

5.2.1 Internal Configuration Analysis

The SA-MOEA/D-AWA will be tested in 12 different variants in the 18 problems described in Section 5.1.1, testing whether the supportive axis is the axis X or if it is the axis X rotated 45^o (denoted as axis M), predicting the surrogate model using RBF or B-SPLINES, or using three different methods to compute the new weights of the posterior MOEA/D evolution. Table 1 and Table 2 display the mean and the standard deviation of these variants in the benchmark problems. For the problems of the families DTLZ and ZDT, the table shows the difference between the hypervolume of the known optimal distribution of points and the predicted Pareto front. While for the 6 generated problems, the total HV is shown with reference point (1.5,1.5), except for DENT and extremely convex, that is (5, 5) and (1.1, 1.1) respectively.
It is worth noting that DTLZ5 has identical results to DTLZ2 in all the variants. This is a known theoretical property, as in the bi-objective case, the angular transformation of DTLZ5 reduces to the identity mapping, making the two problems mathematically equivalent [11]. The results are retained in the table to confirm this equivalence empirically.
To select the best-performing configuration for comparison against the benchmark methods, a Critical Differences (CD) analysis is conducted. The 12 variants are ranked for each problem by their mean performance across the 10 independent runs. The Friedman test is then applied to determine whether all variants are statistically equivalent at a signifi-

Table 1: Mean and standard deviation of the performance metric for all 18 benchmark problems across 10 independent runs, variants 1-6. For ZDT and DTLZ problems, HV difference (lower is better). For generated problems, raw HV (higher is better).

Problem	**RBF(X)+DE**	**RBF(M)+DE**	**BSPL(X)+DE**	**BSPL(M)+DE**	**RBF(X)+BFGS**	**RBF(M)+BFGS**
ZDT1	0.00021 ± 0.00006	0.00010 ± 0.00005	0.00076 ± 0.00010	0.00029 ± 0.00003	0.00020 ± 0.00008	**0.00009 ± 0.00006**
ZDT2	0.00035 ± 0.00030	0.00030 ± 0.00016	0.00017 ± 0.00027	0.00027 ± 0.00022	0.00038 ± 0.00015	0.00034 ± 0.00021
ZDT3	0.00196 ± 0.00101	0.00197 ± 0.00023	0.00445 ± 0.00050	0.00288 ± 0.00003	0.00091 ± 0.00012	0.00202 ± 0.00028
ZDT4	0.00849 ± 0.00520	0.01116 ± 0.00704	0.00921 ± 0.00320	0.01069 ± 0.00447	0.00520 ± 0.00233	0.00930 ± 0.00565
ZDT6	0.00212 ± 0.00112	0.00270 ± 0.00191	0.00337 ± 0.00324	0.00285 ± 0.00227	0.00129 ± 0.00032	0.00227 ± 0.00225
DTLZ1	0.00259 ± 0.00170	0.00221 ± 0.00160	0.00194 ± 0.00138	0.00192 ± 0.00116	0.04069 ± 0.11463	0.00171 ± 0.00087
DTLZ2	**0.00011 ± 0.00003**	0.00026 ± 0.00001	0.00066 ± 0.00006	0.00014 ± 0.00000	0.00015 ± 0.00004	0.00025 ± 0.00001
DTLZ3	0.04927 ± 0.12277	0.05272 ± 0.12170	0.01122 ± 0.00899	0.00855 ± 0.00232	0.00976 ± 0.00729	0.01056 ± 0.00474
DTLZ5	**0.00011 ± 0.00003**	0.00026 ± 0.00001	0.00066 ± 0.00006	0.00014 ± 0.00000	0.00015 ± 0.00004	0.00025 ± 0.00001
DTLZ6	0.13098 ± 0.05834	0.11983 ± 0.05746	0.11845 ± 0.05369	0.11877 ± 0.05399	0.11448 ± 0.04581	0.11194 ± 0.05678
DTLZ7	0.13954 ± 0.21314	0.13955 ± 0.21317	0.13951 ± 0.21311	0.14256 ± 0.21113	0.13955 ± 0.21317	0.13953 ± 0.21314
DENT	17.31237 ± 0.01247	**17.33357 ± 0.00081**	17.29568 ± 0.01649	17.27754 ± 0.00973	17.32391 ± 0.00121	17.33336 ± 0.00081
Ext. Concave	1.39468 ± 0.00207	1.39521 ± 0.00197	1.38745 ± 0.00424	1.39304 ± 0.00268	1.39483 ± 0.00184	**1.39589 ± 0.00133**
Ext. Convex	1.19391 ± 0.00034	**1.19471 ± 0.00001**	1.19307 ± 0.00031	1.19353 ± 0.00023	1.19397 ± 0.00015	**1.19471 ± 0.00000**
Multimodal	1.05489 ± 0.04719	1.04350 ± 0.04743	1.05512 ± 0.04784	1.05493 ± 0.04757	1.05538 ± 0.04812	1.05557 ± 0.04713
DtA	1.50714 ± 0.01057	1.52089 ± 0.01575	1.51468 ± 0.02069	1.52326 ± 0.01900	1.51197 ± 0.01271	1.52114 ± 0.01743
Convergence	1.61963 ± 0.01966	1.63232 ± 0.02107	1.63463 ± 0.01440	1.63754 ± 0.02298	1.63921 ± 0.01740	1.63922 ± 0.01433
Disconnected	0.90368 ± 0.05399	0.89506 ± 0.05563	0.85351 ± 0.05113	0.87714 ± 0.04947	0.91632 ± 0.05286	0.91827 ± 0.05337

Table 2: Mean and standard deviation of the performance metric for all 18 benchmark problems across 10 independent runs, variants 7-12 (continued from Table 1).

Problem	**BSPL(X)+BFGS**	**BSPL(M)+BFGS**	**RBF(X)+DUAL**	**RBF(M)+DUAL**	**BSPL(X)+DUAL**	**BSPL(M)+DUAL**
ZDT1	0.00101 ± 0.00037	0.00028 ± 0.00004	0.00024 ± 0.00009	**0.00009 ± 0.00006**	0.00101 ± 0.00040	0.00030 ± 0.00002
ZDT2	0.00077 ± 0.00076	0.00023 ± 0.00024	0.00037 ± 0.00024	0.00035 ± 0.00017	**0.00015 ± 0.00027**	0.00026 ± 0.00023
ZDT3	0.00441 ± 0.00040	0.00300 ± 0.00014	**0.00078 ± 0.00010**	0.00196 ± 0.00021	0.00457 ± 0.00022	0.00292 ± 0.00009
ZDT4	0.00912 ± 0.01298	0.01539 ± 0.00862	0.00944 ± 0.00383	**0.00305 ± 0.00131**	0.00879 ± 0.00394	0.00397 ± 0.00125
ZDT6	0.00210 ± 0.00326	0.00305 ± 0.00345	0.00226 ± 0.00119	**0.00127 ± 0.00028**	0.00299 ± 0.00407	0.00152 ± 0.00059
DTLZ1	0.07728 ± 0.15130	0.00172 ± 0.00111	0.04054 ± 0.11565	**0.00146 ± 0.00077**	0.00215 ± 0.00136	0.00195 ± 0.00139
DTLZ2	0.00065 ± 0.00006	0.00014 ± 0.00000	0.00015 ± 0.00002	0.00026 ± 0.00002	0.00066 ± 0.00005	0.00015 ± 0.00003
DTLZ3	0.01041 ± 0.00638	0.01431 ± 0.01101	0.01570 ± 0.01448	**0.00786 ± 0.00473**	0.01496 ± 0.00527	0.01178 ± 0.00791
DTLZ5	0.00065 ± 0.00006	0.00014 ± 0.00000	0.00015 ± 0.00002	0.00026 ± 0.00002	0.00066 ± 0.00005	0.00015 ± 0.00003
DTLZ6	0.12502 ± 0.06311	0.11750 ± 0.06036	0.12492 ± 0.04973	0.11946 ± 0.06343	**0.11081 ± 0.05418**	0.11280 ± 0.04845
DTLZ7	0.13951 ± 0.21311	0.13951 ± 0.21311	0.13952 ± 0.21312	0.13957 ± 0.21320	**0.13950 ± 0.21309**	0.14256 ± 0.21113
DENT	17.27082 ± 0.00589	17.26952 ± 0.00646	17.33185 ± 0.00080	17.33330 ± 0.00082	17.29995 ± 0.01022	17.29979 ± 0.00096
Ext. Concave	1.38738 ± 0.00448	1.39559 ± 0.00237	1.39484 ± 0.00146	1.39581 ± 0.00050	1.38800 ± 0.00344	1.39389 ± 0.00225
Ext. Convex	1.19252 ± 0.00094	1.19356 ± 0.00031	1.19425 ± 0.00010	1.19470 ± 0.00001	1.19300 ± 0.00055	1.19359 ± 0.00023
Multimodal	1.04410 ± 0.04711	1.05559 ± 0.04752	1.05508 ± 0.04767	**1.05562 ± 0.04745**	1.05552 ± 0.04798	1.05471 ± 0.04725
DtA	1.51153 ± 0.01544	1.51674 ± 0.01689	1.51197 ± 0.01318	1.52186 ± 0.00648	1.52033 ± 0.02076	**1.52390 ± 0.01501**
Convergence	1.63003 ± 0.02235	1.62486 ± 0.01641	1.63006 ± 0.01925	1.63195 ± 0.02026	1.63531 ± 0.01570	**1.64910 ± 0.01372**
Disconnected	0.86908 ± 0.05548	0.90358 ± 0.05134	0.92196 ± 0.04786	0.91395 ± 0.05050	0.86808 ± 0.05711	**0.92205 ± 0.05103**

cance level of $\alpha = 0.05$. If the null hypothesis is rejected, a post-hoc Wilcoxon signed-rank test with Holm's correction identifies which variants differ significantly. The results are summarised using a CD diagram [14], where variants are positioned by their average rank and methods connected by a horizontal line are not significantly different.
Figure 18 displays this study over the 18 problems. In the best horizontal line, there are 10 variants without a statistical difference. Therefore, the tie is broken by mean ranking, where RBF(M)+DUAL and RBF(M)+BFGS are nearly indistinguishable, with the latter ranked marginally higher. The results show that rotating the support axis consistently improves performance: five of the six best-ranked variants use the rotated axis (Axis M). Among the three optimisers, Differential Evolution produces the worst overall ranking, while RBF outperforms B-Splines as the surrogate model across most problems.
To break the tie, the analysis is repeated considering only the ZDT and DTLZ families, as MOEA/D-AWA results are only available for these problems.
Figure 19 shows that there is no statistical significant difference between the proposed methods with this set of problems, meaning that the Friedman Test could not reject the null hypothesis. The main reason for this result is that the number of problems that we have for 12 different algorithms is too small. However, the model that uses the rotated RBF with Dual Annealing for computing the new weights of MOEA/D has the best ranking out of the 12 variants.
To complete the internal study, Figure 20 shows the CD analysis on the set of problems created from the work [20]. As expected given the reduced number of problems, the Friedman test cannot reject the null hypothesis in this subset either. The limited sample size constrains the statistical power of the test. However, the model with the best ranking in this set of problems is the rotated RBF with BFGS, followed by the variant that uses Dual Annealing.
Based on the combined evidence from the three CD analyses, the variant using a rotated RBF surrogate with Dual Annealing (RBF(M)+DUAL) is selected for comparison against the benchmark methods. This variant achieves the best average rank on the ZDT and DTLZ problems, which are the only problems for which MOEA/D-AWA results are available, and ranks second overall on the full benchmark suite. The slight advantage of RBF(M)+BFGS on the full suite does not outweigh the consistency advantage of RBF(M)+DUAL across the two subsets.

5.2.2 Performance on Standard Benchmarks

In this section, the selected variant, RBF(M)+DUAL, is compared against MOEA/D and MOEA/D-AWA on the ZDT and DTLZ families. The three-way comparison covers the five ZDT problems and DTLZ1, 2, and 3, as MOEA/D-AWA results are only available for these problems. The comparison against MOEA/D alone is extended to the full benchmark suite and reported separately. Performance is measured using the HV error and IGD metrics described in Section 2.2. To determine whether the differences are statistically significant, the nonparametric Mann-Whitney U test is applied at a significance level of $\alpha = 0.05$. Also, as multiple hypotheses are tested simultaneously across various benchmark problems, the Holm-Sidak correction is applied to prevent false positives.
The statistical results for the HV error and IGD metrics are presented in Table 3 and Table 4, respectively.

Table 3: HV error values (median and 10% to 90% quantile range) of the final Pareto front approximations over 10 independent runs with population size 50. For HV error, lower is better, as it is displayed the difference to the problem maximum HV. Mann-Whitney U (two-sided, $\alpha = 0.05$) with Holm-Sidak correction. Arrows are relative to MOEA-D + RBF(AXIS M) + DUAL: ↑ better, ↓ worse, ↔ no significant difference.

