

Original articles

The Directed Multi-Objective Estimation Distribution Algorithm (D-MOEDA)

Salvador Botello-Aceves^a, Arturo Hernandez-Aguirre^a, S. Ivvan Valdez^{b,*}^a Centro de Investigación en Matemáticas A.C., Jalisco S/N, Col. Valenciana, Guanajuato, 36000, Gto, Mexico^b CONAHCYT-Centro de Investigación en Ciencias de Información Geoespacial A.C., Parque Tecnológico Querétaro, San Fandila, 76703, Qro, Mexico

Received 31 October 2022; received in revised form 12 May 2023; accepted 21 July 2023

Available online 27 July 2023

Abstract

Improvement Direction Mapping (IDM) methods have been applied as a local search strategy to hybridize global search algorithms. A natural question is whether this concept could be applied within a global search scheme, so that the stochastic search operators are directed toward promising regions, promoting a more efficient search. This paper introduces a novel Multi-Objective Evolutionary Algorithm (MOEA) that incorporates the IDM into the reproduction operator of an Estimation of Distribution Algorithm (EDA). In this proposal, the search directions of the IDM based on aggregation functions are used to directly steer the search process of a multi-objective evolutionary algorithm based on decomposition, by orienting a local probability distribution towards a search direction, the proposal intends to steer solutions toward the Pareto front (PF) of the Multi-Objective Optimization Problem (MOP), exploiting the search features of the aggregation functions. The proposal is tested using a set of well-known benchmark MOPs and compared to state of the art MOEAs. Results showed statistical evidence about the importance of the orientation of the search probability distribution to improve the convergence to the Pareto front.

© 2023 International Association for Mathematics and Computers in Simulation (IMACS). Published by Elsevier B.V. All rights reserved.

Keywords: Multi-objective optimization; Hybridization; Improvement Direction Mapping

1. Introduction

Estimation of Distribution Algorithms (EDAs) are a set of well-known Evolutionary Algorithms (EAs) that estimate the parameters θ of a probability distribution $\mathcal{G}(\theta)$, computed from a subset of the best elements. Contrary to other EAs described in the literature, EDAs replace the reproduction operator through the simulation from the probability distribution. The central idea of an EDA arose as an approximation of the true probability distribution function of the selected population. This approach has been gradually abandoned, to an approach of acquiring MOP information from the probability distribution in order to adapt the search performance of the algorithm. By the estimation of the parameters of the probability distribution, the sampling explores the neighborhood of the

* Corresponding author.

E-mail addresses: salvador.botello@cimat.mx (S. Botello-Aceves), artha@cimat.mx (A. Hernandez-Aguirre), sergio.valdez@conacyt.mx (S.I. Valdez).

<https://doi.org/10.1016/j.matcom.2023.07.013>

0378-4754/© 2023 International Association for Mathematics and Computers in Simulation (IMACS). Published by Elsevier B.V. All rights reserved.

probability distribution of the best solutions. Therefore, the most relevant characteristics of the optimization problem are generally captured by the probability distribution, which allows the construction of ideal conjectures about the problem and, therefore, better control of the evolution.

Multi-objective optimization based on EDAs has been addressed in the literature. For instance, the Multi-Objective Hierarchical Bayesian Optimization Algorithm (mhBOA) [12,13] replaces the standard selection, crossover, and mutation operators present in the Multi-Objective Genetic Algorithms (MOGA) based on Pareto dominance over a binary domain. The reproductive operator of the mhBOA computes a probability model based on Bayesian inference so that the joint probability of the model over the selected populations is maximized. Notice that as the estimation parameters are computed from the selected population, essential features from the MOP are collected by the probability model, such as the most correlated decision variables and their effects on the objective functions. A similar approach is presented in Laumanns and Ocenasek [14], where the Multi-Objective Mixed Bayesian Optimization Algorithm (mmBOA), where the selection operator and dominance relationship are taken from the Strength Pareto Evolutionary Algorithm II (SPEA-II) [22]. Similar results are presented for both mhBOA and mmBOA, allowing the assumption that the probability model captures the guiding traits to explore the search space.

In Pelikan et al. [15] the Multi-Objective Extended Compact Genetic Algorithm (mECGA) is introduced. The mECGA features the non-dominated sorting selection and the elitism replacement from the Fast Non-dominated Sorting Genetic Algorithm (NSGA-II) [6] for real value domain. At the same time, the reproductive operator samples new candidate solutions through the parameter estimation of a probability model known as the marginal product model, derived from the single-objective optimization method ECGA [8]. Furthermore, the Multi-Dimensional Bayesian Network Estimation of Distribution Algorithm (MBN-EDA) is introduced in Karshenas et al. [11] to tackle Many-Objective Optimization Problems. Similar to the previous approaches, to sample new candidate solutions from a probability distribution based on a Bayesian network. However, contrary to previous methods, the non-dominated sorting dominance relationship from the NSGA-II is not used, mainly due to the reduction of selection pressure as the number of objectives increases, affecting the effectiveness of this sorting process. Therefore, the selection operator of the MBN-EDA applies four non-Pareto compliant scalarizing functions to sort the current population: Weighted sum, distance to the ideal point, total gain, and profit.

The Decision-Tree-based Multi-Objective Estimation of Distribution Algorithm (DT-MEDA) [20] for continuous variables builds a probabilistic decision tree model from a selected population, computed through the application of the non-dominated sorting selection and the elitism replacement from the NSGA-II, to sample new candidate solutions. The Regularity Model-Based Multi-objective Estimation of Distribution Algorithm (RM-MEDA) [18] uses a local principal component analysis algorithm to build a probability distribution whose centroid is an $(m - 1)$ dimension piecewise continuous manifold. Due to the application of the principal component analysis, the RM-MEDA adequately deals with MOP with variable linkages, outperforming the compared optimization methods.

Furthermore, it is worth noting that the application of MOEDAs to address time-variant multi-objective optimization problems, commonly known as Dynamic Multi-Objective Optimization Problems (DMOPs) [10], has increased in recent years, mainly due to their capabilities to collect features from the original DMOP and predicting the next non-dominated set in the subsequent time step.

This paper studies and discusses the search features and properties of Multi-objective optimization algorithms based on EDAs. Three steering cases for a probability distribution of MOEDAs are studied. Overall, it is shown non-oriented probability distribution tends to a rapid stagnation. Thus, this work introduces the Directed Multi-Objective Estimation of Distribution Algorithm (D-MOEDA). This approach addresses the premature stagnation on locally Pareto fronts present on the Multi-Objective Estimation of Distribution Algorithm by reorienting the covariance matrix of a multivariate Gaussian distribution. Based on a rank-one matrix, the search direction of an Improvement Direction Mapping (IDM) method based on aggregation functions is introduced to the covariance matrix in order to direct the search toward promising regions.

IDM methods are a family of novel Multiobjective Local Search algorithms (MOLSAs), that steer the approximation set towards the Pareto set by first computing an improvement direction in the objective functions space, and then transforming it into an improvement direction in the decision variables space, for each solution in the current solution set.

The organization of this work is as follows. Section 2 presents the background and notations used to present the sequel of the proposed Multi-Objective Estimation of Distribution Algorithm. Section 3 introduces the Directed

Multi-Objective Estimation of Distribution Algorithm and the orientation features are presented. To study the performance of the proposed method, a set of experiments are conducted, and their results are discussed in Section 4. Concluding remarks and future work are presented in Section 5.

2. Background

For the sake of the following discussion, this section briefly introduces theoretical insight of Multi-Objective Optimization Problems (MOP) and Multi-Objective Local Search Algorithms (MOLSA). Topics presented in this section will be later referred throughout the text.

2.1. Multi-objective optimization problems

A MOP is composed of a set of m conflicting objectives. This work is focused on continuous and unconstrained MOPs. An unconstrained MOP consists of two basic mathematical entities: (1) The $n \geq 1$ *decision variables* and (2) the $m \geq 2$ *objective functions*. In consequence, an unconstrained MOP is described as in Eq. (1).

$$\min_{\mathbf{x} \in \mathbb{R}^n} \mathcal{F}(\mathbf{x}), \quad (1)$$

where $\mathbf{x} \in \Omega \subset \mathbb{R}^n$ are the decision variables, $\mathcal{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the objectives vector, and $f_i \in \mathcal{F} : \mathbb{R}^n \rightarrow \mathbb{R}$ is the i th objective functions vector. To confine the search, the decision variables are bounded by box constraints, $x_i^L \leq x_i \leq x_i^U \forall i = 1, \dots, n$.

The Jacobian of the MOP $\mathcal{F}(\mathbf{x})$ is given by the first derivative of all objective functions, as:

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \nabla f_1(\mathbf{x})^\top \\ \vdots \\ \nabla f_m(\mathbf{x})^\top \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad (2)$$

where $\mathbf{J}(\mathbf{x})$ denotes the Jacobian evaluated at \mathbf{x} and $\nabla f_i(\mathbf{x})$ the gradient of objective f_i .

2.2. Improvement direction mapping methods

Improvement Directions Mapping (IDM) methods are a family of line search MOLSAs that proposes improvement directions in the objective space and transforms them to search directions in the variable space. The IDM methods present a proper control of the convergence toward the Pareto front (PF), distribution of the solutions, and their spread of the Approximation front (AF).

The prediction of a favorable search direction in the search space of an IDM method is driven through the mapping of improvement directions \mathbf{p} in the objective space by applying a linear transformation,

$$\mathbf{d} = \mathbf{H}(\mathbf{x})\mathbf{p}, \quad (3)$$

where $\mathbf{H}(\mathbf{x}) \in \mathbb{R}^{n \times m}$ is the spatial transformation tensor and $\mathbf{p} \in \mathbb{R}^m$ is an improvement direction.

