Using the Averaged Hausdorff Distance as a Performance Measure in Evolutionary Multiobjective Optimization

Oliver Schütze, Xavier Esquivel, Adriana Lara, and Carlos A. Coello Coello, *Fellow, IEEE*

***Abstract*—The Hausdorff distance d_H is a widely used tool to measure the distance between different objects in several research fields. Possible reasons for this might be that it is a natural extension of the well-known and intuitive distance between points and/or the fact that d_H defines in certain cases a metric in the mathematical sense. In evolutionary multiobjective optimization (EMO) the task is typically to compute the entire solution set—the so-called Pareto set—respectively its image, the Pareto front. Hence, d_H should, at least at first sight, be a natural choice to measure the performance of the outcome set in particular since it is related to the terms *spread* and *convergence* as used in EMO literature. However, so far, d_H does not find the general approval in the EMO community. The main reason for this is that d_H penalizes single outliers of the candidate set which does not comply with the use of stochastic search algorithms such as evolutionary strategies. In this paper, we define a new performance indicator, Δ_p, which can be viewed as an "averaged Hausdorff distance" between the outcome set and the Pareto front and which is composed of (slight modifications of) the well-known indicators generational distance (GD) and inverted generational distance (IGD). We will discuss theoretical properties of Δ_p (as well as for GD and IGD) such as the metric properties and the compliance with state-of-the-art multiobjective evolutionary algorithms (MOEAs), and will further on demonstrate by empirical results the potential of Δ_p as a new performance indicator for the evaluation of MOEAs.**

***Index Terms*—Averaged Hausdorff distance, generational distance, inverted generational distance, multiobjective optimization, performance indicator.**

I. Introduction

IN MANY applications, it is desired to optimize several conflicting objectives at once leading to a multiobjective optimization problem (MOP). Typically, the solution set of a MOP—the Pareto set—is not given by a single point but forms a $(k-1)$-dimensional object, where k is the number of objectives involved in the MOP. Hence, a natural question that arises is how to measure the performance of an (evolutionary) algorithm aiming for the approximation of the *entire* Pareto set and respectively its image, the Pareto front. One way to do this is to measure the distance of the outcome set of the algorithm to the set of interest.

One such distance function is the Hausdorff distance d_H [23], which is already established in several research fields such as image matching (e.g., [7], [24], [50]), the approximation of manifolds in dynamical systems [2], [12], [35], or in fractal geometry [15], among others. One major advantage of d_H is that it defines a metric in the mathematical sense on the set of compact subsets of $\mathbb{R}^n$. The problem at hand (i.e., to measure the distance between two sets) is certainly abstract, and no ultimate fairness can be expected ("How can *one* value give all the required information about the relation of a candidate set consisting of, say, 100 elements to a discretized Pareto set/front consisting of 300 elements?"). One important property of a metric is that the triangle inequality is satisfied which says that given the sets A, B, and C, the distance from A to C via B is at least as great as from A to C directly. If indicators are used that do not have the properties of a metric, unwanted effects can occur (e.g., greedy methods based on such indicators may be guided into wrong directions).

Another advantage of the Hausdorff distance, which is more specific to the optimization problem at hand, is that a low value $d_H(\mathcal{O}, \mathcal{F})$ of the distance between the image of the outcome set $\mathcal{O}$ and the Pareto front $\mathcal{F}$ gives a clear idea of the approximation quality of $\mathcal{O}$ (the same argument holds for Pareto *set* approximations). Since $d_H(\mathcal{O}, \mathcal{F})$ measures the distance of each set to the other one, the decision maker (DM) gets the information about the approximation quality in terms of the distance from $\mathcal{O}$ to $\mathcal{F}$ [which is typically termed as *convergence* in the evolutionary multiobjective optimization (EMO) literature] as well as the distance from $\mathcal{F}$ to $\mathcal{O}$ (which is closely related to what is termed as *spread* in EMO literature in terms of the maximal gap in the approximation). If, for instance, the DM is willing to accept an a priori determined deterioration $\delta > 0$ (resulting by lack of convergence or by the discretization of the Pareto set), every outcome $\mathcal{O}$ with $d_H(\mathcal{O}, \mathcal{F}) < \delta$ is "good enough" for his/her application (see also the results in Section V-C).