Problem	MOEA-D + RBF(AXIS M) + DUAL	MOEA/D-AWA	MOEA-D
ZDT1	**6.1989e-05(1.52e-04)**	1.8300e-03(3.01e-04)↑	1.6141e-03(3.10e-04)↑
ZDT2	**3.3120e-04(3.26e-04)**	2.0410e-03(2.91e-04)↑	2.3140e-03(1.72e-03)↑
ZDT3	**1.8704e-03(3.91e-04)**	2.6550e-03(7.68e-04)↑	1.0397e-02(1.53e-03)↑
ZDT4	2.6403e-03(2.95e-03)	**1.5685e-03(1.70e-03)**↓	6.2926e-03(5.67e-03)↑
ZDT6	**1.3197e-03(6.80e-04)**	2.0970e-03(3.52e-04)↑	3.2585e-03(7.99e-04)↑
DTLZ1	1.2934e-03(1.86e-03)	**5.6250e-04(1.02e-04)**↓	2.3554e-03(2.73e-03)↑
DTLZ2	**2.5287e-04(2.69e-05)**	1.7725e-03(1.66e-04)↑	1.8408e-03(1.61e-06)↑
DTLZ3	7.4373e-03(1.21e-02)	**1.8705e-03(4.61e-04)**↓	1.5943e-02(1.68e-02)↑
+/≈/-		5/0/3	8/0/0

Table 4: IGD values (median and 10% to 90% quantile range) of the final Pareto front approximations over 10 independent runs with population size 50. For IGD, lower is better. Mann-Whitney U (two-sided, $\alpha = 0.05$) with Holm-Sidak correction. Arrows are relative to MOEA-D + RBF(AXIS M) + DUAL.

Problem	MOEA-D + RBF(AXIS M) + DUAL	MOEA/D-AWA	MOEA-D
ZDT1	**7.2318e-03(4.17e-04)**	8.3225e-03(3.17e-04)↑	8.1751e-03(9.98e-05)↑
ZDT2	8.1456e-03(8.68e-04)	8.7840e-03(1.06e-04)↔	**7.9893e-03(6.07e-04)**↔
ZDT3	1.5022e-02(2.58e-03)	**1.0863e-02(4.87e-04)**↓	1.6355e-02(3.44e-04)↑
ZDT4	**8.3044e-03(7.45e-04)**	8.3680e-03(3.85e-04)↔	9.3604e-03(2.01e-03)↑
ZDT6	**6.1787e-03(9.02e-04)**	7.1045e-03(2.07e-04)↑	6.7778e-03(2.94e-04)↔
DTLZ1	**4.2134e-03(2.19e-03)**	4.3090e-03(7.75e-05)↔	5.0417e-03(2.76e-03)↔
DTLZ2	1.0109e-02(5.97e-04)	9.2150e-03(1.69e-04)↓	**7.9044e-03(6.92e-10)**↓
DTLZ3	1.3154e-02(5.11e-03)	**9.3035e-03(4.46e-04)**↓	1.2381e-02(7.06e-03)↔
+/≈/-		2/3/3	3/4/1

5.2.3 Geometrically Challenging Benchmarks

The results for the remaining problems are shown in Table 5. For this set of problems, only the HV metric is reported, as the PF is unknown. For the DTLZ tests, the difference between the optimal set and the predicted set is reported, i.e., the closer to zero the better, while for the rest of the problems, the higher the final HV is reported, so the higher the better.

Table 5: Performance metrics (median and 10% to 90% quantile range) of the final Pareto front approximations over 10 independent runs with population size 50. For DTLZ5, DTLZ6, and DTLZ7, the HV difference is reported (lower is better). For the seven generated problems, the raw HV is reported (higher is better). Mann-Whitney U (two-sided, $\alpha = 0.05$) with Holm-Sidak correction. Arrows are relative to MOEA-D + RBF(AXIS M) + DUAL: ↑ better, ↓ worse, ↔ no significant difference.

Problem	MOEA-D + RBF(AXIS M) + DUAL	MOEA-D
DTLZ5	**2.5287e-04(2.69e-05)**	1.8408e-03(1.61e-06)↑
DTLZ6	**1.2410e-01(1.48e-01)**	1.3017e-01(1.41e-01)↔
DTLZ7	**1.0000e-10(4.65e-01)**	4.7628e-04(4.66e-01)↔
DENT	**1.7333e+01(1.90e-03)**	1.7040e+01(7.77e-03)↑
Ext. Concave	**1.3959e+00(1.19e-03)**	1.3867e+00(2.10e-02)↑
Ext. Convex	**1.1947e+00(1.39e-05)**	1.1737e+00(1.40e-03)↑
DtA	**1.5217e+00(1.60e-02)**	1.4637e+00(1.22e-02)↑
Conv. Problem	**1.6316e+00(5.12e-02)**	1.5621e+00(3.56e-02)↑
Disconnected	**9.4050e-01(1.09e-01)**	8.8332e-01(1.21e-01)↑
Multimodal	**1.0355e+00(1.04e-01)**	1.0318e+00(1.01e-01)↔
+/≈/-		7/3/0

5.2.4 Convergence Analysis

To complement the aggregate statistics reported in Tables 3, 4 and 5, we now examine the convergence behaviour of SA-MOEA/D-AWA against MOEA/D along the full evolutionary run. Figure 21 displays the hypervolume convergence on six representative problems, selected to cover the full benchmark suite. For every problem, the y-axis shows the difference between the best hypervolume on the problem and the current population's hypervolume on a logarithmic scale, so that lower values indicate better approximations. Solid lines represent the median across the 10 independent runs, and the shaded regions cover the 10% to the 90% quantile band. We plot the quantiles and not the standard deviation, using a logarithmic scale on the y-axis. MOEA/D-AWA is not included in this analysis, as the available implementation only stores the final approximation set and does not provide the statistics in each generation. Therefore, the comparison against MOEA/D-AWA is confined to the final-front metrics already reported in Tables 3 and 4.

ZDT1 represents the standard behaviour observed across the ZDT family. During the first 400 generations, both algorithms behave identically, as the initial convergence phase of SA-MOEA/D-AWA is itself MOEA/D. After the first update of the weights, the new directions take effect, and the median gap of SA-MOEA/D-AWA drops by roughly one

order of magnitude below MOEA/D, reaching 10^{-4} by generation 1000 while MOEA/D gets stuck near $2 \cdot 10^{-3}$.
On DENT, SA-MOEA/D-AWA improves by roughly two orders of magnitude after the first weight update, while MOEA/D stabilizes from generation 50 onwards.
ZDT3 extends the previous observation to disconnected Pareto fronts. SA-MOEA/D-AWA reduces the gap well past the point where MOEA/D gets stuck. The quantile band is wider than in ZDT1, reflecting occasional runs in which the surrogate misrepresents the disconnections.
On DTLZ1, the curve bands of both methods are wide due to the multimodal behaviour of the problem, but the median of SA-MOEA/D-AWA remains consistently below MOEA/D from generation 450.
DTLZ7 marks the boundary of the method's effectiveness. The sharp drop of the median around generation 450 shows that several runs successfully reach all the disconnected components of the front, but the wide quantile band reveals that other runs remain trapped in other components. The Mann-Whitney U test on the final hypervolume yields a p-value of 0.07, just above the corrected significance threshold, consistent with the result reported in Table 5. The difference does not reach significance under Holm-Sidak correction at $\alpha = 0.05$.
On Multimodal, both methods converge to comparable final values throughout the run, and the difference is not statistically significant. The reasons for this behaviour are discussed in Section 6.

5.2.5 Visual Analysis of Pareto fronts

While previous sections quantify performance, they do not reveal how the final approximations distribute across the Pareto front. Figure 22 presents the final populations of SA-MOEA/D-AWA (left column) and MOEA/D (right column) on four problems with non-trivial geometries. For each method, we display the population of the run whose final HV is closest to the median across the 10 independent runs. The true Pareto front is displayed as a thin grey line, except in DTLZ7, where it is not known.
On DENT, SA-MOEA/D-AWA distributes its solutions across the full extent of the front, with points reaching $f_1 \approx 4.5$ on the x-axis and $f_2 \approx 4.5$ on the y-axis. On the other hand, MOEA/D concentrates its solutions in the central region, and the right tail does not extend further than $f_1 \approx 2.5$, leaving the extreme regions of the front uncovered.
SA-MOEA/D-AWA extends the Convergence Problem further along the front, reaching points of $f_1 \approx 0$ close to the real PF, while MOEA/D has more problems to find the left extreme. Both algorithms produce approximations that lie slightly above the true Pareto front, indicating that neither algorithm achieves full convergence under the present budget, but the proposed method reaches, overall, a closer approximation.
On Extremely Convex, the contrast is most significant. SA-MOEA/D-AWA predicts the full convex geometry, with solutions distributed along the entire Pareto front. On the other hand, MOEA/D remains trapped in the central portion of the front, missing the entire horizontal and vertical sections.
On the disconnected DTLZ7, both methods discover the same two of the four disconnected components of the front. Neither algorithm reaches the two components with $f_1 \leq 2.5$. The distributions within the reached components differ: SA-MOEA/D-AWA spreads its solutions more uniformly across both components, while MOEA/D concentrates a larger

fraction of its solutions in the fourth component.

Chapter 6 Discussion

The experimental results presented in Section 5 allow us to understand the problems in which SA-MOEA/D-AWA brings a measurable benefit and those in which it does not. In this section, we analyse these scenarios, followed by the trade-off between hypervolume and IGD performance, the practical implications of the results, and the limitations of the present study.

6.1 Performance characterisation by problem geometry

The strongest gains of SA-MOEA/D-AWA over the baselines are observed on problems whose Pareto front exhibits sharp or otherwise irregular geometric features, where the uniform initial definition of the weights is not enough to capture the shape of the Pareto front. Against MOEA/D, the proposed method wins in all ZDT problems, and on all three DTLZ problems where both methods are compared (Table 3), and on 7 out of 10 of the extended benchmarks with non-standard PF geometries (Table 5). Against MOEA/D-AWA, the proposed method wins on 5 out of 8 problems in terms of hypervolume (Table 3). The mechanism underlying these gains is consistent with the design motivation of the rotated-axis surrogate described in Section 4.3. The rotation by 45° transforms vertical regions of the Pareto front into well-behaved, single-valued functions, allowing a good approximation of the RBF surrogate even on geometries that would destabilise a direct interpolation in the original objective space. The analytic hypervolume maximisation on the surrogate then yields reference directions that target the extreme regions of the front, including those that uniform weight vectors of standard MOEA/D cannot reach.
The visual analysis in Figure 22 provides a direct evidence of this mechanism. On DENT, SA-MOEA/D-AWA distributes its solutions across the full extent of the front, while MOEA/D cannot explore outside the central region. On Extremely Convex, the contrast is even more pronounced, with SA-MOEA/D-AWA reconstructing the entire convex boundary, whereas MOEA/D remains trapped in the central region of the front, missing the entire horizontal and vertical section. The convergence plots displayed in Figure 21 confirm these observations analytically, with SA-MOEA/D-AWA achieving improvements of one to two orders of magnitude in HV gap on the geometrically challenging benchmarks. On the standard ZDT and DTLZ benchmarks, the benefit is also visible, although it is less pronounced. The Pareto fronts of these problems do not exhibit the geometric pathologies that the method is designed to address, but the surrogate-derived directions still distribute the search effort more efficiently than uniform reference vectors, producing consistent, significant statistical improvements in HV throughout the run.

6.2 Neutral scenarios and their causes

Two distinct failure reasons emerge from the experimental results, and in both cases, the cause of the failure can be traced to the design of the method.
The first failure mode comes from problems whose difficulty lies in the decision space rather than in the geometry of the Pareto front. The Multimodal problem is the clearest

example, with SA-MOEA/D-AWA and MOEA/D converging to comparable final hypervolume values, and the convergence curves do not show a significant difference, because of the high standard deviation values during the whole run, as reported in the last subplot of Figure 21. The Pareto front of this problem is geometrically well-behaved, so the surrogate model fits it accurately, and the analytic HV maximisation produces sensible reference directions. The challenge, however, is escaping local minima in the decision space, and adapting reference directions in objective space cannot solve that. The proposed method is therefore neutral with respect to multimodality of the decision space, neither helping nor harming, and any improvement on such problems would have to come from complementary modifications to the variation operators or the search dynamics.

The second failure reason comes from severely disconnected problems with many Pareto front disconnected components. The rotated-axis surrogate fits a continuous curve to the non-dominated solutions, so, currently, it cannot represent the gaps between disconnected components. The algorithm, therefore, only benefits if the population after Phase 1 already covers the disconnected regions. DTLZ7 illustrates this failure problem at the edge of the method's capability. The fifth subplot of the convergence plot, Figure 21, shows a clear drop in the median after the first weight update is done. This indicates that in a subset of the runs, the new reference directions successfully push the search beyond the initial component of the front and reach additional disconnected components. The wide quantile band, however, shows that this success is not reproduced across all 10 runs, as in some runs, the population remains confined to the components already covered after Phase 1, and the surrogate-based reweighting cannot discover them. The method, therefore, helps when the disconnected geometry can be partially bridged by the new directions, and fails otherwise.

The visual analysis of DTLZ7 on the last row of Figure 22 further qualifies this result. None of the methods finds all the components of the front, with both algorithms covering only the first and the second components. However, in these two connected components, SA-MOEA/D-AWA divides its solutions more uniformly, whereas MOEA/D concentrates a larger fraction of its population on a single component. The Mann-Whitney U test on the final HV yields $p = 0.07$, which does not reach significance under Holm–Sidak correction at $\alpha = 0.05$. The difference in intra-component diversity is therefore observable but not statistically confirmed under the present sample size.

6.3 The Hypervolume/IGD trade-off

A consistent pattern over the standard benchmarks is that SA-MOEA/D-AWA tends to win on hypervolume and tie or lose on IGD relative to MOEA/D-AWA. Against this baseline, the win/tie/loss counts are 5/0/3 for HV but 2/3/3 for IGD, as shown in Table 3 and Table 4 respectively.