Overall, an IDM method is decoupled into two main tasks. For every i th solution in the current Approximation set, an IDM method performs (1) The computation of an improvement direction \mathbf{p} , on the objective space, within the improvement direction set P . (2) The computation of a transformation tensor $\mathbf{H}(\mathbf{x})$, such that the displacement on the objective space is aligned to the improvement direction. Each task is briefly introduced. For a deeper review, refer to Botello-Aceves et al. [1,2,3].

2.2.1. Improvement directions computation

For a given solution \mathbf{x} in the current Approximation set (AS), an improvement direction of its corresponding image $\mathcal{F}(\mathbf{x})$ must meet at least one of the following three goals:

- **Minimization in the objective space:** The convergence of the AF towards the PF requires the improvement direction to outperform the image of the current solution
- **Distribution maintenance:** Maintaining the quality of the solutions within the current AF by meeting distribution criteria is essential to any improvement direction.

- **Front full-extension:** The extension of the solutions on the objective space gives one quality aspect of the AF. Therefore, an improvement direction should extend the range of the current AF.

Notice that each goal is related to one of the three characteristics of the AF. Typically, a dominance relationship is used by the MOEAs to control the three characteristics. Therefore, the control of the three goals of the improvement directions is related to a dominance relationship, providing the necessary information to steer the search, maintaining an adequate distribution, increasing the spread of solutions, and improving the current AF.

Related work have classified the computation of improvement directions into three different methods.

1. **Geometric approach:** In Botello-Aceves et al. [1], the computation of improvement directions is approached as the alignment with the normal vectors to the manifold of the Approximation front. Since the real manifold is unknown, a local approximation is applied using a quadratic regression, a tangent vector \mathbf{t} to the quadratic regression is derived from the trade-off description, and an improvement direction \mathbf{p} is computed as the perpendicular vector that maximizes its projection to the reference vector \mathbf{z}_r , as:

$$\mathbf{p} = \mathbf{z}_r - \frac{\mathbf{z}_r^\top \mathbf{t}}{\mathbf{t}^\top \mathbf{t}} \mathbf{t}. \quad (4)$$

2. **Minimization of aggregation function:** In Bouter et al. [4] and Botello-Aceves et al. [2], the search directions are derived from a Deepest Descent and Quasi-Newton approach applied to an aggregation function, respectively. For the sake of the discussion, let us derive the search directions by minimizing the aggregation function through a second-order Taylor approximation, such as:

$$\mathbf{d} = - \left(\left(\sum_{k=1}^m \frac{\partial g}{\partial f_k} \nabla^2 f_k \mathbf{J}^\dagger \right)^\top + \nabla^2 g \mathbf{J} \right)^\dagger \nabla g. \quad (5)$$

where ∇g and $\nabla^2 g$ are the gradient vector and the Hessian of the aggregation function with respect to the objective functions, respectively.

To reduce additional costs and avoid numerical errors, the contribution term of the weighted sum of the second-order derivative of the objective functions is neglected, so the improvement direction is given as:

$$\mathbf{p} = -\nabla^2 g^{-1} \nabla g. \quad (6)$$

3. **Maximization of indicator function:** An indicator function measures the contributions of each solution within the Approximation set, using a scalar indicator. Intuitively, the improvement direction of a solution is computed as the direction that maximizes its contribution. In Botello-Aceves et al. [2], the improvement directions based on indicators are computed by the first derivative of the R2 indicator [7], given as:

$$\nabla_{\mathbf{f}} C_{R_2}(\mathbf{x}) = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \begin{cases} 0 & \text{if } u(\mathbf{x}, v, \mathbf{z}) \neq \min_{\mathbf{a} \in \mathbf{A}} u(\mathbf{a}, v, \mathbf{z}) \\ \nabla_{\mathbf{f}} u(\mathbf{a}, v, \mathbf{z}) & \text{else} \end{cases} \quad (7)$$

where $\nabla_{\mathbf{f}} u(\mathbf{a}, v, \mathbf{z})$ is the utility gradient. Notice that if solution \mathbf{a} does not contribute to the approximation set \mathbf{A} , no direction is computed, thus, an alternative direction is given as $\mathbf{p} = \mathbf{z} - \mathbf{a}$. In Botello-Aceves et al. [2], the standard weighted Tchebycheff function is used as the utility function, given by $u(\mathbf{a}, v, \mathbf{z}) = \max_{j=1, \dots, m} v_j |z_j - a_j|$, hence, the utility gradient is computed as:

$$\frac{\partial}{\partial f_j} u(\mathbf{a}, v, \mathbf{z}) = \begin{cases} v_j & \text{if } j | \max_{j=1, \dots, m} v_j |z_j - a_j| \\ 0 & \text{else} \end{cases} \quad (8)$$

2.2.2. Transformation tensor

A tensor is used as a local coordinate transformation to relate the improvement direction in the objective (\mathbb{R}^m) and the search direction in the search space (\mathbb{R}^n). In the context of similar approaches in the literature, the most commonly used computation methods for transformation tensors are reviewed in Botello-Aceves et al. [3]. Additionally, the process for the computation of transformation tensors is studied and an iterative approach based on the Broyden method is introduced, demonstrating a lesser transformation mismatch and a computational complexity decrease. Overall, transformation tensors are classified into three main categories: (1) Analytic approach:

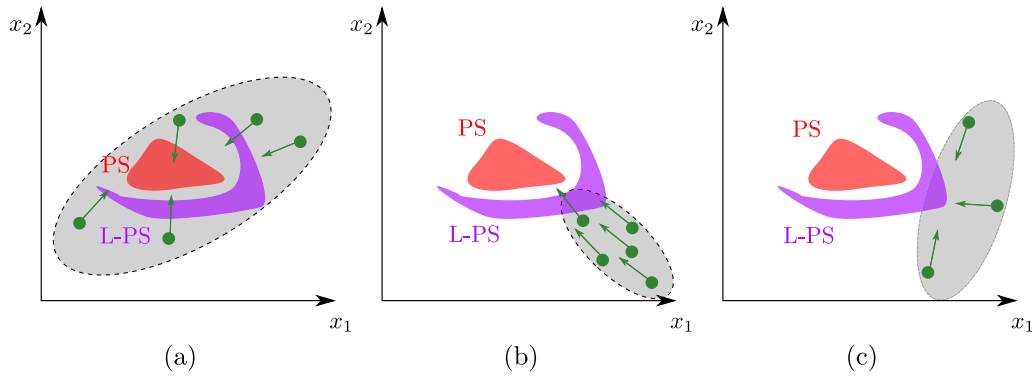


Fig. 1. Visual representation of the three steering cases for a probability distribution. (a) Cases 1: The 90% confidence interval engulf the PS, driving the search towards it. (b) Case 2: Search driven by a well-represented probability distribution, steering the search towards the L-PS and the PS. (c) Cases 3: Search driven by an under-represented probability distribution, plausibly deviating the search towards non-Pareto optimal solutions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

An operator applied to the analytic expression of the Jacobian, (2) Numerical approach: An operator applied to a numerical approximation of the Jacobian, and (3) Iterative update: Updates the transformation tensor improving parallelism between the displacement of the non-dominated solutions in the objective space and improvement directions. The applied Operators are the Moore–Penrose pseudoinverse, the matrix transpose and the normalized pseudoinverse [2]. For the sake of completeness, the transformation tensor applied in this work is given as the pseudoinverse of a numerical approximation of the Jacobian evaluated at the solution \mathbf{x} through a quadratic regression, as:

$$\mathbf{H}(\mathbf{x}) = \hat{\mathbf{J}}(\mathbf{x})^\dagger \quad (9)$$

For a deeper review refer to Botello-Aceves et al. [3].

3. Multi-objective estimation of distribution algorithms

Overall, it can be argued that any instances of a MOEDA present different combinations of an estimation of a probability function (probability distribution) and a dominance relationship within the selection operator. The probability of sampling a new candidate solution from the estimated probability distribution is related to the distribution of the selected solutions within the search space, which in turn influences the search performance of the MOEDAs. Based on such an argument, the role of the probability distribution in a MOEDA is discussed. On behalf of the following discussion, let us assume two different scenarios involving the relative position between the probability distribution and the Pareto set (PS), and that either one of them may occur in the evolution process: (1) The PS is within the 90% level contour of the probability distribution allows us to deduce that the distribution will eventually fit the PS. (2) On the contrary, suppose that the Pareto set (PS) is not within the 90% level contour. Therefore, the probability distribution must be gradually displaced through the search space towards the PS until the probability distribution reaches the PS or a locally Pareto optimal set. The first assumption only occurs when the number of observations in the selected population is large enough in relation to the dimension of the observations so that a large portion of the search space is covered, otherwise the second scenario arises. Similarly, the second assumption converges towards the PF if the probability distribution is well-represented so that the captured traits of the MOP are applied to steer the search. Therefore, it is presumed that classic MOEDAs cannot adequately address high-dimensional problems (depending on the optimization problem, it may vary between a hundred, up to a thousand or more decision variables) and require special operators to steer the probability distribution as presented in Sastry et al. [16]. It is worth recognizing that in practice it is known that the poor performance of classic EDAs to address MOPs is shown long before having hundreds of variables, so it is necessary to add special operators to resist premature stagnation and misrepresentation issues.