Manuscript received October 8, 2010; revised February 25, 2011; accepted June 18, 2011. Date of publication February 10, 2012; date of current version July 26, 2012. The work of O. Schütze was supported by DAAD, under Project 146776. The work of X. Esquivel and A. Lara was supported by CONACyT through a scholarship to pursue graduate studies at the Department of Computer Science, CINVESTAV-IPN. The work of C. A. Coello Coello was supported by CONACyT, under Project 103570.

O. Schütze and A. Lara are with the Department of Computer Science, CINVESTAV-IPN, Mexico City, DF 07360, Mexico (e-mail: schuetze@cs.cinvestav.mx; alara@computacion.cs.cinvestav.mx).

X. Esquivel is with the Oracle Mexico Development Center, Zapopan Jalisco 45110, Mexico (e-mail: xavies@computacion.cs.cinvestav.mx).

C. A. Coello Coello is with the Department of Computer Science, CINVESTAV-IPN, Mexico City, DF 07360, Mexico, and also with Unité Mixte Internationale LAFMIA 3175 CNRS, CINVESTAV-IPN, Mexico City 07360, Mexico (e-mail: ccoello@cs.cinvestav.mx).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TEVC.2011.2161872

On the other hand, the Hausdorff distance is yet scarcely used by the EMO community except for rather theoretical works [13], [40], [46], [47]. One mayor reason for this is probably that d_H penalizes the largest outlier of the candidate set which makes "good" approximations that contain at least one outlier to appear "bad." Hence, a large value of $d_H(\mathcal{O}, \mathcal{F})$ can indicate both that $\mathcal{O}$ is indeed a bad approximation of $\mathcal{F}$ and that $\mathcal{O}$ is "good" but contains at least one outlier.

One possible remedy is to average the distances of the elements of the sets leading to an "averaged Hausdorff distance." However, one has to be aware of the fact that such an averaging of the distances leads to violations of the triangle inequality, and hence, to a loss of the metric property.

To motivate the need for a fair incorporation of outliers for distance assignments in the context of evolutionary multiobjective optimization, we consider the following three examples: the first (academic) example shows that once points near to weakly optimal solutions that are far from the Pareto set are generated, it might not be easy to eliminate them from the archive/population (such points are also called dominance resistant points in the EMO literature, see [20], [31]): consider the MOP

$$F : [0, 1]^n \to \mathbb{R}^k$$
$$F(x) = \begin{pmatrix} x_1 \\ \mathrm{g(x)} \end{pmatrix} \tag{1}$$

where $g : [0, 1]^n \to \mathbb{R}^{k-1}$ (i.e., the first objective is given by x_1 as in the Okabe [34] or ZDT benchmark models [52] which are widely used in the EMO literature). Further, assume a point $x = (\epsilon, \tilde{x})$, where $\epsilon > 0$ is "small" and $\tilde{x} \in [0, 1]^{n-1}$ arbitrarily, is generated by the evolutionary search. Depending on g, x can be "far" from P_Q as well as $F(x)$ be "far" from $F(P_Q)$. Clearly, a point $z \in [0, 1]^n$ can only dominate x if $z_1 \leq x_1$. The probability for that might be low when using stochastic search (the probability is ϵ when z is chosen uniformly at random from $[0, 1]^n$—not counting the required improvement according to g). Note that this does in contrast not hold for mathematical programming techniques: given any feasible solution, a descent direction can be computed (e.g., [18], [37]), and hence, a sequence of dominating solutions can be generated leading to a (local) solution of the MOP. The integration of local search, however, is not an issue in this paper but will be left for future investigation.