This is not a contradiction but a consequence of the optimisation criterion used to derive the new reference directions. The set of points $Z_{\max}$ on the surrogate is obtained by maximising hypervolume analytically, solving Equation 12. Therefore, the resulting weight vectors are HV-optimal by construction. MOEA/D-AWA, in contrast, adjusts the weights to enforce uniform spread along the front via a crowding-based heuristic, which is the property that the IGD indicator rewards. Also, it is important to highlight that while MOEA/D-AWA performs the weight adjustment heuristically and indicator-based methods such as SMS-EMOA improve HV by selection over evaluated solutions, the proposed

method computes the HV-optimal directions analytically on the surrogate model. This requires no additional evaluations of the true objective function, which is the dominant cost in expensive optimisation scenarios, and obtains directions that are optimal with respect to the surrogate instead of being approximated through stochastic search.
This discrepancy is most pronounced on DTLZ2 and DTLZ3. Both problems have convex Pareto fronts, and on convex geometries the HV-optimal distribution of a finite set is non-uniform, as solutions concentrate near the extreme regions of the front, where each point dominates a larger exclusive volume. IGD, by contrast, rewards uniform spread across the front. Therefore, on convex problems, any method that explicitly maximises hypervolume will produce distributions that IGD penalises, regardless of convergence quality. This effect is particularly visible on DTLZ2, where SA-MOEA/D-AWA is the only case in which the proposed method loses against both baselines simultaneously, MOEA/D-AWA and standard MOEA/D, in IGD, and on DTLZ3, where it loses against MOEA/D-AWA and ties against MOEA/D. These results are therefore not indicative of a convergence failure but of the theoretical incompatibility between the HV-optimal distribution and the uniform spread that IGD rewards on convex fronts.
This makes SA-MOEA/D-AWA an HV-driven method by construction. The trade-off comes directly from how the new directions are computed and from what role hypervolume plays in the formulation. Therefore, the choice between SA-MOEA/D-AWA and MOEA/D-AWA depends on what the user intends to optimise.

6.4 Practical implications

The patterns identified in the previous subsections suggest a simple positioning of SA-MOEA/D-AWA relative to the tested baselines.
SA-MOEA/D-AWA is most appropriate when the geometry of the Pareto front is the main source of difficulty: extreme curvature, sharp boundaries, or regions of high non-uniform density. In such a scenario, the method consistently outperforms both MOEA/D and MOEA/D-AWA in hypervolume, and the surrogate-driven reweighting reaches regions of the front that uniform weight vectors do not cover. Applications where the value of an approximation is proportional to the dominated volume of the objective space, such as engineering design, where corner solutions are valuable trade-offs, can be considered in this category.
MOEA/D-AWA remains a competitive alternative when the priority is a uniform distribution of solutions along the Pareto front. This is the case in applications where the approximation set is presented to a human decision-maker who needs a representative pool of trade-offs, rather than the maximum dominated volume. The 2/3/3 IGD record of SA-MOEA/D-AWA against MOEA/D-AWA reflects this, as when uniform coverage is the goal, the explicit spread-enforcement of MOEA/D-AWA gives it an edge.
Standard MOEA/D is dominated by both methods on the problems studied. Its main practical role is as a baseline against which more sophisticated weight-adjustment schemes can be evaluated, rather than as a recommended choice for direct application.
Problems whose difficulty lies in the decision space, such as Multimodal, are not addressed by any of the three compared methods. Their resolution requires modifications to the variation operators or to the search dynamics that are orthogonal to the weight-adjustment mechanism studied here.
A final practical advantage of the proposed framework concerns its modularity. SA-

MOEA/D-AWA does not replace MOEA/D but augments it, as the weight-adjustment loop wraps around an underlying decomposition-based MOEA without modifying its variation operators, selection mechanism, or convergence guarantees between updates. The same surrogate-driven adjustment could, in principle, be combined with other decomposition-based algorithms, for example, MOEA/D-DE or MOEA/D-DRA, by replacing the inner algorithm in Phase 1 and Phase 4. This makes the contribution of the present work compatible with, rather than competitive against, current improvements to the decomposition family of MOEAs.

6.5 Limitations

There are two main limitations of the present study that should be discussed: the dimensionality and the hyperparameter selection.

6.5.1 Dimensionality

Currently, SA-MOEA/D-AWA is tested for 2 dimensions. For $m \geq 3$ objectives, the rotation is no longer uniquely defined, and the surrogate would have to model a hyperplane in R^m. The analytic hypervolume maximisation also becomes substantially more expensive as the dimension of the objective space grows. Therefore, the method as presented here is restricted to the bi-objective case.

6.5.2 Hyperparameter tuning

The initial convergence generations, 400 iterations for Phase 1, and the per-update budget, 75 generations for Phase 4, repeated 8 times, were tuned empirically on a subset of the benchmark suite and held fixed across all problems. The method does not currently include a mechanism to adapt these values to problem-specific characteristics such as the difficulty of the initial convergence or the rate of improvement after each weight update.

Chapter 7 Conclusion

This thesis covered the problem of distributing solutions across geometrically irregular Pareto fronts in decomposition-based multi-objective evolutionary algorithms. The proposed method, SA-MOEA/D-AWA, fits a surrogate model directly in the objective space to reconstruct the geometry of the Pareto front, and uses analytic hypervolume maximisation on this surrogate to derive new reference directions for MOEA/D, without evaluating the true objective functions. The main finding is that this HV-driven reweighting consistently reaches the extreme regions of the Pareto front that uniform reference vectors do not cover, producing statistically significant improvements in hypervolume on problems where the geometry of the Pareto front is the main source of difficulty.
The work makes four concrete contributions. First, a 45° rotation applied before surrogate fitting guarantees that the bi-objective Pareto front is always single-valued in the transformed space, resolving the instability caused by near-vertical regions. Second, an optimisation procedure derives hypervolume-optimal reference directions directly from the surrogate, requiring zero additional evaluations of the true objective functions. Third, the

weight-adjustment loop is flexible, as it wraps around standard MOEA/D without modifying its variation operators or convergence guarantees. Making it, in principle, compatible with other decomposition-based variants. Fourth, an empirical study on 18 benchmark problems confirms that the method outperforms MOEA/D on all five ZDT problems and all three DTLZ problems in the three-way comparison, and wins on 7 out of 10 problems with non-standard Pareto-front geometries. Against MOEA/D-AWA, the method wins on 5 out of 8 problems in hypervolume, while the IGD losses on convex fronts such as DTLZ2 and DTLZ3 are a direct consequence of the HV-optimality of the derived directions, not of convergence failure.

The experimental results also reveal the limits of the method's effectiveness. Problems whose difficulty lies in the decision space, such as multimodal landscapes, are not solved by weight adjustment in the objective space. Severely disconnected Pareto fronts with many components present a second limitation, as the continuous surrogate cannot represent the gaps between components, and the method only benefits when the initial MOEA/D phase already covers the relevant regions. Both failure modes are consistent with the design of the method and suggest concrete directions for future work.

Several extensions follow naturally from the present work. The most direct is the generalisation to three or more objectives. In the bi-objective case, the 45° rotation aligns one coordinate axis with the diagonal direction $(1,1)/\sqrt{2}$, which is approximately normal to the Pareto front of standard benchmark problems. This rotation admits a more general reading, as the diagonal $(1,1)/\sqrt{2}$ is precisely the normal of the standard simplex Δ^1 (the segment joining $(1,0)$ and $(0,1)$), while the orthogonal axis runs along that simplex. Fitting $v = \hat{f}(u)$ is therefore nothing but parametrising the front by its position along Δ^1 and modelling its displacement away from the simplex. Seen this way, the 45° rotation is the two-dimensional instance of a general construction: project the non-dominated set onto the hyperplane that supports the scaled simplex Δ^{m-1}, and model the front as a graph over it.

The generalisation then follows naturally. For m objectives, the vector $\overrightarrow{N} = (1,\ldots,1)/\sqrt{m}$ is the normal of Δ^{m-1} and is approximately normal to the Pareto front, and a rotation matrix $R \in R^{m\times m}$ can be constructed whose last row is $\overrightarrow{N}$ and whose remaining rows form an orthonormal basis of the hyperplane perpendicular to it. Projecting the non-dominated solutions onto the first $m-1$ coordinates of this rotated system yields a representation in which the front is single-valued, and a surrogate can be fitted as a function of $m-1$ variables. It is important to highlight that the choice of an orthonormal basis for the hyperplane is not unique. Still, any basis produces a different parametrisation of the same reconstructed graph. Hence, the geometric object recovered is invariant to that choice. Second, the single-valuedness guarantee still holds, as if we have two non-dominated points projected onto the same point of the hyperplane, their difference would be parallel to $(1,\ldots,1)$, making one of them strictly worse in every objective and therefore dominated, a contradiction. The main remaining challenge, however, is the hypervolume maximisation in Section 4.4, whose cost grows with m. Efficient approximations such as Monte Carlo estimators would be needed to keep the overhead tractable. In short, the simplex view shows that the difficulty of extending the method is no longer geometric but computational, and is concentrated entirely in the hypervolume step.

A second direction concerns hyperparameter selection, as the Phase 1 budget and the per-update iteration count were tuned empirically and held fixed during all problems. An adaptive policy that triggers weight updates based on HV stagnation would remove this

manual tuning.
SA-MOEA/D-AWA demonstrates that fitting a surrogate directly in the objective space, rather than in the decision space, is a viable and effective strategy for guiding the weight adjustment of decomposition-based MOEAs towards geometrically difficult regions of the Pareto front.

References

[1] Nicola Beume, Boris Naujoks, and Michael Emmerich. Sms-emoa: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007.

[2] Wei-Mei Chen, Hsien-Kuei Hwang, and Tsung-Hsi Tsai. Maxima-finding algorithms for multidimensional samples: A two-phase approach. *Computational Geometry*, 45(1):33–53, 2012.

[3] Ran Cheng, Yaochu Jin, Markus Olhofer, and Bernhard Sendhoff. A reference vector guided evolutionary algorithm for many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 20(5):773–791, 2016.

[4] Tinkle Chugh, Yaochu Jin, Kaisa Miettinen, Jussi Hakanen, and Karthik Sindhya. A surrogate-assisted reference vector guided evolutionary algorithm for computationally expensive many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 22(1):129–142, 2018.

[5] Tinkle Chugh, Karthik Sindhya, Jussi Hakanen, and Kaisa Miettinen. Handling computationally expensive multiobjective optimization problems with evolutionary algorithms : A survey, 05 2015.

[6] Carlos A. Coello Coello and Margarita Reyes Sierra. A study of the parallelization of a coevolutionary multi-objective evolutionary algorithm. In Raúl Monroy, Gustavo Arroyo-Figueroa, Luis Enrique Sucar, and Humberto Sossa, editors, *MICAI 2004: Advances in Artificial Intelligence*, pages 688–697, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[7] Indraneel Das and J. Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Structural Optimization*, 14:63–69, 01 1997.

[8] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[9] Kalyan Deb. *Multiobjective Optimization Using Evolutionary Algorithms. Wiley, New York.* 01 2001.

[10] Kalyan Deb, Kaisa Miettinen, and Shamik Chaudhuri. Toward an estimation of nadir objective vector using a hybrid of evolutionary and local search approaches. *Evolutionary Computation, IEEE Transactions on*, 14:821 – 841, 01 2011.

[11] Kalyanmoy Deb, Lothar Thiele, Marco Laumanns, and Eckart Zitzler. *Scalable Test Problems for Evolutionary Multiobjective Optimization*, pages 105–145. 03 2006.

[12] C.M. Fonseca, Luis Paquete, and Manuel López-Ibáñez. An improved dimension-sweep algorithm for the hypervolume indicator. pages 1157 – 1163, 08 2006.

[13] Peter I. Frazier. A tutorial on bayesian optimization, 2018.

[14] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.

[15] J. Knowles. Parego: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 10(1):50–66, 2006.

[16] Gary B. Lamont and David A. van Veldhuizen. Multiobjective evolutionary algorithms: classifications, analyses, and new innovations. 1999.

[17] Yuri Cossich Lavinas, Abe Mitsu Teru, Yuta Kobayashi, and Claus Aranha. MOEA/D with adaptative number of weight vectors. *CoRR*, abs/2109.05741, 2021.

[18] Yutao Qi, Xiaoliang Ma, Fang Liu, Licheng Jiao, Jianyong Sun, and Jianshe Wu. Moea/d with adaptive weight adjustment. *Evolutionary computation*, 22, 06 2013.

[19] Rui Wang, Qingfu Zhang, and Tao Zhang. Decomposition-based algorithms using pareto adaptive scalarizing methods. *IEEE Transactions on Evolutionary Computation*, 20(6):821–837, 2016.

[20] Zhenkun Wang, Yew-Soon Ong, Jianyong Sun, Abhishek Gupta, and Qingfu Zhang. A generator for multiobjective test problems with difficult-to-approximate pareto front boundaries. *IEEE Transactions on Evolutionary Computation*, 23(4):556–571, 2019.

[21] Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *Evolutionary Computation, IEEE Transactions on*, 11:712 – 731, 01 2008.

[22] Qingfu Zhang, Wudong Liu, Edward Tsang, and Botond Virginas. Expensive multiobjective optimization by moea/d with gaussian process model. *IEEE Trans. Evolutionary Computation*, 14:456–474, 06 2010.

[23] Eckart Zitzler, Kalyan Deb, and Lothar Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary computation*, 8:173–95, 02 2000.

[24] Eckart Zitzler, Lothar Thiele, Marco Laumanns, C.M. Fonseca, and Viviane Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *Evolutionary Computation, IEEE Transactions on*, 7:117 – 132, 05 2003.