Fig. 1 presents a visual representation of the aforementioned discussion, where the three steering performance cases of the probability distribution are shown. For all three cases, the 90% confidence intervals of the probability

distribution, represented by the gray shaded ellipse, the Pareto set (PS) in red, and a local Pareto optimal set (L-PS) in violet are shown in the search space. A selected solution set in green, from which the parameters of the probability distribution are estimated, presents the trend of the steering performance from the probability distribution with green arrows. The three steering performance cases are:

1. **Case 1:** In Fig. 1(a), the probability distribution envelops the PS and the L-PS. As the MOEDA evolves, the sampling of the reproductive process may generate candidate solutions near the PS, which in turn makes them suitable to be selected to compute the next probability distribution so that at each step, it encloses and intensifies its search, eventually fitting to the PS.
2. **Case 2:** In Fig. 1(b), a probability distribution estimated from a well-represented set, steers the search towards the PS, gradually advancing through the search space towards the PS until the probability distribution reaches the PS or the L-PS.
3. **Case 3:** In Fig. 1(c) a probability distribution estimated from an under-represented set, deviates from the search path that a well-represented set is guided through. Since the most representative traits of the MOP are not fully captured, the probability distribution lacks the information to properly steer the search. Therefore, the search may be steered towards non-Pareto optimal solutions, deteriorating the converge feature of the optimization method.

To address the issue of under-represented set in MOEDAs from Case 3, efficiently steering the solutions towards the PF in Case 2, while maintaining the intensification process in Case 1, the following section introduces and discusses the Directed Multi-Objective Estimation of Distribution Algorithm (D-MOEDA). The proposal estimates a multivariate Gaussian distribution for each solution using a neighborhood of the closest solutions. Once the probability distribution is estimated, it is oriented towards the search direction obtained from the IDM method based on aggregation functions (List 2), using a rank-one matrix orientation over the covariance matrix of the multivariate Gaussian distribution. The sampling of each probability distribution computes an offspring population. The Approximation set is updated by the solutions in the current Approximation set and the offspring population that minimize the aggregation function for each aggregation coefficients vector associated with a solution.

3.1. Directed multi-objective estimation of distribution algorithm

As previously mentioned, the proposed MOEDA drives the Approximation Set (AS) toward the PS by updating a probability distribution of each solution in the current AS through its neighborhood. It can be argued that the traits from the optimization problem captured by the probability distribution directly affect the performance of the algorithm. Additionally, the sampling of a probability distribution estimated over a set of non-dominated solutions may propose candidate solutions in the same non-dominated set, affecting the convergence towards the PF while improving the distribution and spread of the solutions in the AF, deriving into the third case of search performance for MOEDAs. Fig. 2 shows a visual representation of the non-dominated sampling. In the left figure, a current AS is represented as the collection of green dots in the search space, where all non-dominated solutions to the current AS are enclosed in the green region. Their corresponding images on the objective space are shown in green in the right figure. The gray shaded ellipse on the search space represents the estimated probability distribution, and a candidate solution sampled from the estimated probability distribution is denoted in red. Notice that as the probability distribution encloses and resembles a non-dominated region, its sampling tends to generate non-dominated candidate solutions, which may improve the distribution of the solutions, but undoubtedly affects the convergence towards the PF.

Therefore, to improve the convergence of solutions towards the PS, the orientation of the probability distribution must be incorporated. Therefore, this section is devoted to introducing a novel Multi-Objective Evolutionary Algorithm (MOEA) based on an Estimation of Distribution Algorithm (EDA) approach, orienting a multivariate Gaussian probability distribution, named the Directed Multi-Objective Estimation Distribution (D-MOEDA). Fig. 3 presents the central idea of the D-MOEDA. The search space presents a current AS in blue and the PS in red on the left side. Their corresponding images in the objective space are presented on the right side. For a solution $\mathbf{x} \in X$ in the current AS, the parameters of a Gaussian multivariate distribution are computed from a set of neighboring solutions, where the dotted ellipse represents a 95% probability mass of the probability distribution. Simultaneously, the corresponding search direction, d , from the solution is computed with the IDM method based on aggregation

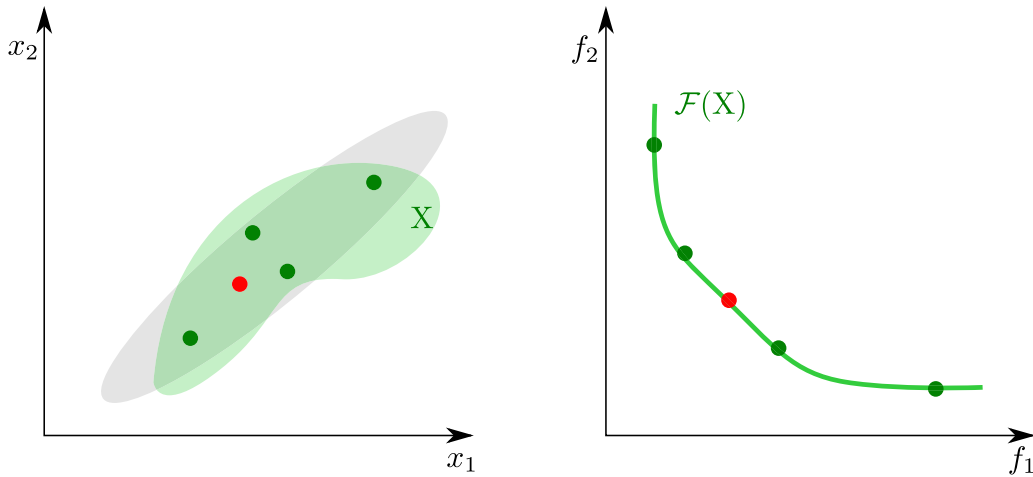


Fig. 2. Visual representation of the non-dominated sampling. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

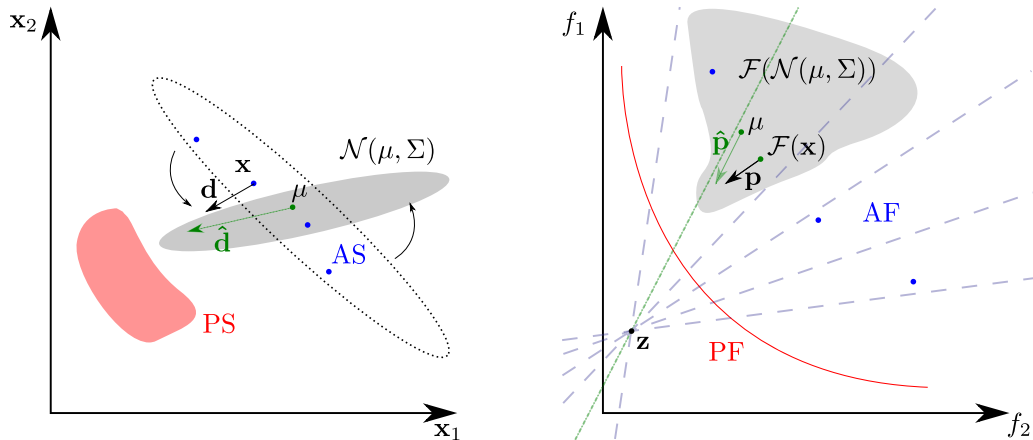


Fig. 3. Visual representation of the orientation of the multivariate Gaussian distribution on the D-MOEDA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

functions. Finally, the probability distribution is oriented towards the PS using the search direction, as presented by the gray ellipse. An adaptive orientation of the probability distribution presents an ideal behavior of the algorithm. As it will be further discussed, a full orientation of the probability distribution presents a fast convergence towards the PS. In contrast, a null orientation improves the distributions and spread the solutions within the current AS.

Algorithm 1 presents in detail the proposal. For a population size N and a set of well-distributed set of coefficient vectors $\lambda_i \quad \forall i = 1, \dots, N$, at the generation $g = 0$. An initial population \mathcal{X}_g of size N is randomly generated from a uniform distribution, while the space of an offspring population is instantiated, \mathcal{O}_g . Each solution is correlated to an aggregation coefficients vector λ_i , and a neighborhood N_i , compound by the collection of solutions with the closest $|N_i|$ aggregation coefficients vectors. Until a stop condition is met, for an adaptive orientation of the probability distribution a *orientation rate* c is computed. For each solution $\mathbf{x}_i \in \mathbb{R}^n$ in the current AS, the improvement direction $\mathbf{p} \in \mathbb{R}^m$ is computed on Line 6. Then, Line 8 presents the computation of the search direction $\mathbf{d} \in \mathbb{R}^n$ by mapping the improvement direction \mathbf{p} via a linear transformation tensor, computed on Line 7, given as the normalized pseudo-inverse of the Jacobian quadratic approximation. A set of weight coefficients concerning the fitness value of the aggregation function for the aggregation coefficients vector λ_i are computed as a mean to settle the importance of each neighboring solution within the probability distribution. Based on the weight coefficient, the mean vector, and the oriented covariance matrix of a Gaussian multivariate distribution are estimated using the neighboring solutions

Algorithm 1 Directed Multi-Objective Estimation of Distribution Algorithm

```

1: Set generation counter  $g = 0$ 
2: Initialize population, offspring and step size  $[X_g, O_g] = \text{Initialize}(N)$ 
3: while Stop criteria  $\neq$  True do
4:   Compute orientation rate  $c$ 
5:   for  $i = 1$  to  $N$  do
6:     Compute improvement direction  $\mathbf{p}$  as in Eq. (6)
7:     Compute transformation tensor  $\mathbf{H}$  as in Eq. (9)
8:     Compute search directions  $\mathbf{d}$  as in Eq. (3)
9:     Compute weight coefficient  $\omega_i$  as in Eq. (12)
10:    Compute mean vector  $\mu$  as in Eq. (10)
11:    Compute covariance matrix  $\Sigma$  as in Eq. (11)
12:    Compute offspring  $\mathbf{o}_i \sim \mathcal{N}(\mu, \Sigma)$ 
13:  end for
14:  Update Population  $X_{g+1} = \text{argmin } g(X_g, O_g | \lambda_i) \forall i = 1, \dots, N$ 
15:  Increase Counter  $g = g + 1$ 
16: end while
17: return  $X_g$ 

```

and the search directions on Lines 10 and 11. From the estimated probability distribution, an offspring solution $\mathbf{o}_i \in \mathbb{R}^n$ is created by a simulation of the probability distribution. Once the offspring population is updated, it is appended to the current population. A new population is computed as the set of solutions that minimize the aggregation function for each aggregation coefficients vector. The generation counter is updated, and once the stop criterion is met, the proposal outputs the current population.