The next empirical result confirms the above considerations. Fig. 1 shows two typical results using the well-known state-of-the-art MOEA NSGA-II [11] on the three-objective benchmark model DTLZ1 [9]. The Pareto front of DTLZ1 is given by the triangle with the corners $(1/2, 0, 0)$, $(0, 1/2, 0)$, and $(0, 0, 1/2)$. Hence, both approximations F_1 and F_2 can be considered to be "good," however, both of them contain several outliers. If d_H is used to measure the distance of F_i, $i = 1, 2$, to the Pareto front, none of the two values represents this.

Finally, we consider one example that illustrates the averaging effect in the evaluation of the outcome set (compare to Fig. 2): assume a hypothetical discrete Pareto front is given by P where $p_i = ((i - 1) \cdot 0.1, 1 - (i - 1) \cdot 0.1)^T$, $i = 1, \ldots, 11$. Further, we are given two approximations of P: X_1 is identical to P except for the first element $x_{1,1} = (\epsilon, 10)^T$

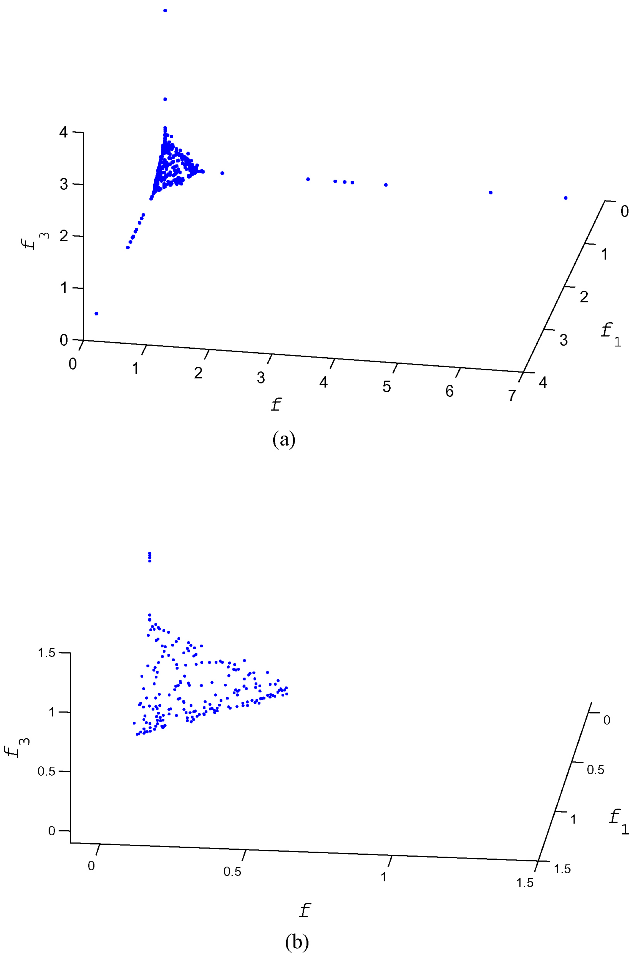


Fig. 1. Two typical results from NGSA-II on the benchmark model DTLZ1 with three objectives. (a) Front F_1. (b) Front F_2.