Chapter A Convergence plots for the full benchmark suite

Section 5.2.4 reported the hypervolume convergence on six representative problems selected to cover the full benchmark suite. For completeness, this appendix provides the convergence curves for all 18 benchmark problems. As in the main text, the y-axis shows the HV gap on a logarithmic scale, so that lower values indicate better approximations; solid lines denote the median across the 10 independent runs, and the shaded bands cover the 10%–90% quantile range. The trends discussed in Section 6 hold across the full suite: SA-MOEA/D-AWA separates from MOEA/D after the first weight update on the geometrically challenging problems, while on the decision-space-hard and severely disconnected problems, the two methods remain comparable.

Chapter B Final Pareto front approximations for the full benchmark suite

Section 5.2.5 compared the final populations of SA-MOEA/D-AWA and MOEA/D on four problems with non-trivial geometries. This appendix extends that visual comparison to the full benchmark suite. For each method and problem, the population shown is the run whose final hypervolume is closest to the median across the 10 independent runs. The true Pareto front is drawn as a thin grey line where known, and the reference point used for the HV computation is reported in each subplot.

[b]0.48

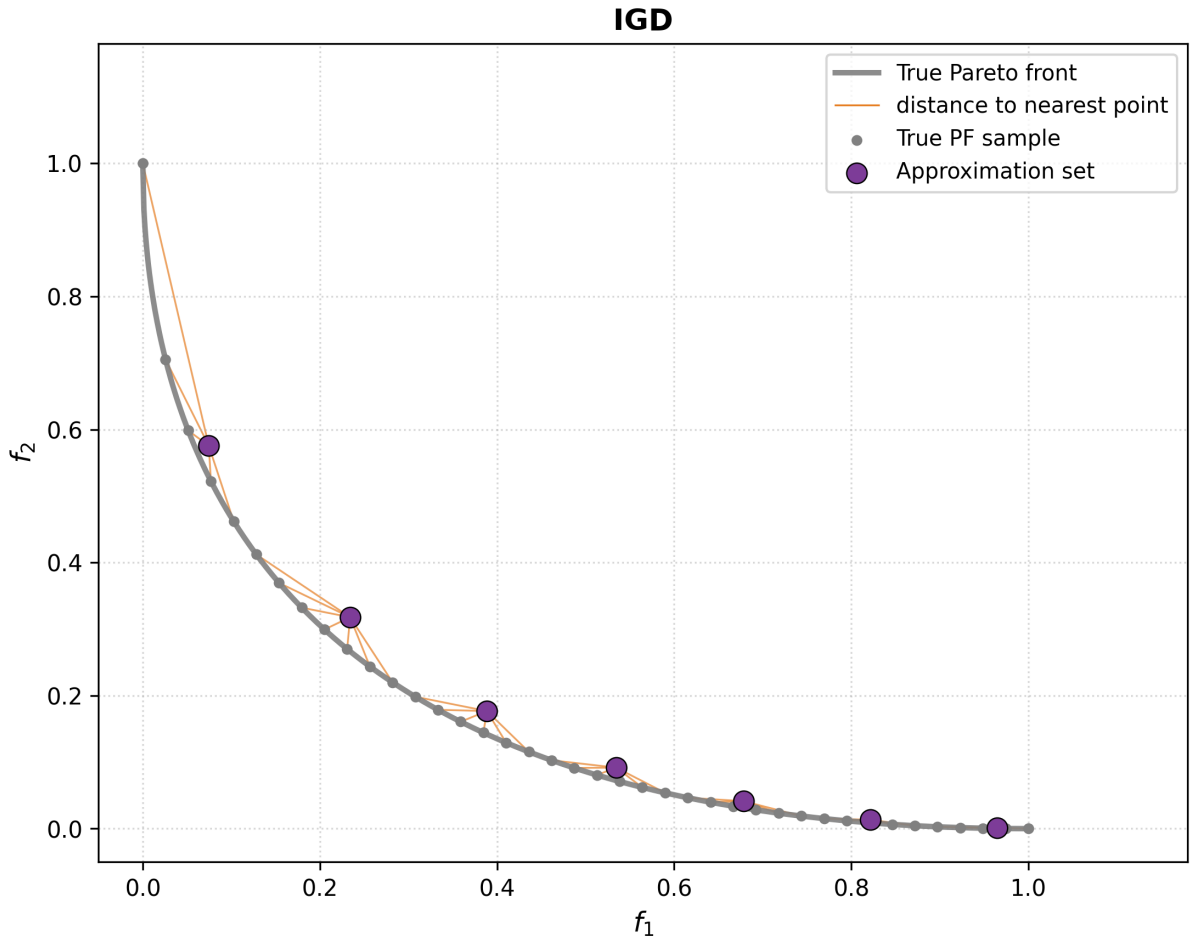
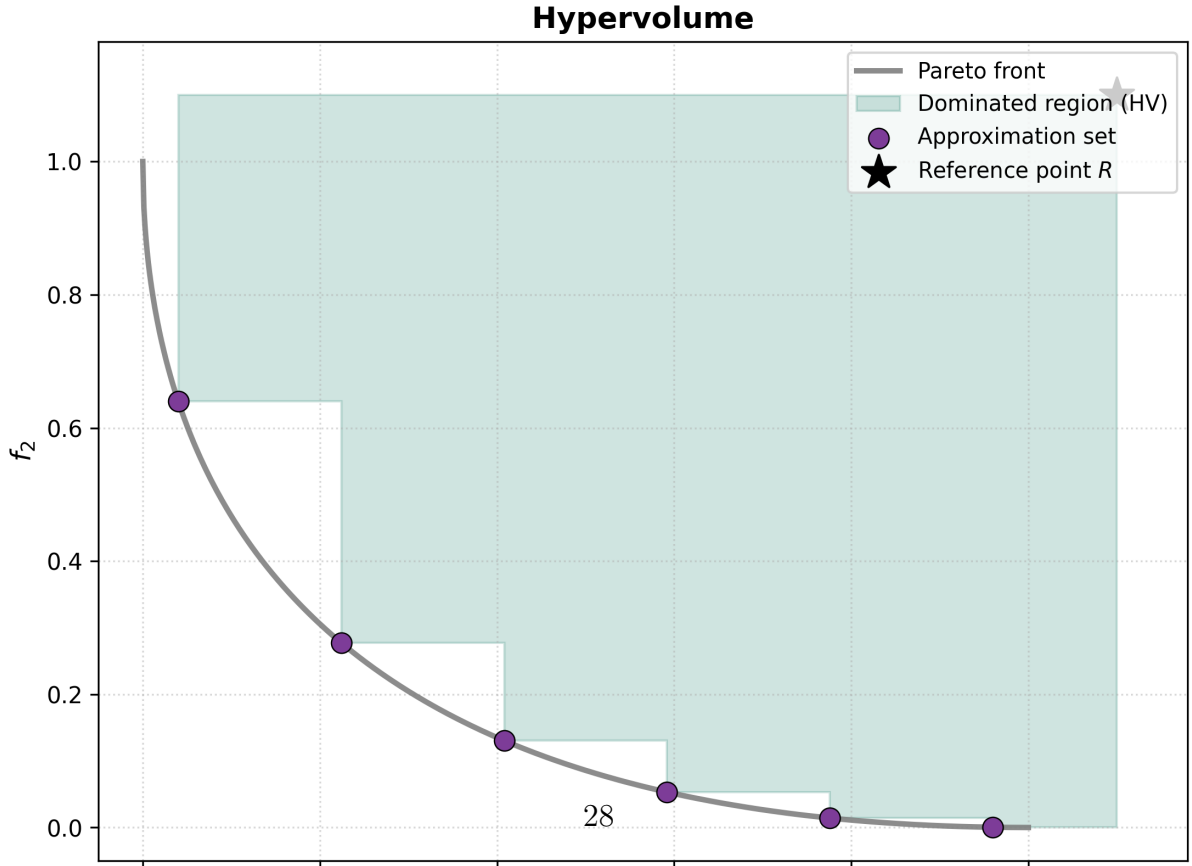


Figure 2: IGD

[b]0.48



[b]0.48

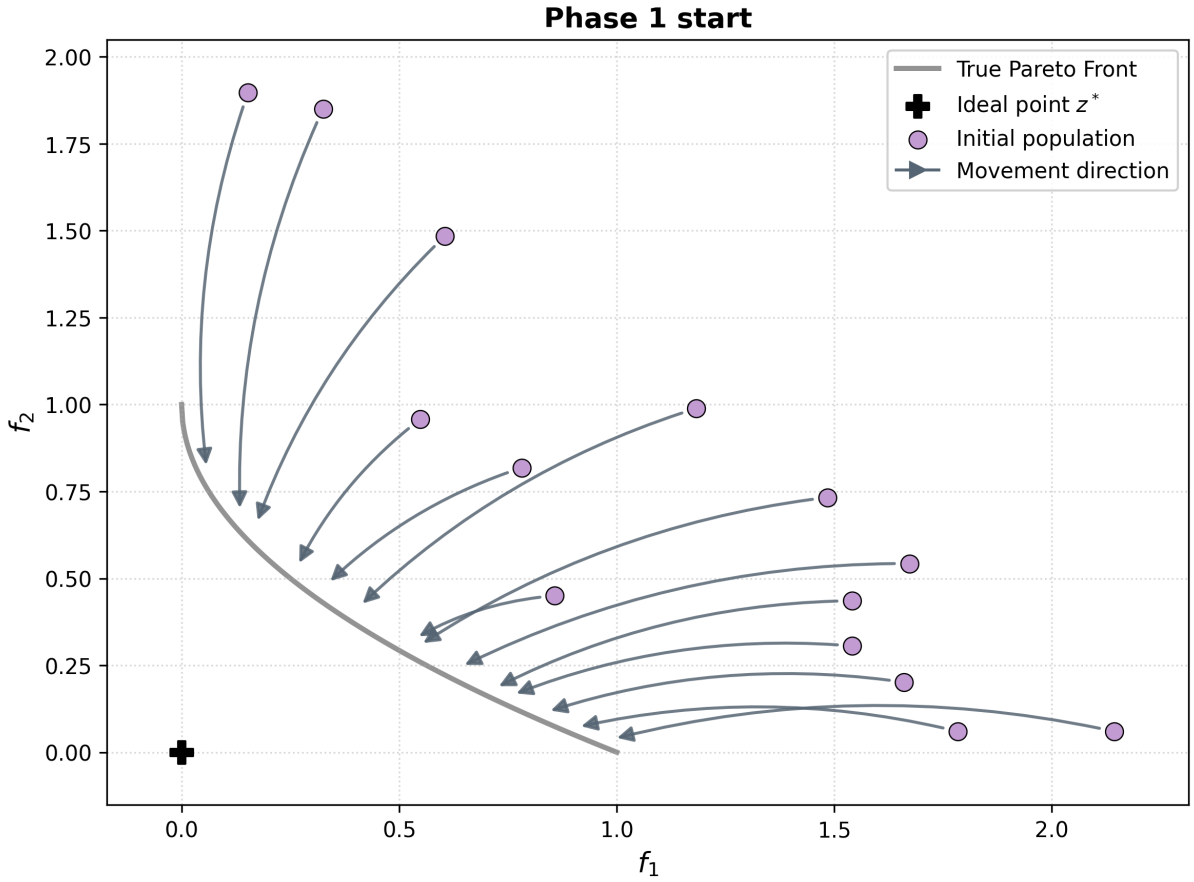
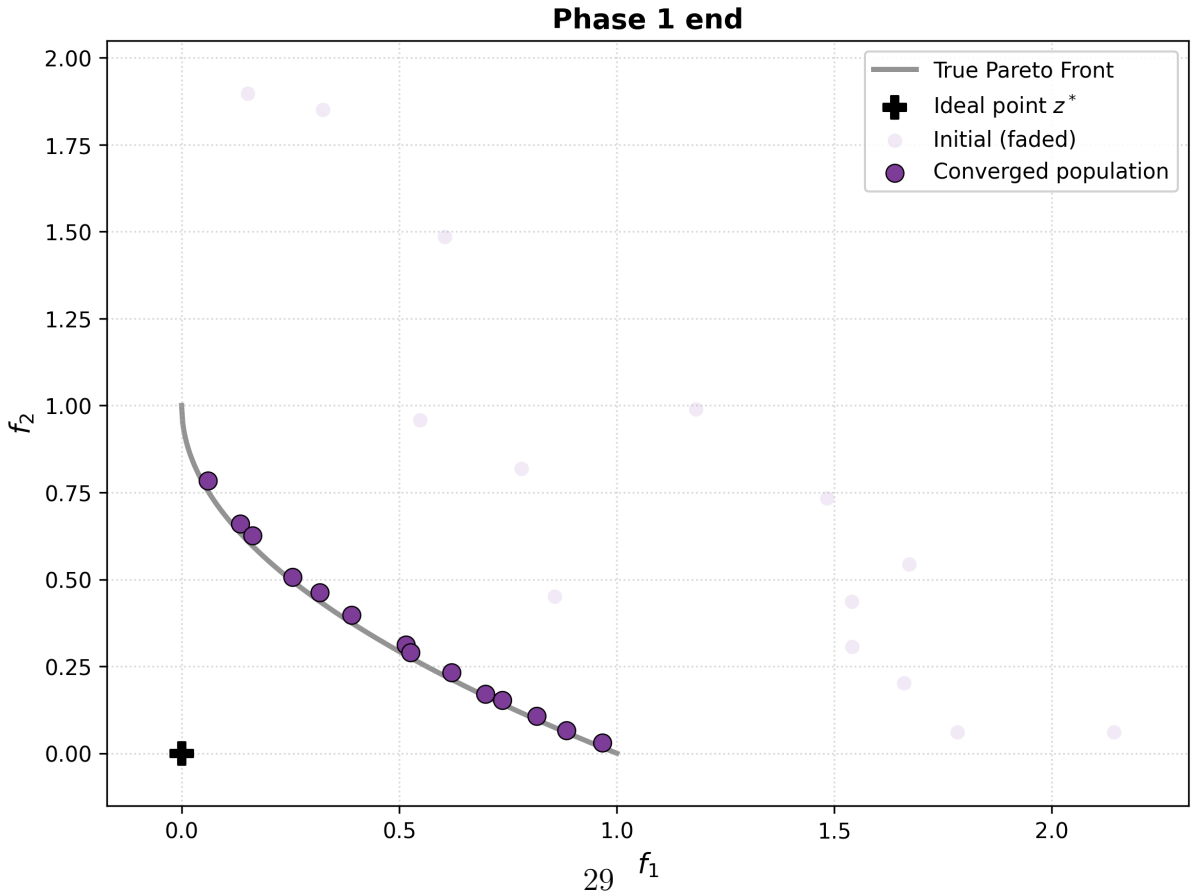