Notice that the reorientation of the probability distribution is embedded in the computation of the mean vector and covariance matrix of a Gaussian multivariate distribution. For the sake of clarity, the following section presents the formulae for the estimation of the parameters of an oriented multivariate Gaussian distribution.

3.2. Directing the probability distribution

Line 12 from Algorithm 1 computes an offspring solution through the simulation over a local multivariate Gaussian distribution. The proposal estimates the parameters of the local probability distributions (Lines 10 and 11) for each solution from their corresponding subset of neighboring solutions. Nevertheless, the orientation of the local probability distribution towards a search direction is computed via a rank-one matrix at the center of the probability distributions. Therefore, this section delves into the reorientation of each local probability distribution.

Let $\mathbf{x} \in X$ be a solution within the current AS, associated to the aggregation coefficients vector λ and a neighborhood $N \subset X$. Let the solutions in the neighborhood be sorted based on the values of the aggregation function g with respect to the aggregation coefficients vector λ and a reference point \mathbf{z} , such that $g(\mathbf{x}_1) \leq \dots \leq g(\mathbf{x}_{|N|})$. Furthermore, let \mathbf{d} be the search direction of the IDM method on \mathbf{x} , computed through the quasi-Newton approach presented in List 2, a multivariate Gaussian distribution is directed towards the line of action of the aggregation coefficients vector by the estimating the parameter distribution as:

$$\mu = \sum_{j=1}^{|N|} \omega_j \mathbf{x}_j, \quad (10)$$

$$\Sigma = \frac{1-c}{1-\omega^\top \omega} \sum_{j=1}^{|N|} \omega_j (\mathbf{x}_j - \mu_{k+1})(\mathbf{x}_j - \mu_{k+1})^\top + c \cdot (\mathbf{d}_\mu \mathbf{d}_\mu^\top), \quad (11)$$

where $\mathbf{x}_j \in \mathcal{N}$ are the neighboring solutions, μ_{k+1} and Σ_{k+1} are the updated mean vector and covariance matrix of a multivariate Gaussian distribution, respectively. ω_j is a weight coefficient to each solution within the neighborhood, related to the distance to the line of action of the aggregated coefficients vector and the reference point through

the aggregation function. For all the experiments in this article the reference point is computed each generation as the current known minimum of each objective function. The weight coefficient centers the distribution on the values closest to the solution with the minimum aggregation function value while closing the distribution on the solutions with the minimum values to the aggregation function. Therefore, this work introduces an approach for the computation of the weight coefficients based on the value of the aggregation function of its correspondent solution and the solution in the neighborhood, as:

$$\omega_i = \frac{1}{|N| - 1} \left(1 - \frac{g(\mathbf{x}_i | \lambda, \mathbf{z})}{\sum_{j=1}^{|N|} g(\mathbf{x}_j | \lambda, \mathbf{z})} \right). \quad (12)$$

For the sake of clarity and completeness, let us delve into each parameter independently. The mean vector is computed as the weighted sum of the solutions within the neighborhood, as presented in Eq. (10). Notice that all solutions $\mathbf{x}_j \in N$ drag the probability distributions proportionally to the value of the aggregation function so that the nearest solution to the action line of the aggregation coefficients vector “pulls” harder than the farthest solution, driving the probability distribution towards the action line of the aggregation coefficient vector.

Additionally, the covariance matrix in Eq. (11) is computed as a weighted sum of two terms. The first is given as the covariance matrix with a weighted Gaussian multivariate distribution. The probability distribution is supplemented by each neighboring solution by its contribution, related to the weight coefficients. The second presents the rank-one orientation, given as the external product of a search direction computed at the centroid of the probability distribution \mathbf{d}_μ . Each search direction \mathbf{d}_i is computed over each neighboring solution $\mathbf{x}_i \in N$. Any search direction computed over the neighboring solutions does not ensure an adequate direction to drive the orientation of the probability distribution, since it is expected that the mean vector does not align with any of the solutions, $\nexists \mathbf{x}_i \in N | \mathbf{x}_i = \mu_{k+1}$. The most precise solution is through the evaluation of the mean vector and its corresponding computation of the search direction through the IDM procedure. The unnecessary evaluation is avoided by computing a proper search direction at the centroid of the probability distribution as the weighted sum of all the search directions in the neighborhood. In order to compute the search direction for each neighboring solution, a transformation tensor must be computed, increasing the computational cost. To reduce the computational cost, the weighted sum of the search directions is approximated as the transformation, computed over the mean of the probability distribution, of the weighted sum of the improvement directions, as presented in Eq. (13). The transformation tensor is computed as the pseudo-inverse of the Jacobian evaluated at the mean vector, approximated through a local quadratic regression of the MOP over the neighborhood, $\mathbf{H}(\mu) = \mathbf{J}_{QR}^\dagger(\mu)$.

$$\mathbf{d}_\mu = \sum_{i=1}^{|N|} \omega_i \mathbf{H}(\mathbf{x}_i) \mathbf{p}_i \approx \mathbf{H}(\mu) \sum_{i=1}^{|N|} \omega_i \mathbf{p}_i \quad (13)$$

Notice that the orientation rate c weights the importance of each term in the computation of the covariance matrix. The adaptation of this coefficient allows the algorithm to shift from exploration, by orienting the probability distribution towards the line of action of the aggregation coefficients vector accordingly to the search features of the aggregation function, to intensification, improving the distribution and spread of the solution within the AF by exploring the vicinity of the AS.

3.3. Computational complexity

The computational complexity of our proposal per iteration is $O(Nnm^2)$ for a population size N , n variables, and m objective functions, considering an iterative computation of the transformation tensor using a constant number of iterations k , and $O(Nnkm^2)$ in the case of using a variable number of iterations k . In contrast, the MOEA/D presents a computational complexity of $O(Nmt)$, for t target points, and the computational complexity of NSGA-II is $O(N^2 m)$. This last one presents, possibly (depending on the N and m values), the greatest computational complexity among these comparing algorithms [9]. On one hand, algorithms based on improvement directions such as the MOEA/D and this proposal present the advantage of avoiding a quadratic comparison among all individuals in the population, on the other hand, they add computational cost due to the computation of improvement directions or target displacements. In the particular case of this proposal, we require operations for computing the transformation tensor, applying the transformation to the improvement direction in the objective function space to compute the

search direction in the variable space and integrate the search direction in the covariance matrix to sample new candidate solutions. Thus our proposal presents a higher complexity than MOEA/D as a payoff of benefits such as the analytical concepts and tools that are introduced and could be used in future research, for instance, notice that the transformation tensor is a model to map the objective function space to the variable function space, even if it is not the subject of this article, the transformation tensor could be studied to infer the number of active objective functions, that is to say, whether the improvement direction depends on all or some objective functions, the importance of the variables to move forward in the improvement direction, to compute surrogate models considering that it represents local properties in the objective as well as the variable space, and, together with the covariance matrix, it could be used to quantify the spread or convergence of the current population, due to the rank reduction of the transformation tensor computed at points over the Pareto front.

4. Results

The previous section introduced the Directed Multi-Objective Estimation of Distribution Algorithm. It discussed the behavior of a multivariate Gaussian distribution on a MOEDA on MOP, where the probability distribution aligns with the current AS so that the convergence to the PF deteriorates whilst the distribution and spread improve. To improve the convergence toward the PF, the reorientation of the probability distribution using the search direction from the IDM framework is proposed. To improve the convergence at the exploration stage (orienting the probability distribution) and to improve the distribution and spread at the intensification stage, a *orientation rate* is introduced into the estimation of the parameter of the probability distribution. The value of the orientation rate influences the orientation of the probability distribution.

The following experiments are intended to visualize the performance of the oriented multivariate Gaussian distribution within the D-MOEDA and compare it with well-known and similar MOEAs. The first experiment studies the performance of the orientation rate applied to the probability distribution in order to establish the best-performance coefficient to the search features of the D-MOEDA. The second compares the search feature of the D-MOEDA with a set of state-of-the-art MOEAs showing the competitiveness of the D-MOEDA to deliver AFs that converge and adequately distribute the solutions over the PF.

4.1. Experiment 1

The experiment features the D-MOEDA with a set of different orientation rates and measures their performance on the UF functions of the CEC09 - competition problem [19]. The experiment weakens the evolutionary process of the D-MOEDA by reducing the number of evaluation functions and the population size to $FE = 50000$ evaluation functions to the UF[1-7] and $FE = 100000$ for the UF[8-10], and a population size $n_{pop} = 50$ for the UF[1-7] and $n_{pop} = 100$ for the UF[8-10]. The main objective of the weakening of the evolutionary process is to visualize the capabilities of the D-MOEDA for each orientation rate to reach the PF while improving the distribution and spread of each solution in the objective space before reaching the PF.

Table 1 compiles the mean and standard deviation of the Hypervolume ratio (HV_r) and Modified Inverted Generational Distance (IGD+) of 30 independent runs for the D-MOEDA with a set of different orientation rates $c = 0.0, 0.25, 0.5, 0.75, 1.0$. Notice that for $c = 0.0$, the probability distribution is not oriented, and the parameters are estimated only with the neighboring solutions. However, when $c = 1.0$, the probability distribution is fully oriented, and the estimation of the parameters only takes into account the search direction of the IDM without acknowledging the contribution of the neighboring solutions, such that its performance resembles an IDM method.

The presented results show that a mid-value for the orientation of the probability distribution has a reasonable balance between convergence, distribution, and spread of the output AFs. Notice that the D-MOEDA with an orientation rate value of 0.5 is the best, outperforming the other compared orientation rates in 7 out of 20 performance metrics. The second-best is the D-MOEDA with an orientation rate value of 0.75, outperforming other compared orientation rates in 6 out of 20 performance metrics. Lastly, the D-MOEDA with an orientation rate of 1.0 and 0.25 are respectively the thirty-best and fourth-best compared methods being the best in 5 out of 20 and 1 out of 20 performance metrics. Notice that for a non-oriented value, the D-MOEDA does not reach a suitable AF to outperform any other orientation rate, ensuring that the orientation of the probability distribution is an outstanding feature of the D-MOEDA to control its performance.

Table 1

HV_r and IGD+ mean and standard deviation values of 30 independent executions of the D-MOEDA with five orientation rates ($c = 0.0, 0.25, 0.5, 0.75$ and 1.0). The best mean value is displayed in bold for HV_r (bigger is better) and the IGD+ (lower is better).

		$c = 0$		$c = 0.25$		$c = 0.5$		$c = 0.75$		$c = 1.0$	
		mean	(\pm std dev)	mean	(\pm std dev)	mean	(\pm std dev)	mean	(\pm std dev)	mean	(\pm std dev)
HV _r	UF1	0.0649	(± 0.0648)	0.6145	(± 0.0837)	0.6214	(± 0.0629)	0.6181	(± 0.0685)	0.5745	(± 0.0682)
	UF2	0.7326	(± 0.0391)	0.8805	(± 0.0184)	0.8825	(± 0.0142)	0.8827	(± 0.0125)	0.8483	(± 0.0123)
	UF3	0.0887	(± 0.0240)	0.4479	(± 0.1068)	0.4671	(± 0.0693)	0.4221	(± 0.0573)	0.3880	(± 0.0598)
	UF4	0.5867	(± 0.0247)	0.5924	(± 0.0382)	0.5914	(± 0.0378)	0.6009	(± 0.0285)	0.6028	(± 0.0354)
	UF5	0.0000	(± 0.0000)	0.0000	(± 0.0000)	0.0000	(± 0.0000)	0.0000	(± 0.0000)	0.0000	(± 0.0000)
	UF6	0.0000	(± 0.0000)	0.0000	(± 0.0000)	0.0000	(± 0.0001)	0.0000	(± 0.0000)	0.0004	(± 0.0024)
	UF7	0.0193	(± 0.0369)	0.6727	(± 0.1571)	0.7051	(± 0.1594)	0.6952	(± 0.1222)	0.6417	(± 0.1535)
	UF8	0.1839	(± 0.0423)	0.3618	(± 0.0159)	0.3680	(± 0.0160)	0.3714	(± 0.0154)	0.3257	(± 0.0629)
	UF9	0.1874	(± 0.0573)	0.3086	(± 0.0929)	0.3217	(± 0.0812)	0.2919	(± 0.0494)	0.2503	(± 0.0731)
	UF10	0.0000	(± 0.0000)	0.0000	(± 0.0000)	0.0000	(± 0.0000)	0.0000	(± 0.0001)	0.0035	(± 0.0112)
$\Delta\mu$		0.1860	(± 0.0292)	0.3883	(± 0.0517)	0.3960	(± 0.0451)	0.3888	(± 0.0363)	0.3637	(± 0.0485)
IGD+	UF1	0.6573	(± 0.1302)	0.1816	(± 0.0510)	0.1777	(± 0.0399)	0.1779	(± 0.0435)	0.1954	(± 0.0410)
	UF2	0.1356	(± 0.0198)	0.0665	(± 0.0142)	0.0641	(± 0.0089)	0.0631	(± 0.0095)	0.0865	(± 0.0091)
	UF3	0.5802	(± 0.0308)	0.3051	(± 0.0795)	0.2957	(± 0.0426)	0.3144	(± 0.0366)	0.3428	(± 0.0472)
	UF4	0.1014	(± 0.0070)	0.0986	(± 0.0120)	0.0973	(± 0.0116)	0.0932	(± 0.0086)	0.0923	(± 0.0111)
	UF5	3.2034	(± 0.5154)	2.1207	(± 0.5093)	2.2845	(± 0.5816)	2.5839	(± 0.6148)	2.4721	(± 0.4714)
	UF6	3.0348	(± 0.5079)	1.2637	(± 0.7108)	1.2399	(± 0.6164)	1.2762	(± 0.9561)	1.0634	(± 0.3732)
	UF7	0.6990	(± 0.1439)	0.1258	(± 0.0728)	0.1410	(± 0.2369)	0.1093	(± 0.0586)	0.1339	(± 0.0853)
	UF8	0.3513	(± 0.0507)	0.2009	(± 0.0114)	0.1949	(± 0.0081)	0.1932	(± 0.0068)	0.2213	(± 0.1106)
	UF9	0.4246	(± 0.1255)	0.3341	(± 0.1149)	0.3091	(± 0.0427)	0.3191	(± 0.0281)	0.3595	(± 0.1069)
	UF10	1.7826	(± 0.2641)	0.6529	(± 0.0636)	0.7685	(± 0.7712)	0.6417	(± 0.0721)	1.1247	(± 1.1837)
$\Delta\mu$		1.0970	(± 0.1795)	0.5350	(± 0.1639)	0.5573	(± 0.2360)	0.5772	(± 0.1835)	0.6092	(± 0.2439)

The last rows of each section of Table 1 present the average value of the means and standard deviations of all problems. Even though $c = 0.5$ is the best for 5 of 10 problems in HV_r and for 3 in IGD+, $c = 0.25$ is more consistent in the IGD+ metric for all problems, because the averages of means and standard deviations are lower, in addition, recall that the IGD+ metric is Pareto compliant while the HV_r is not. Thus, based on these observations we recommend to use $c = 0.25$, and it is used in the second experiment, because even if it is not the best for some particular problems, it is the most competitive for a wider variety of problems, hence this encourages its use where not a priori information about the adequate value is known.

Fig. 4 presents the distribution of the Hypervolume ratio (HV_r) and the Modified Inverted Generational Distance (IGD+) for the 30 independent runs and orientation rates in each benchmark problem. Similar to Table 1, Fig. 4 compared the five approaches using the non-parametric Wilcoxon rank-sum test, remarking in green the fittest approach, that is the biggest median for the HV_r and the lowest median for the IGD+. Results support the previous assertions, where the orientation rate of 0.5 and 0.75 presents the fittest HV_r for 4 out of 10 and 3 out of 10 benchmark problems. It is worth noticing that in the UF benchmark [4, 5, 10], the output AFs of all approaches do not reach the reference point needed by the Hypervolume metric. Nevertheless, the orientation rate of 0.25 and 0.5 presents the fittest IGD+ performance metric, being the best in 4 out 10 and 3 out of 10 benchmark problems, respectively. These results support the statement of the performance of the oriented probability distribution, where an oriented probability distribution converges towards the PF whiles the non-oriented probability distribution scatters the solutions evenly over the AF.

To visualize the performance of the oriented probability distribution statement previously discussed, Figs. 5 and 6 present the AF of each approach with the best Hypervolume ratio (HV_r) and best modified inverted generational distance (IGD+) for UF[1, 2, 5, 6, 9, 10] problems. A common pattern arises from all the figures, where the most oriented probability distribution approximates, without fully reaching, a heavily crowded AF to the PF. In contrast, the non-oriented probability distribution scatters all the solutions through the AF with even approaching the PF. These patterns perfectly describe the performance of the orientation applied to the probability distribution, where a completely oriented distribution accelerates the convergence towards the PF (Case 2 steering performance), affecting