Figure 5:

[b]0.48



[b]0.48

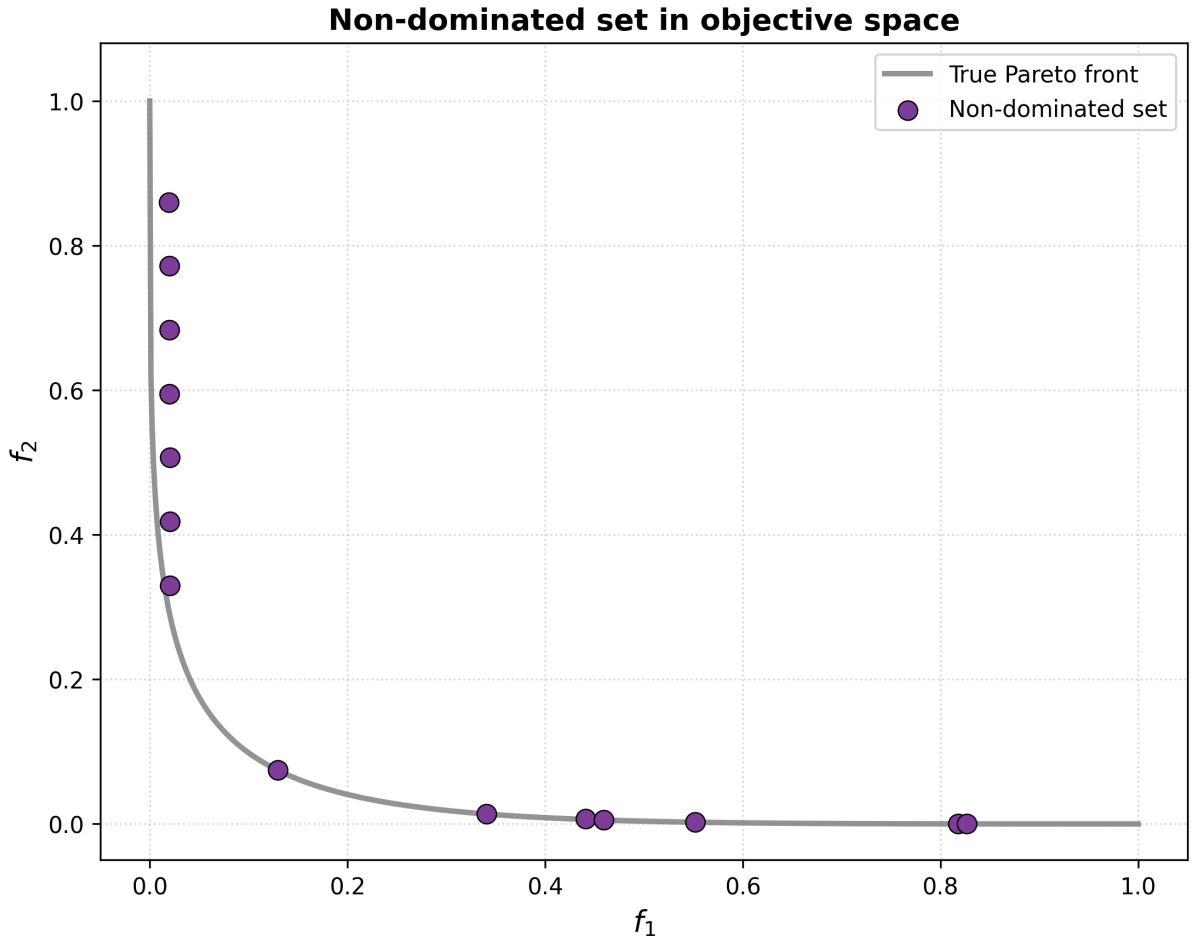
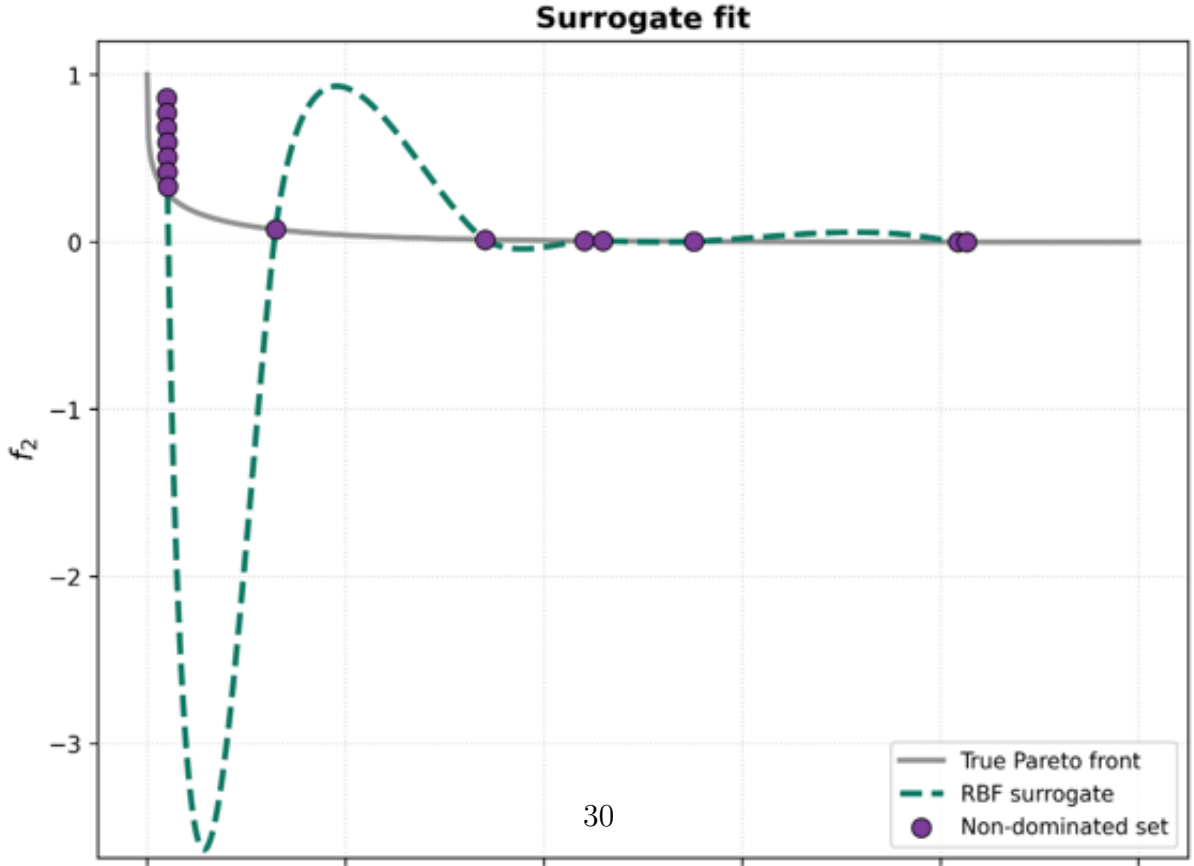


Figure 8: Non-dominated set in the original space.

[b]0.48



[b]0.48

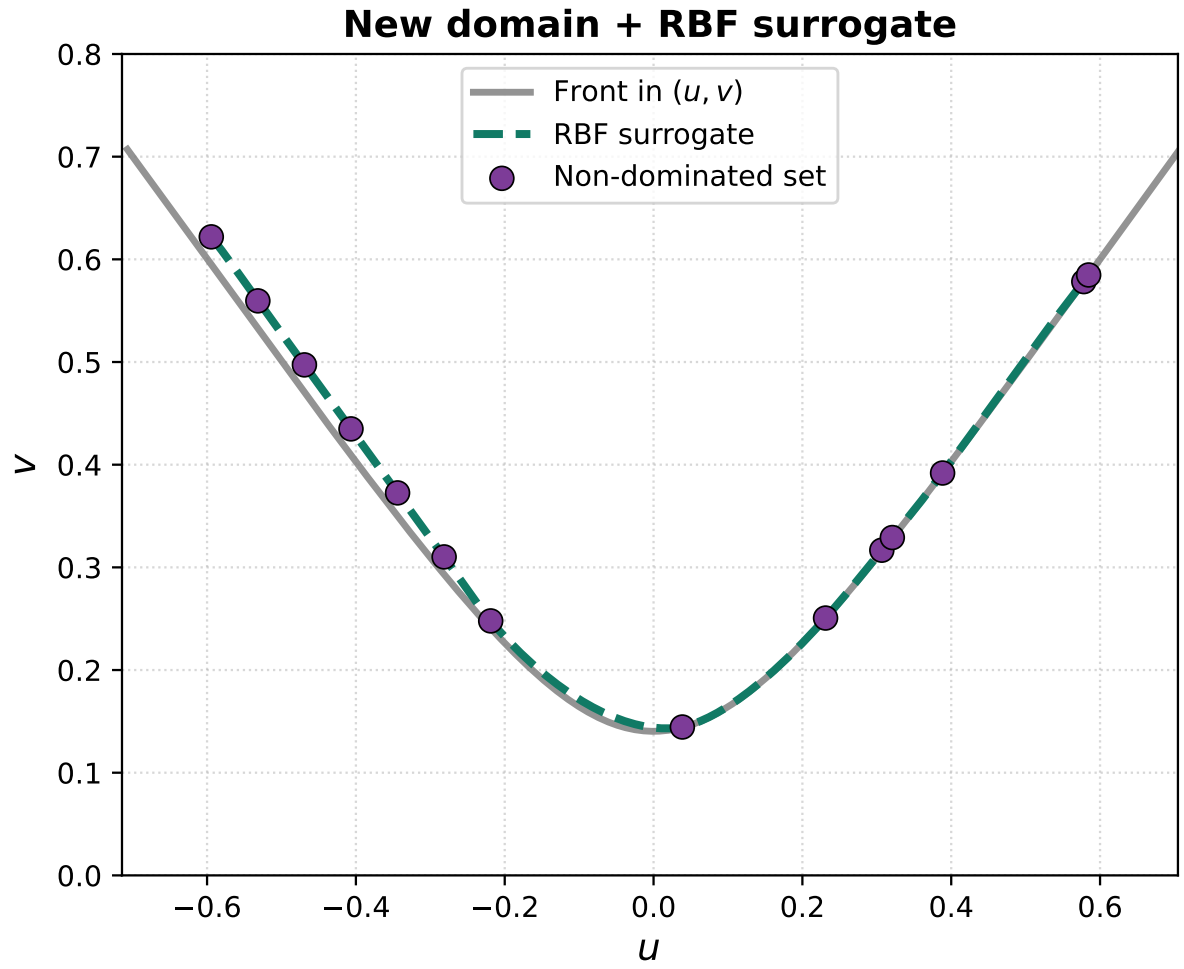
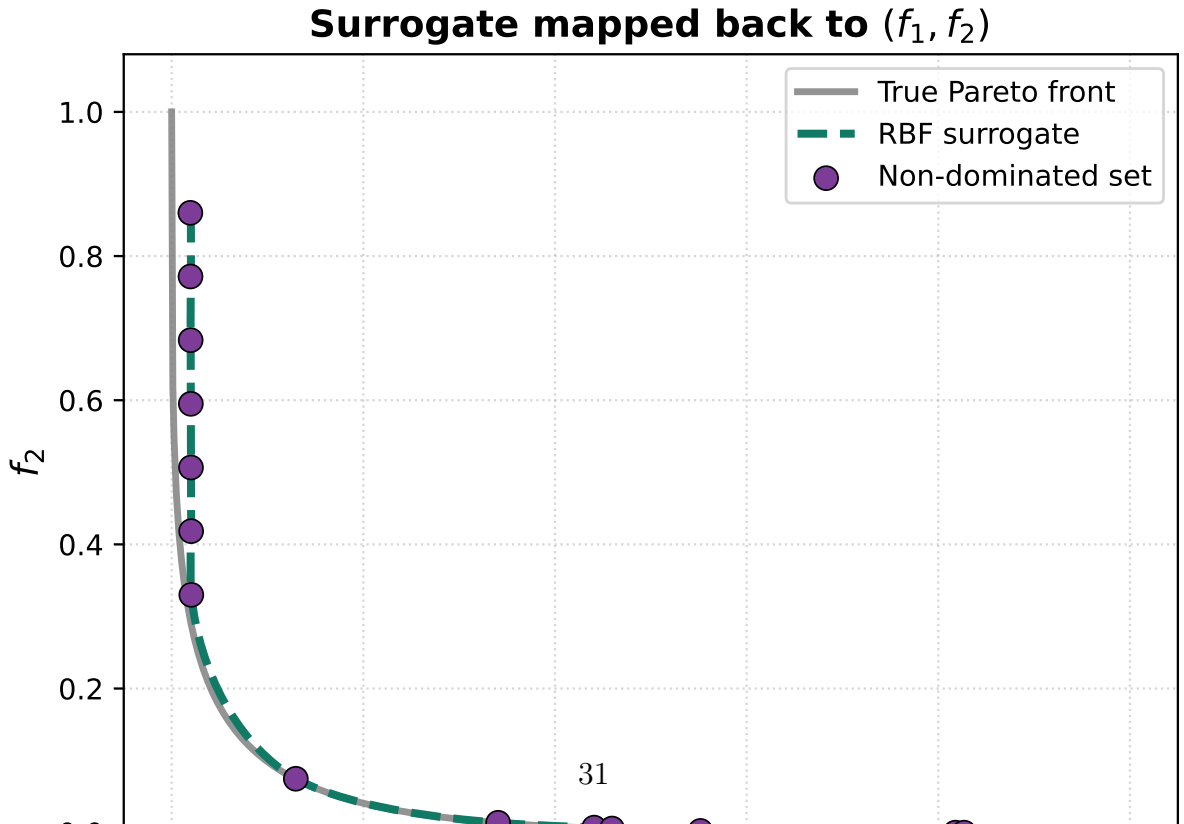


Figure 11: RBF fit in the rotated space.