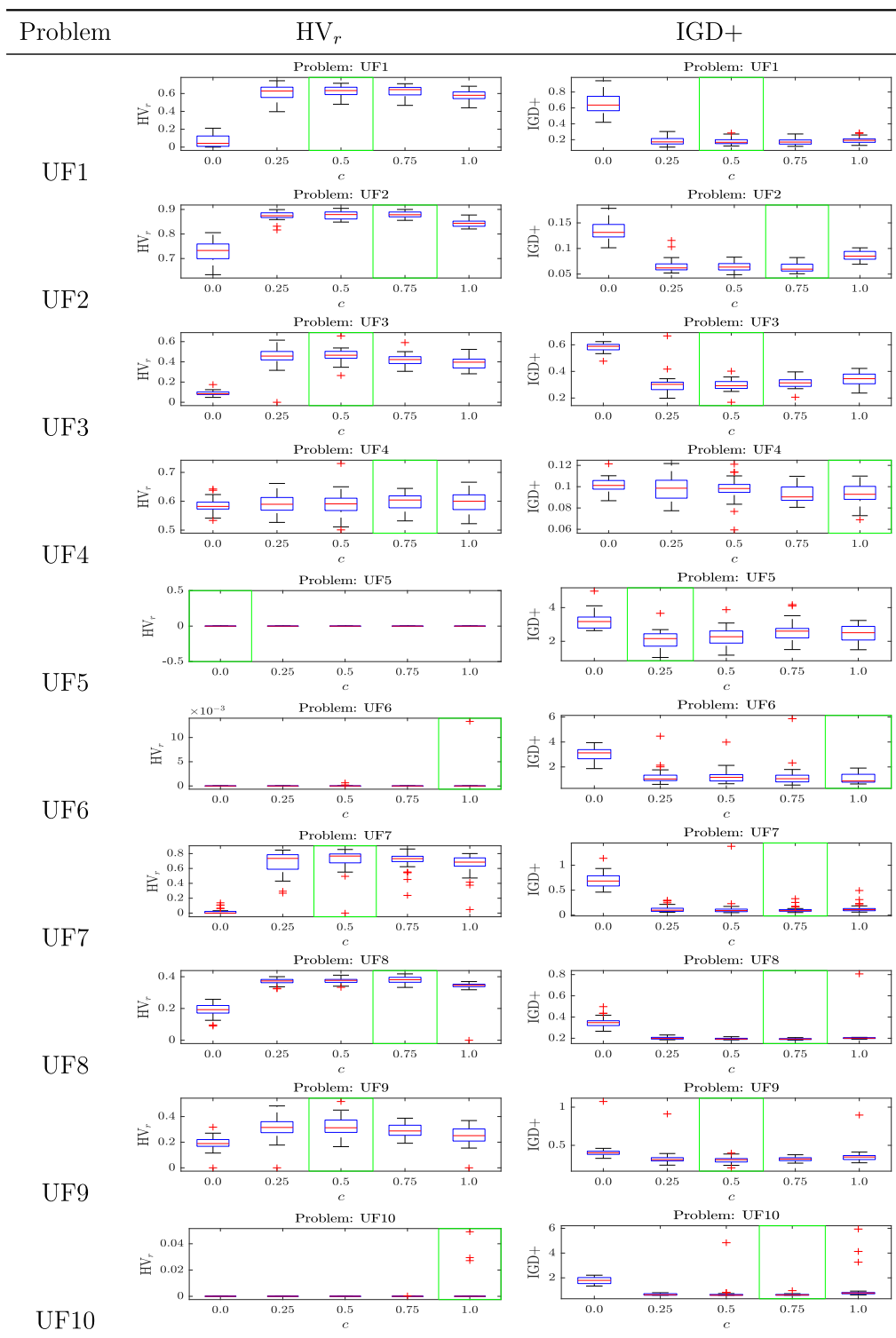


Fig. 4. HV_r and IGD+ distribution the of 30 independent executions of the D-MOEDA for five orientation rates ($c = 0.0, 0.25, 0.5, 0.75$ and 1.0), where a bigger HV_r is better and a lower IGD+ is better. The most outperforming coefficient is remarked in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

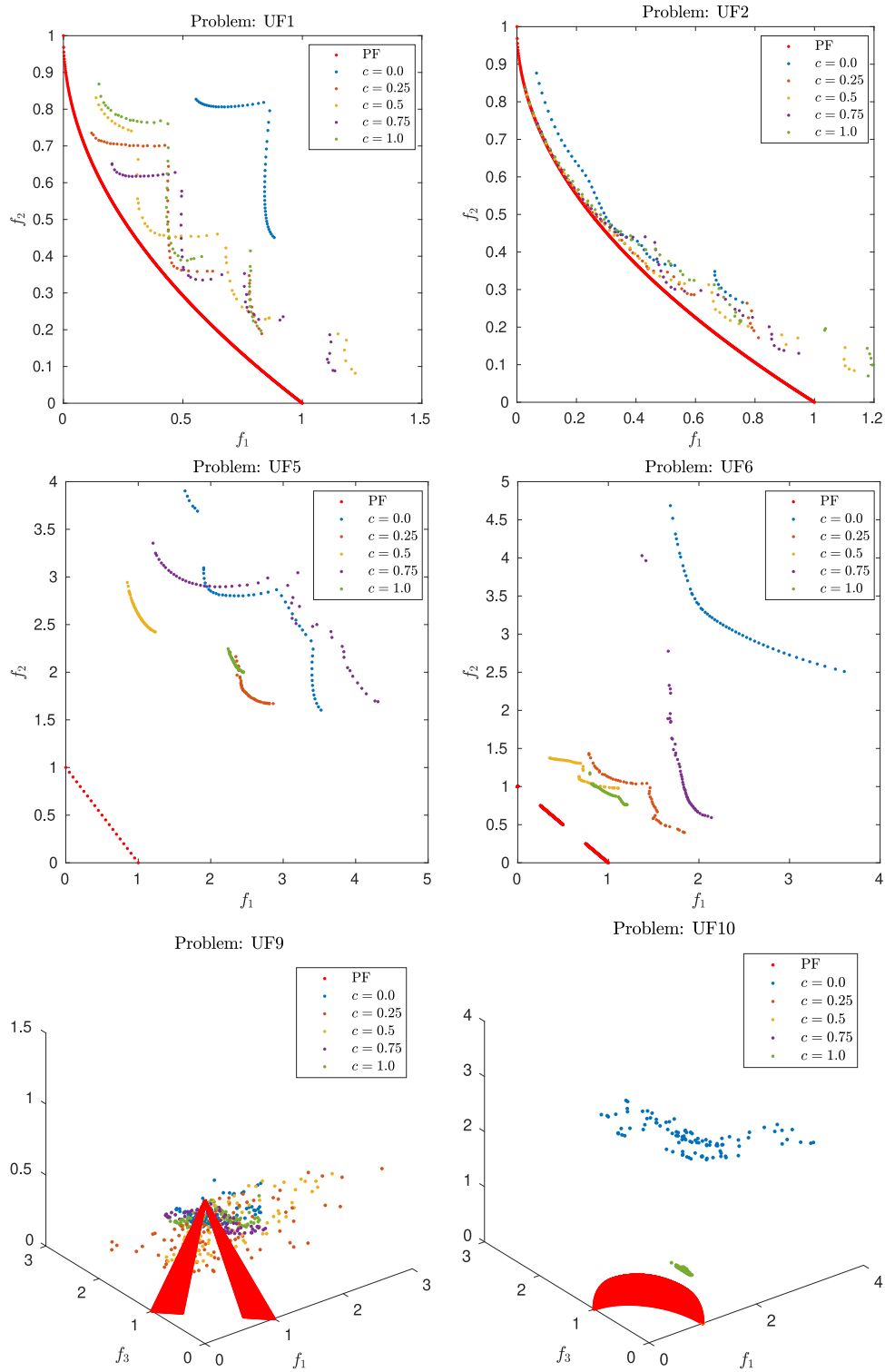


Fig. 5. Best HV_r approximation fronts for the five orientation rates for the UF[1, 2, 5, 6, 9, 10] benchmark problems.