[b]0.48



[b]0.48

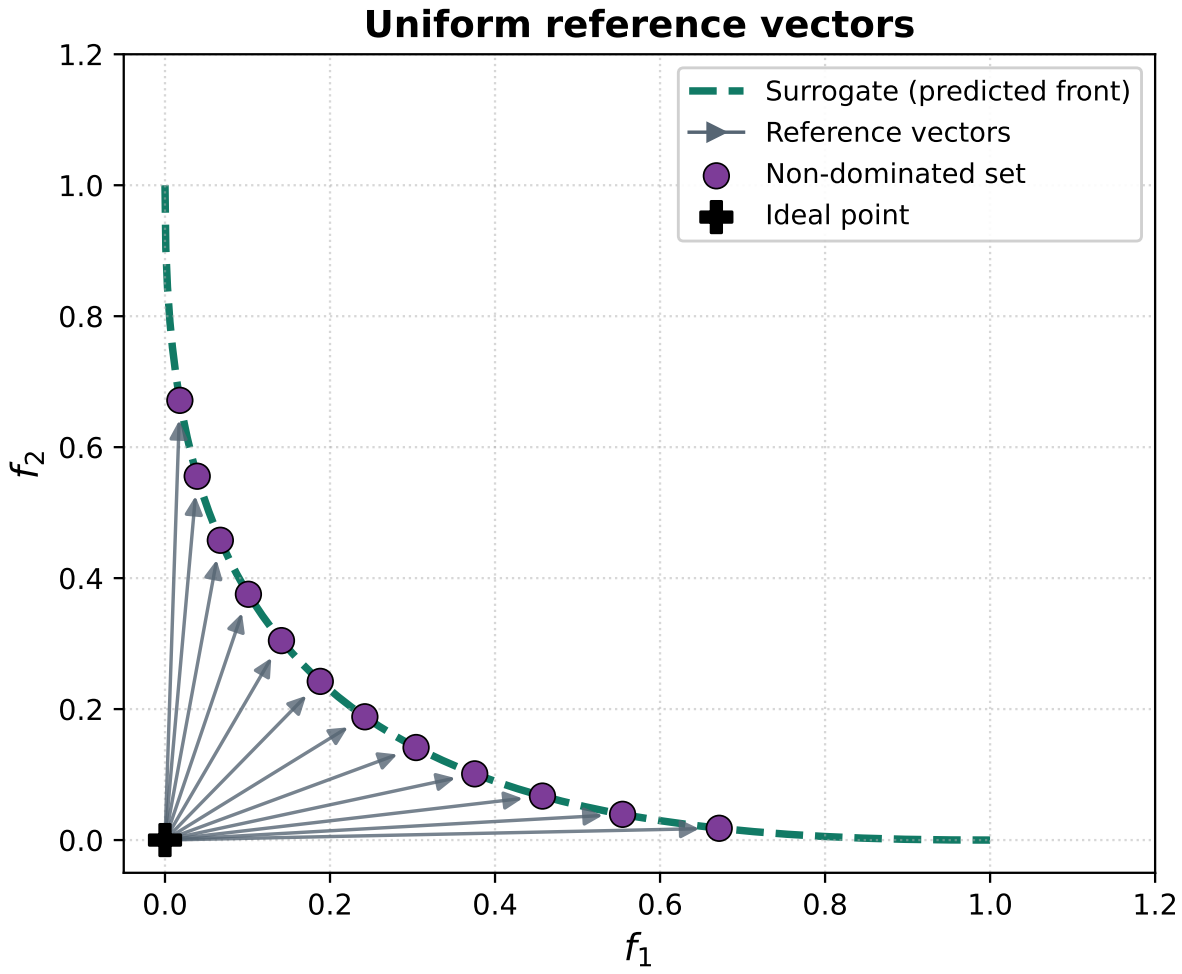
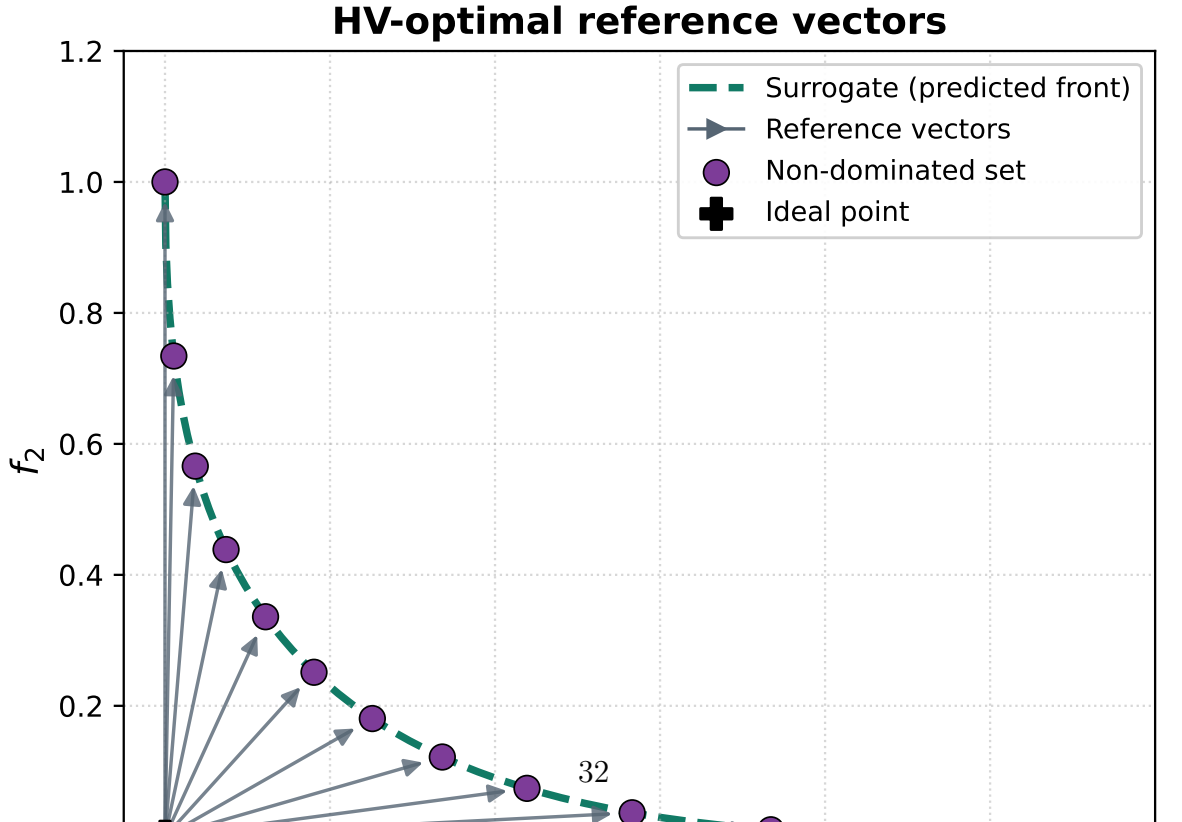


Figure 14: Uniform reference vectors.

[b]0.48



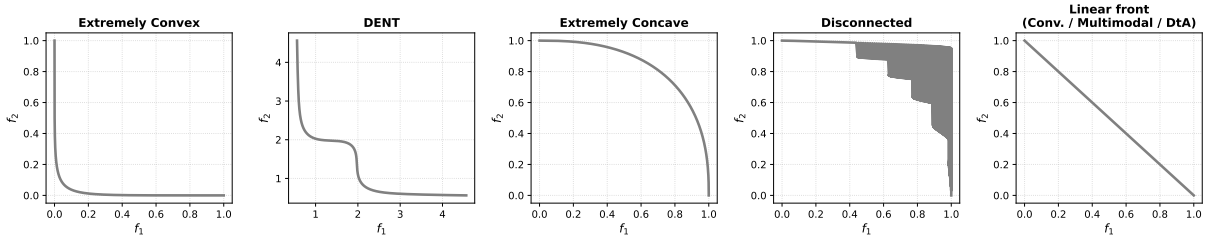


Figure 17: Pareto fronts of the geometrically challenging benchmark problems. The Convergence, Multimodal, and Difficult-to-Approximate problems share the same linear front and are shown once. Their difficulty lies in the decision space rather than in the geometry of the front.

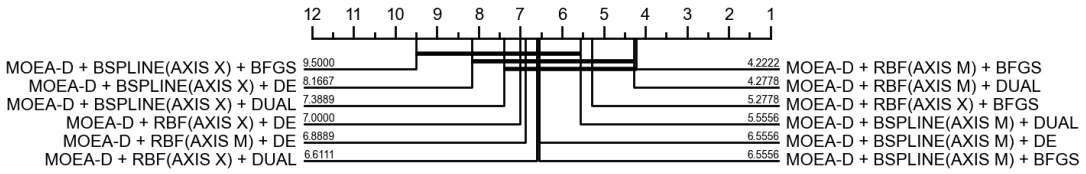


Figure 18: Critical Differences plot for all the problems together.

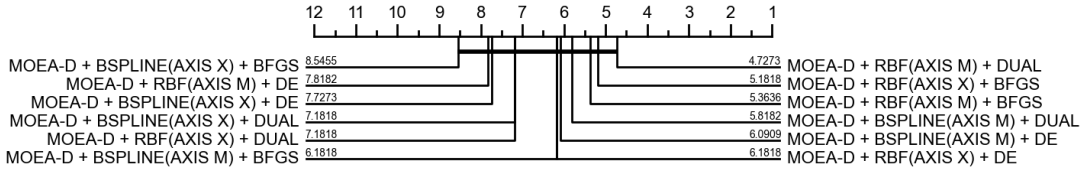


Figure 19: Critical Differences plot for the families of DTLZ and ZDT problems.

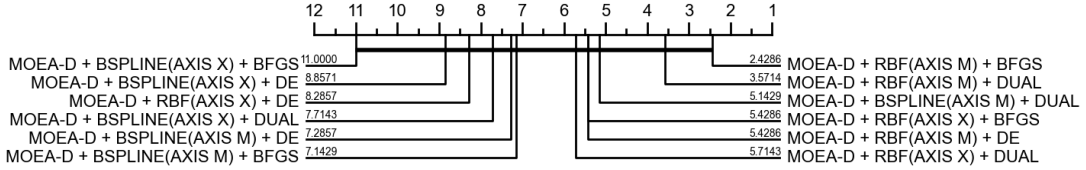


Figure 20: Critical Differences plot for all the Difficult to Approximate problems.

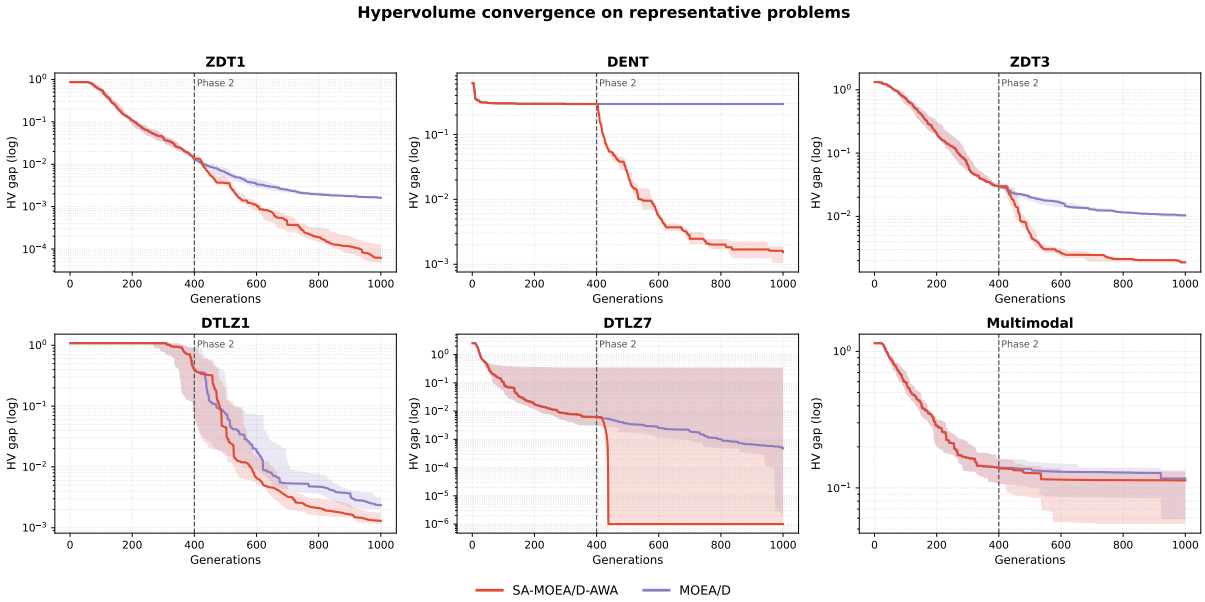


Figure 21: Hypervolume convergence on six representative problems. The y-axis shows the HV gap (lower is better) in a logarithmic scale. Curves are the median over 10 independent runs; shaded bands cover the 10%–90% quantile range.

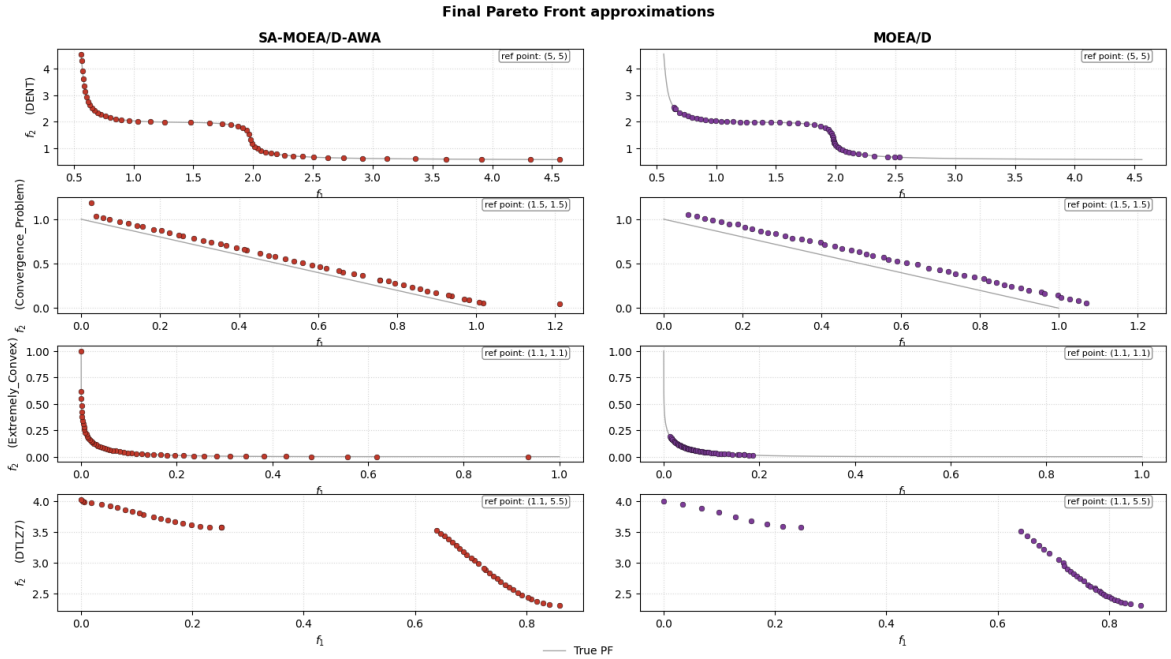


Figure 22: Final Pareto front approximations of SA-MOEA/D-AWA (left) and MOEA/D (right) on four representative problems. Each method is shown for the run with HV closest to the median across 10 independent runs. The grey line indicates the true Pareto front; the reference point used for HV computation is reported in the upper-right inset of each subplot.

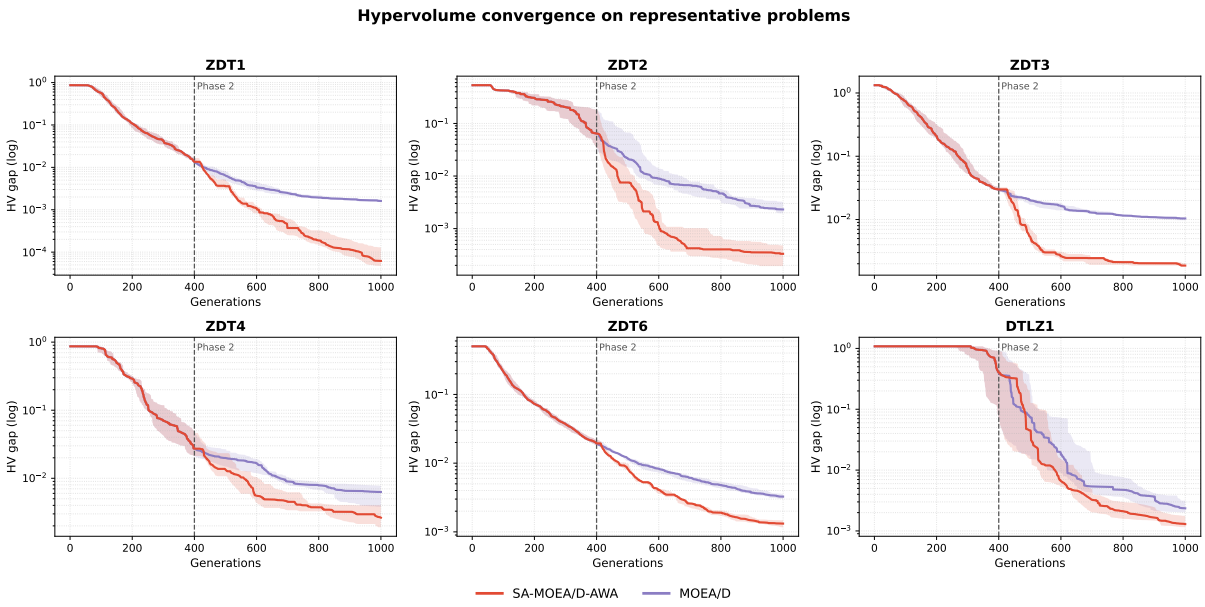


Figure 23: Hypervolume convergence on all the ZDT problems and DTLZ1.

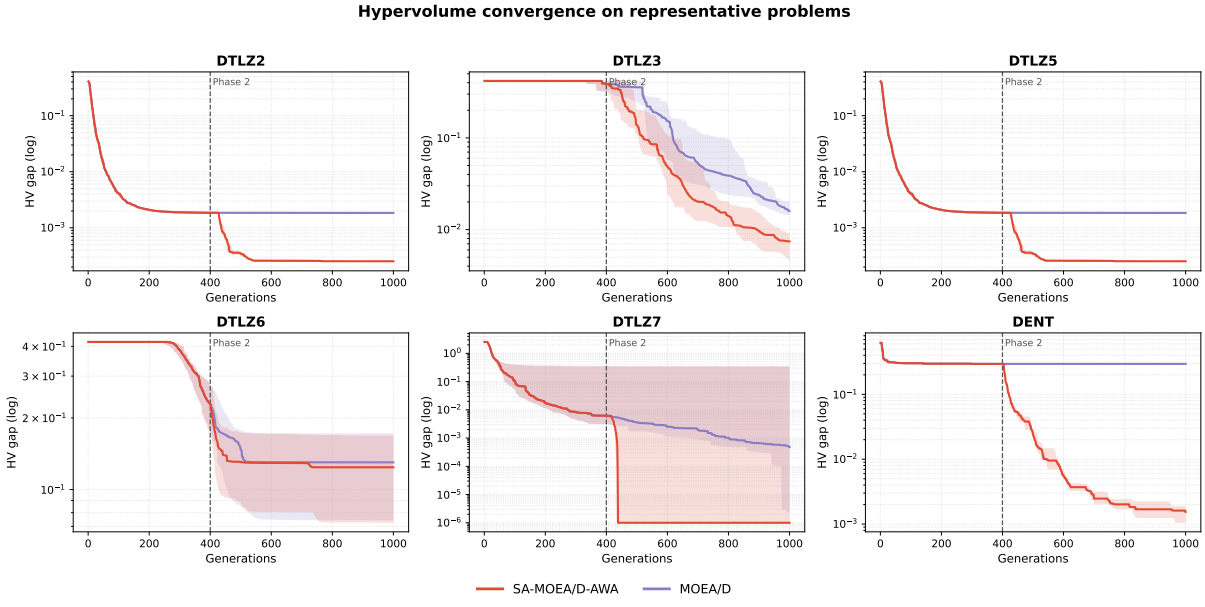


Figure 24: Hypervolume convergence on DENT and all the DTLZ problems, except DTLZ1 and DTLZ4.

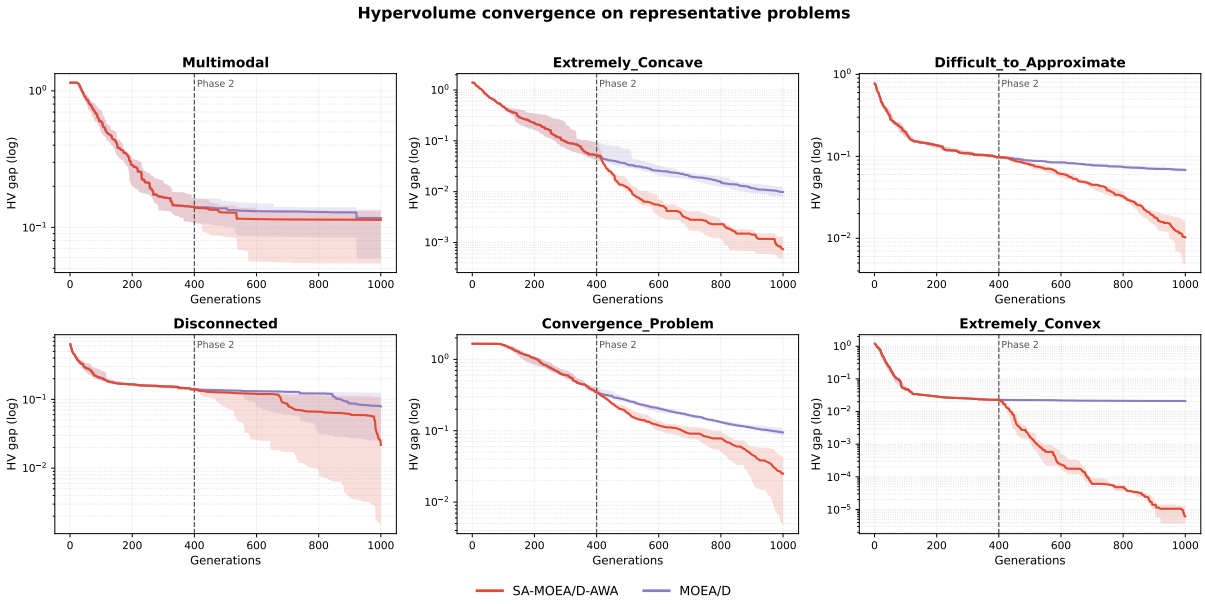


Figure 25: Hypervolume convergence on the rest of the benchmark problems.

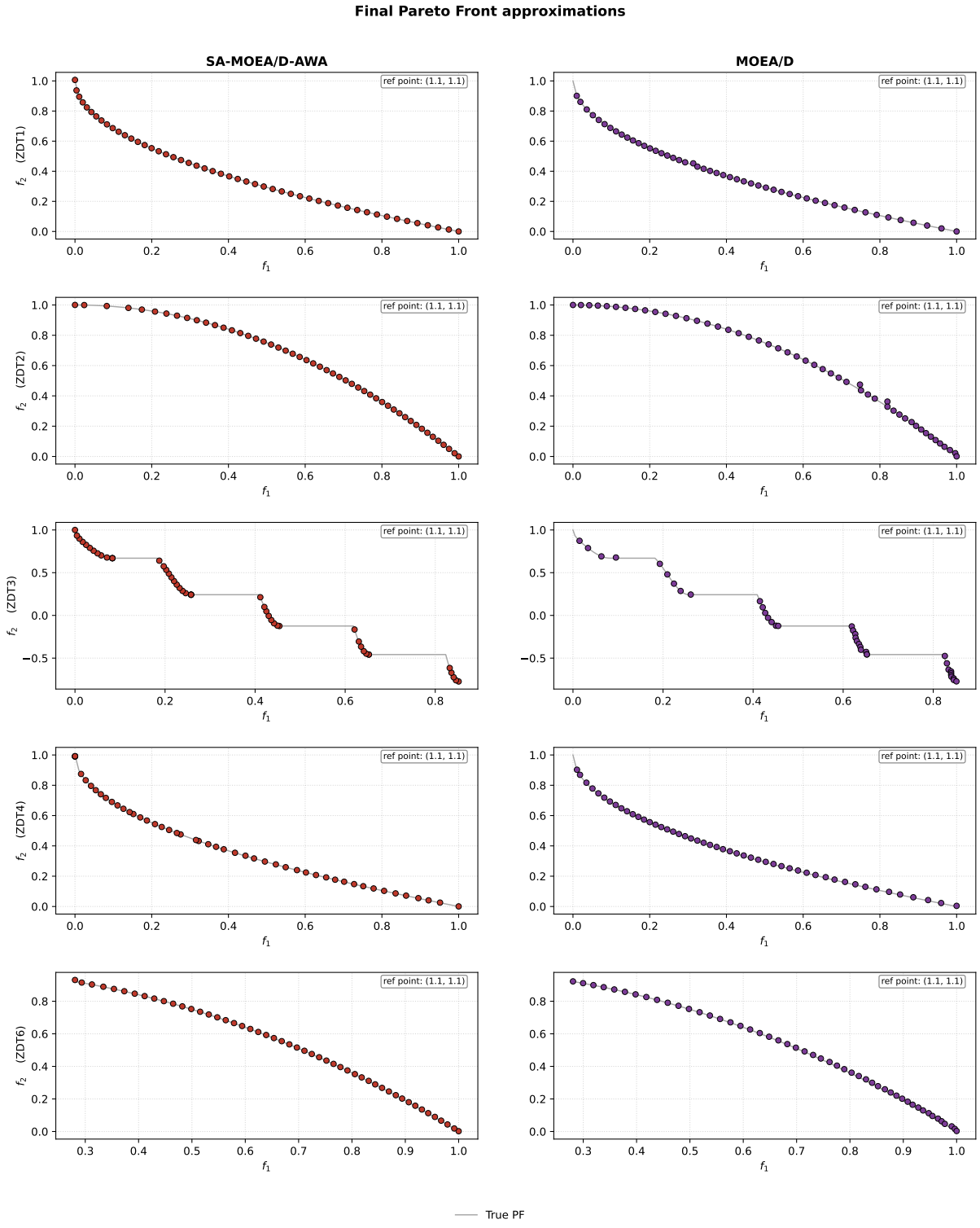


Figure 26: Final Pareto front approximations of SA-MOEA/D-AWA (left) and MOEA/D (right) of all the ZDT problems.