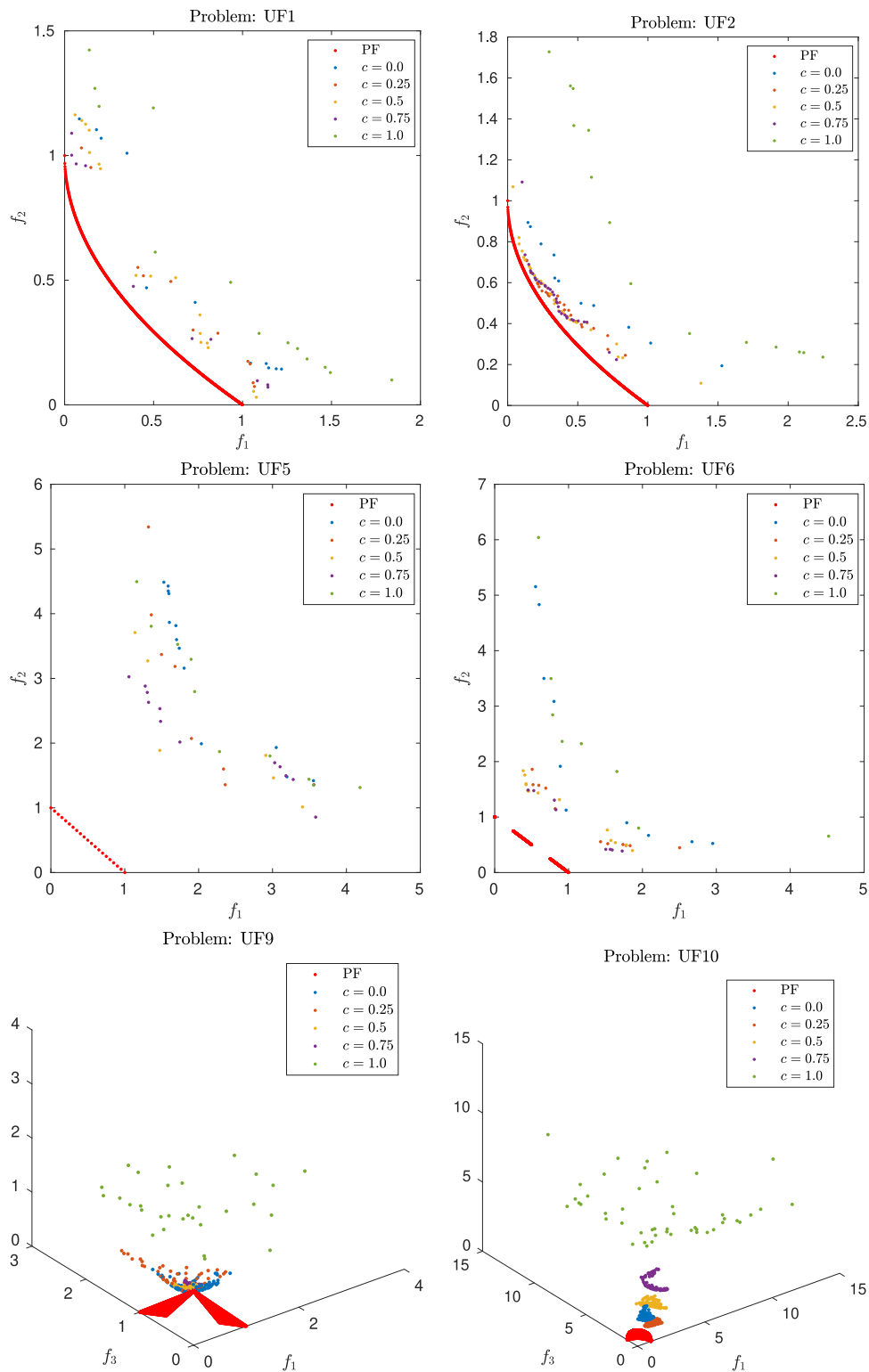


Fig. 6. Best IGD+ approximation fronts for the five orientation rates for the UF[1, 2, 5, 6, 9, 10] benchmark problems.

Table 2

Mean and standard deviation of the hypervolume (HV) of 30 independent executions of the NSGA-II, NSGA-III, MOEA/D, MO-RV-GOMEA and the D-MOEDA with orientation rate $c = 0.25$. The best mean value is displayed in bold for HV (bigger is better).

		NSGA-II		NSGA-III		MOEA/D		D-MOEDA		MO-RV-GOMEA	
		mean	(\pm std dev)	mean	(\pm std dev)	mean	(\pm std dev)	mean	(\pm std dev)	mean	(\pm std dev)
HV	ZDT1	0.8891	(± 0.2162)	0.0017	(± 0.0091)	3.4827	(± 0.0762)	3.3852	(± 0.1466)	3.4563	(± 0.0813)
	ZDT2	0.0000	(± 0.0000)	0.0000	(± 0.0000)	2.2342	(± 0.4229)	3.1616	(± 0.2260)	3.1275	(± 0.0777)
	ZDT3	1.7083	(± 0.2503)	0.0665	(± 0.1307)	4.5834	(± 0.1431)	4.0536	(± 1.1576)	4.3143	(± 0.2287)
	ZDT4	0.0000	(± 0.0000)	0.0000	(± 0.0000)	1.9839	(± 0.5894)	0.0000	(± 0.0000)	0.0101	(± 0.0384)
	ZDT6	0.0000	(± 0.0000)	0.0000	(± 0.0000)	2.4931	(± 0.1386)	2.7170	(± 0.0708)	3.0136	(± 0.0148)
	genMED	1.7455	(± 0.2173)	0.0350	(± 0.0959)	3.6961	(± 0.0505)	3.4101	(± 0.6510)	3.7904	(± 0.0100)
IGD+	ZDT1	1.1631	(± 0.1511)	2.3765	(± 0.1951)	0.032	(± 0.0064)	0.0959	(± 0.0403)	0.9457	(± 0.1614)
	ZDT2	2.0143	(± 0.2213)	3.6466	(± 0.2568)	0.2155	(± 0.1037)	0.0883	(± 0.027)	0.6731	(± 0.6304)
	ZDT3	0.9453	(± 0.1632)	2.1233	(± 0.2065)	0.0364	(± 0.0111)	0.1291	(± 0.0395)	0.9883	(± 0.1675)
	ZDT4	58.5145	(± 9.5746)	93.2329	(± 11.7614)	0.0687	(± 0.0534)	3.0553	(± 0.8458)	29.1233	(± 14.2527)
	ZDT6	4.8864	(± 0.5914)	6.9748	(± 0.2219)	0.0829	(± 0.0226)	0.0171	(± 0.006)	0.0493	(± 0.028)
	genMED	–	–	–	–	–	–	–	–	–	–

the distribution and extension of the solutions in the AF. In contrast, a probability distribution on the neighboring solutions improves the distribution and spread of the solutions within the AF, decelerating the convergence towards the PF (Case 1 steering performance). Consequently, the adaptation of the orientation rate to adapt the performance of the D-MOEDA based on the evolution of the AF is an attractive line of research. Furthermore, notice that the performance of the D-MOEDA due to the orientation rate is closely related to the features of the MOP. That is, if the objective space is large enough, the convergence must be prioritized (Case 2 steering performance), while a narrow objective space must prioritize the distribution and spread of the solutions (Case 1 steering performance). Therefore, the adaptation of the orientation rate based on the characteristics of the MOP and the evolution of the AF must be further studied.

4.2. Experiment 2

This experiment compares the D-MOEDA with the fittest orientation rate with a set of well-known, the Fast Elitist Non-Dominated Sorting Genetic Algorithm (NSGA-II) [6], the Reference-Point-Based Non-Dominated Sorting Genetic Algorithm (NSGA-III) [5], the Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D) [17] and a MOEA that similarly replaces the reproductive process by applying a sampling method from a probabilistic model, the Multi-Objective Real-Valued Gene-Pool Optimal Mixing Evolutionary Algorithm (MO-RV-GOMEA) [4].

The experiment features the ZDT[1-4,6] [21] and the genMED [4] benchmark problems. For all the benchmark problems the stop criteria are set at 5000 function evaluations, and a population size of $n_{pop} = 50$ individuals. The reproductive parameters used by the NSGA-II, NSGA-III and the MOEA/D are: crossover coefficient $\eta_{cross} = 5$, mutation coefficient $\eta_{mut} = 5$, crossover probability $p_{cross} = 0.9$, mutation probability $p_{mut} = 0.1$. The MOEA/D uses the Standard Tchebycheff as the aggregation function to apply a dominance relationship, while the D-MOEDA uses the Parabolic Scalarizing Function as a dominance relationship and to direct the probability distribution. For a fair comparison, the MO-RV-GOMEA is used as a black-box method, regulating additional information from the objective functions to limit further function evaluation needed by the gray-box method.

Table 2 presents the hypervolume (HV) and the modified inverted generational distance (IGD+) from the AF of the 30 independent executions for each MOEA. Notice that the IGD+ of genMED problem is not presented since its PF is unknown, and at the best of our knowledge, no reference front is reported in the literature. The algorithm with the best performance for each benchmark problem is highlighted in bold. Notice that the MOEA/D and both MOEDAs outperform the Pareto dominance-based algorithms. Overall, the MOEA/D performs better than the other MOEAs, outperforming all MOEAs in 3 out of 6 benchmark problems. The MO-RV-GOMEA is the second-best optimization method, outperforming all MOEAs in 2 out of 6 benchmark problems. Lastly, the D-MOEDA outperforms all MOEAs in 1 out of 6 benchmark problems. Nevertheless, notice that the D-MOEDA shows competitive results, presenting hypervolume values near the fittest MOEAs.