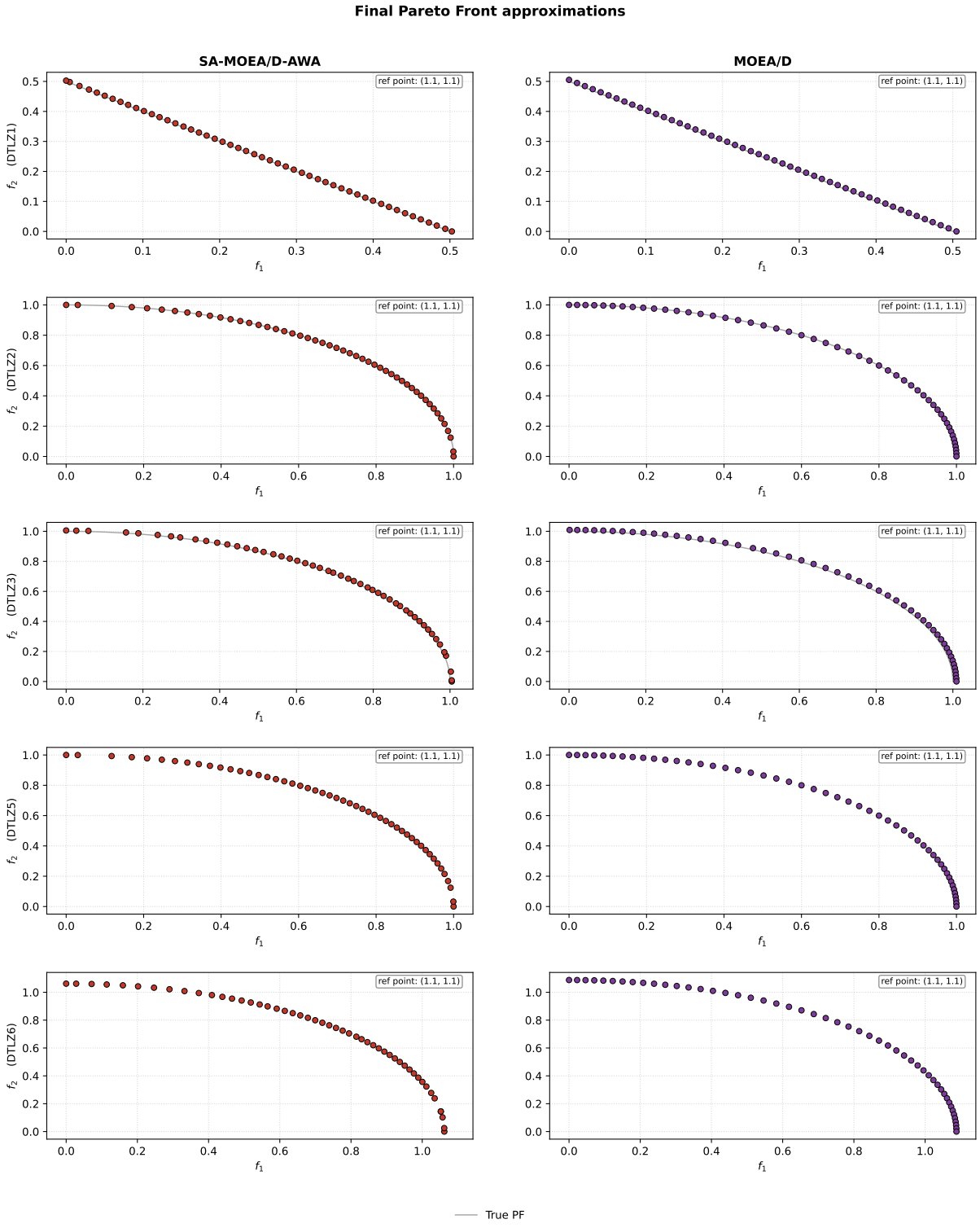


Figure 27: Final Pareto front approximations of SA-MOEA/D-AWA (left) and MOEA/D (right) of all the DTLZ problems, except DTLZ4 and DTLZ7.

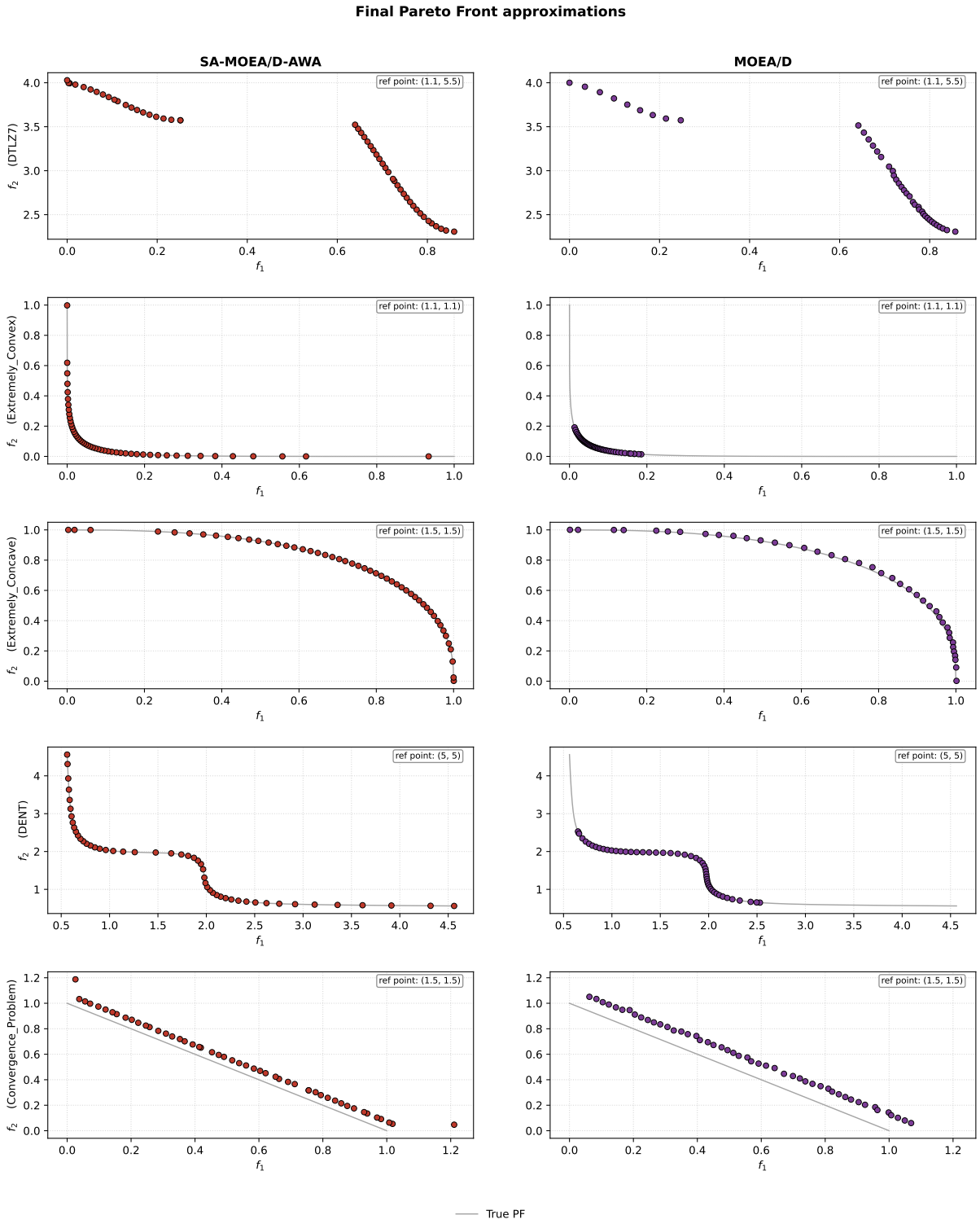


Figure 28: Final Pareto front approximations of SA-MOEA/D-AWA (left) and MOEA/D (right) of DTLZ7, Extremely Convex, Extremely Concave, DENT, and Converge problems.

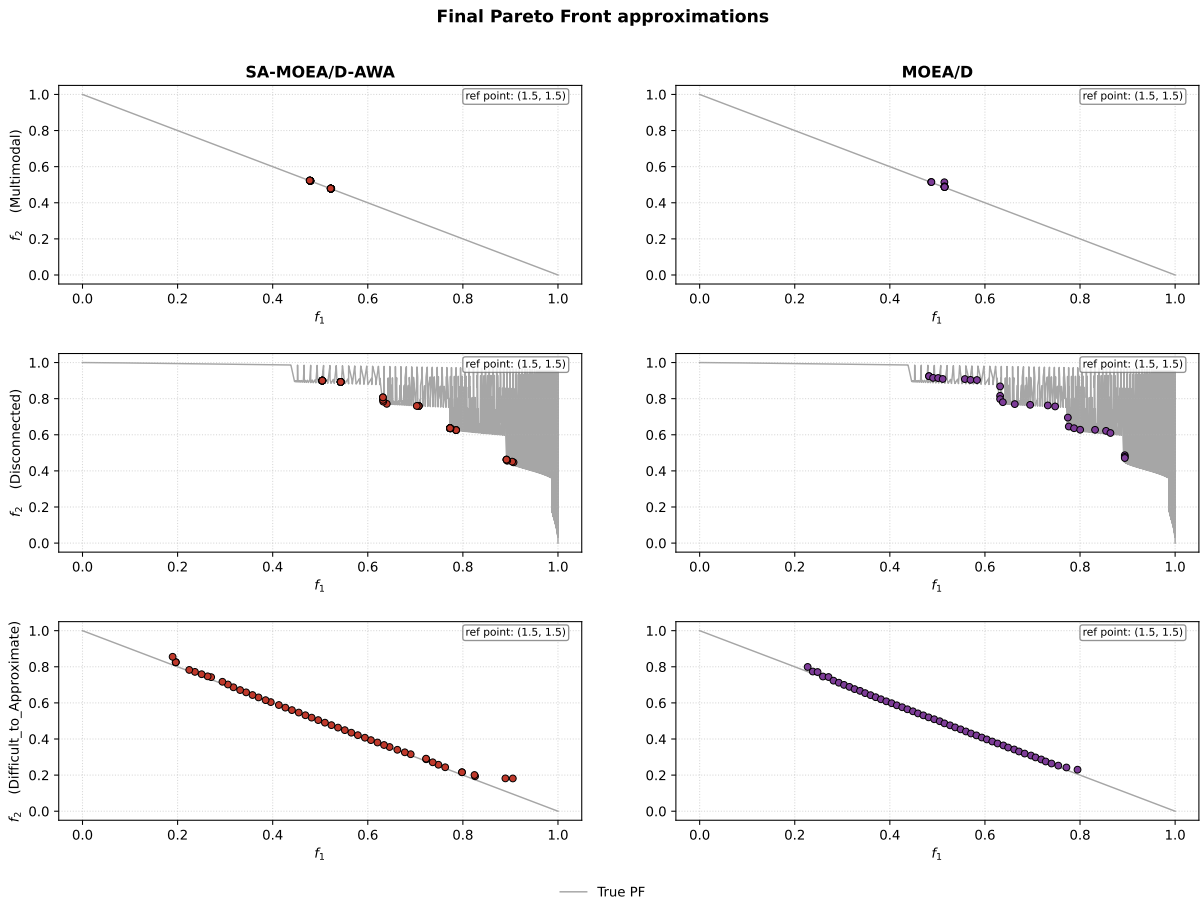


Figure 29: Final Pareto front approximations on the full benchmark suite. On the Multimodal problem (top row), both methods recover only a small portion of the front, as the difficulty of this problem lies in the decision space rather than in the front geometry, so both algorithms remain trapped in local optima under the given budget.