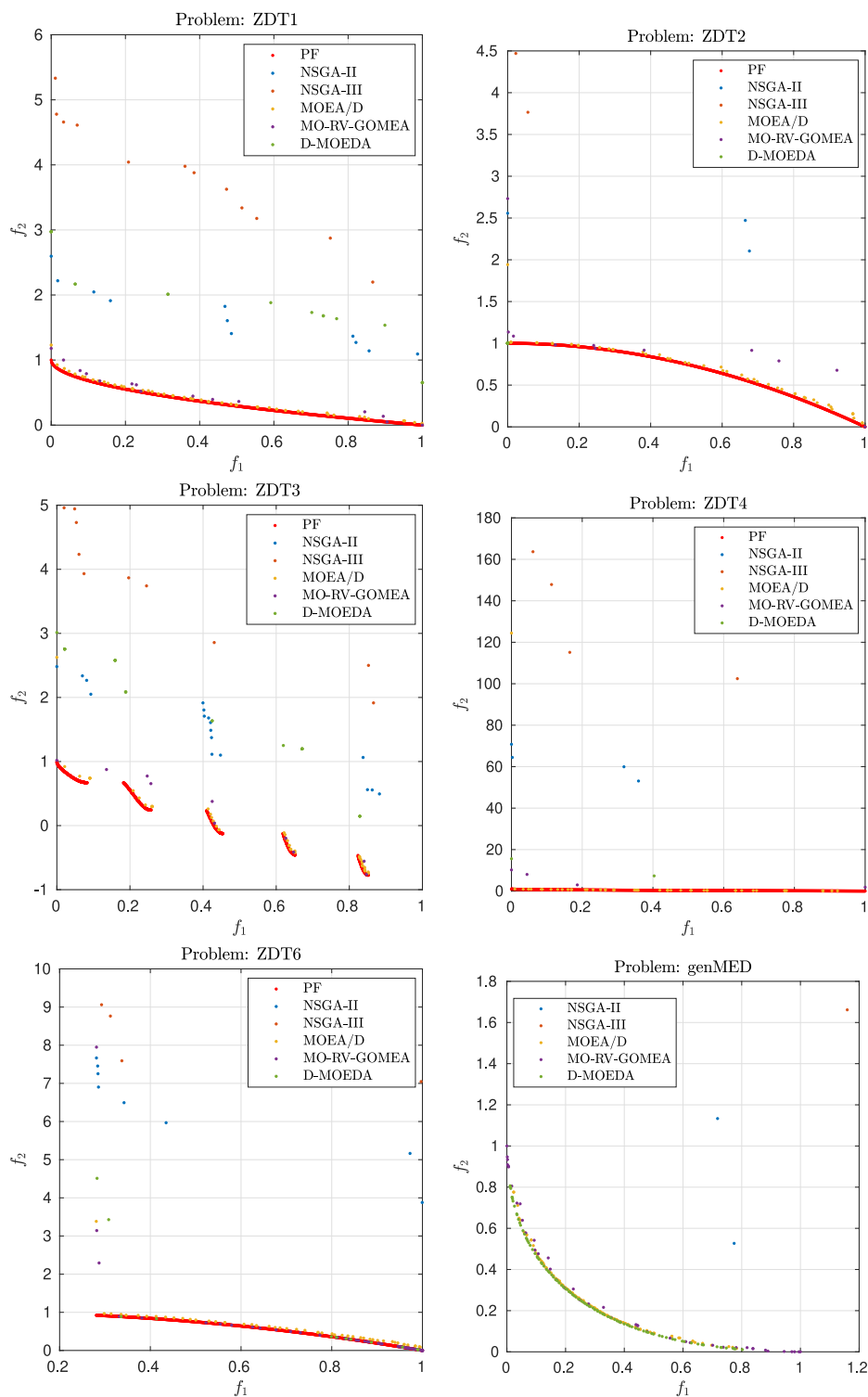


Fig. 7. Best HV approximation fronts for the five MOEAs for the ZDT[1-4,6] and genMED benchmark problems. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 7 presents the PF in red and the AF for each MOEA with the best hypervolume (HV) for ZDT[1-4,6] and genMED benchmark problems. The PF of the genMED problem is not presented due to the same complication previously mentioned. Fig. 7 supports the discussion mentioned lines above. Notice that the MO-RV-GOMEA approaches the PF of each problem in pools of solutions, mainly due to its clustering operators. As for the D-MOEDA, it reaches extreme solutions in the PF, while distributing other solutions through the objective space, with its best performance in the genMED problem.

5. Final remarks

This article discusses the performance of Estimation of Distribution Algorithms (EDAs) in Multi-Objective Optimization Problems (MOPs). The attributes and issues of steering performance in EDAs are examined. To exploit the favorable properties while alleviating their main issues of EDAs, this work introduces the Directed-Multi-Objective Estimation of Distribution Algorithm (D-MOEDA). The proposal intends to orient a local probability distribution applying the search direction from the IDM framework, in order to improve the convergence features of the optimization method, while preserving the distribution and spread features of the MOEDAs. The proposal aims to balance the convergence towards the PF while spreading the solutions in the AF through an orientation rate. To characterize the performance of the proposal, the orientation rate is studied through a first experiment, where a set of well-known benchmark problems are solved by the D-MOEDA with different orientation rates. Results show that the coefficient controls the search features of the proposal by either improving the convergence or the distribution of the solutions. Encouraged by the results of the first experiment, the proposal presents a competitive performance, compared to three well-known MOEAs and a state-of-the-art MOEDA. Results motivate the studies of adaptively direct the stochastic search by means of the search direction computed from the IDM framework in order to promote the convergence to the PF and the distribution of the solutions in the AF at different stages of the evolutionary process.

CRedit authorship contribution statement

Salvador Botello-Aceves: Conceptualization, Methodology, Software, Writing – original draft. **Arturo Hernandez-Aguirre:** Writing – reviewing & editing, Conceptualization, Methodology, Supervision. **S. Ivvan Valdez:** Writing – reviewing & editing, Conceptualization, Methodology, Supervision.

Acknowledgments

The authors of this paper acknowledge support from the Consejo Nacional de Ciencia y Tecnología of México (CONACyT) through the “Ciencia Básica” project No. 285599. S. Ivvan Valdez is supported by Cátedra CONACyT No. 7795.

References

- [1] S. Botello-Aceves, A. Hernandez-Aguirre, S.I. Valdez, Computation of the improvement directions of the pareto front and its application to moeas, in: Proceedings of the 2020 Genetic and Evolutionary Computation Conference, GECCO '20, Association for Computing Machinery, New York, NY, USA, 2020a, pp. 480–488, <http://dx.doi.org/10.1145/3377930.3390165>.
- [2] S. Botello-Aceves, S.I. Valdez, A. Hernández-Aguirre, The improvement direction mapping method, in: L. Martínez-Villaseñor, O. Herrera-Alcántara (Eds.), *Advances in Soft Computing*, Springer International Publishing, Cham, 2020b, pp. 264–283.
- [3] S. Botello-Aceves, S.I. Valdez, A. Hernandez-Aguirre, A broyden-based algorithm for multi-objective local-search optimization, *Inform. Sci.* 594 (2022) 264–285, <http://dx.doi.org/10.1016/j.ins.2022.02.017>, <https://www.sciencedirect.com/science/article/pii/S0020025522001505>.
- [4] A. Bouter, N.H. Luong, C. Witteveen, T. Alderliesten, P.A.N. Bosman, The multi-objective real-valued gene-pool optimal mixing evolutionary algorithm, in: Proceedings of the Genetic and Evolutionary Computation Conference, GECCO '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 537–544, <http://dx.doi.org/10.1145/3071178.3071274>.
- [5] K. Deb, H. Jain, An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints, *IEEE Trans. Evol. Comput.* 18 (2014) 577–601, <http://dx.doi.org/10.1109/TEVC.2013.2281535>.
- [6] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2002) 182–197, <http://dx.doi.org/10.1109/4235.996017>.
- [7] M.P. Hansen, A. Jaszkiewicz, Evaluating the Quality of Approximations To the Non-Dominated Set, IMM, Department of Mathematical Modelling, Technical University of Denmark, 1994.
- [8] G. Harik, et al., Linkage learning via probabilistic modeling in the ecga, *IlligAL report 99010*, 1999.

- [9] W. Huang, Y. Zhang, L. Li, Survey on multi-objective evolutionary algorithms, *J. Phys.: Conf. Series* 1288 (2019) 012057.
- [10] M. Jiang, L. Qiu, Z. Huang, G.G. Yen, Dynamic multi-objective estimation of distribution algorithm based on domain adaptation and nonparametric estimation, *Inform. Sci.* 435 (2018) 203–223, <http://dx.doi.org/10.1016/j.ins.2017.12.058>, <https://www.sciencedirect.com/science/article/pii/S0020025517311775>.
- [11] H. Karshenas, R. Santana, C. Bielza, P. Larrañaga, Multiobjective estimation of distribution algorithm based on joint modeling of objectives and variables, *IEEE Trans. Evol. Comput.* 18 (2014) 519–542, <http://dx.doi.org/10.1109/TEVC.2013.2281524>.
- [12] N. Khan, Bayesian optimization algorithms for multiobjective and hierarchically difficult problems, (Ph.D. thesis), University of Illinois at Urbana-Champaign Urbana, IL, 2003.
- [13] N. Khan, D.E. Goldberg, M. Pelikan, Multi-objective bayesian optimization algorithm, in: *Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computation*, Citeseer, 2002, p. 684.
- [14] M. Laumanns, J. Ocenasek, Bayesian optimization algorithms for multi-objective optimization, in: J.J.M. Guervós, P. Adamidis, H.-G. Beyer, H.-P. Schwefel, J.-L. Fernández-Villacañás (Eds.), *Parallel Problem Solving from Nature – PPSN VII*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2002, pp. 298–307.
- [15] M. Pelikan, K. Sastry, D.E. Goldberg, Multiobjective estimation of distribution algorithms, in: *Scalable Optimization Via Probabilistic Modeling*, Springer, 2006, pp. 223–248.
- [16] K. Sastry, D. Goldberg, M. Pelikan, Limits of scalability of multiobjective estimation of distribution algorithms, in: *2005 IEEE Congress on Evolutionary Computation*, Vol. 3, 2005, pp. 2217–2224, <http://dx.doi.org/10.1109/CEC.2005.1554970>.
- [17] Q. Zhang, H. Li, MOEA/D: A multiobjective evolutionary algorithm based on decomposition, *IEEE Trans. Evol. Comput.* 11 (2007) 712–731, <http://dx.doi.org/10.1109/TEVC.2007.892759>.
- [18] Q. Zhang, A. Zhou, Y. Jin, Rm-meda: A regularity model-based multiobjective estimation of distribution algorithm, *IEEE Trans. Evol. Comput.* 12 (2008) 41–63, <http://dx.doi.org/10.1109/TEVC.2007.894202>.
- [19] Q. Zhang, A. Zhou, S. Zhao, P.N. Suganthan, W. Liu, S. Tiwari, Multiobjective optimization test instances for the cec 2009 special session and competition, Working Report CES-887, School of Computer Science and Electrical Engineering, University of Essex, 2008.
- [20] X. Zhong, W. Li, A decision-tree-based multi-objective estimation of distribution algorithm, in: *2007 International Conference on Computational Intelligence and Security (CIS 2007)*, 2007, pp. 114–118, <http://dx.doi.org/10.1109/CIS.2007.136>.
- [21] E. Zitzler, K. Deb, L. Thiele, Comparison of multiobjective evolutionary algorithms: Empirical results, *Evol. Comput.* 8 (2000) 173–195, <http://dx.doi.org/10.1162/106365600568202>.
- [22] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: Improving the strength pareto evolutionary algorithm, TIK-report 103, 2001.