(an "outlier," for numerical evaluations we will use $\epsilon = 0.001$), i.e., $X_1 = \{x_{1,1}, p_2, \ldots, p_{11}\}$. X_2 is a translation of P defined by $x_{2,i} = p_i+(\epsilon/2, 5)^T$, $i = 1, \ldots, 11$. Now, we have to ask ourselves which approximation is "better." This certainly depends on our preference. However, when designing an indicator (i.e., reducing the constellation of two different sets down to one scalar value) we have to answer this question. X_1 is nearly perfect but contains one outlier, while none of the elements of X_2 are "near" to P (though the difference of each element is less than given by the single outlier in X_1). When considering the worst case scenario, X_2 is certainly better than X_1. When taking the Hausdorff distance d_H (see the definition in Section II) we obtain $d_H(X_1, P) \approx 9$ and $d_H(X_2, P) \approx 5$, i.e., X_2 is "better" than X_1 when considering d_H (which penalizes outliers). The situation changes when averaging the distances: when using, e.g., the averaged Euclidean distance from X_1 to P (i.e., using GD with $p = 1$ as described in Section III-A) we obtain $GD(X_1, P) \approx 0.81$ and $GD(X_2, P) \approx 4.54$. Hence, in this case X_1 is a "better" approximation than X_2.

The aim of this paper is to present Δ_p, an indicator that evaluates the averaged Hausdorff distance from the image of the output set to the Pareto front of the given MOP. This is intended to give EMO researchers a fair basis to evaluate their MOEA with respect to an approximation of the Pareto front

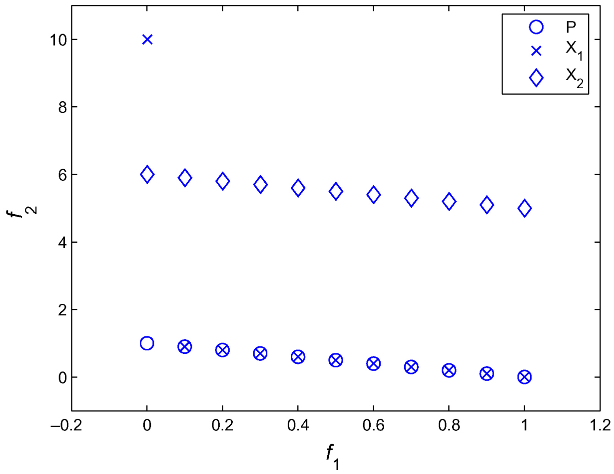


Fig. 2. Hypothetical example for a Pareto front (P) and two different approximations X_1 and X_2.

in the Hausdorff sense. In particular, the contributions of this paper are the following.

a) We will argue that both indicators GD and IGD have to be modified slightly and discuss its properties. As results, we will see that the new variant of GD, called GD_p, can be put in a more positive light with respect to its compliance with Pareto optimality, and IGD_p has certain relations to other distance measures used in the EMO literature. Furthermore, both indicators seem to be more fair when comparing outcome sets with different magnitudes. This is, in particular, interesting when comparing the performance of different archive-based MOEAs (e.g., ϵ-MOEA [8] or ELMA [14]).
b) We will propose Δ_p which consists of GD_p and IGD_p and which can be viewed as an averaged Hausdorff distance. Here, we address the (averaged) distance between the image of the outcome set and the Pareto front. We also address one possibility to handle the "outlier tradeoff" (i.e., penalizing single outliers but having a metric versus diminishing the influence of outliers by considering averaged results while losing the advantages of a metric by violating the triangle inequality). We show the potential of the new indicator on theoretical and empirical results.
c) We will next to discrete or discretized problems also address the problem of how to handle continuous models which has to our best knowledge not been done before for GD and IGD. The knowledge of the indicator for the continuous case is in particular interesting to estimate the approximation error when discretizing the Pareto front.

A preliminary study of this paper can be found in [39].

The remainder of this paper is organized as follows. Section II gives the required background for the understanding of the sequel. In Section III, we argue that a slight modification of GD and IGD leads to more fair indicators and discuss further on the variants GD_p and IGD_p. Based on these two indicators, we construct and discuss the averaged Hausdorff distance Δ_p. In Section IV, we address the extension of the three performance measurements to the case where the MOP is continuous. In Section V, we present some numerical results, and finally, we conclude in Section VI.

II. Background

In the following, we consider MOPs which are of the form:

$$\min_{x\in Q}\{F(x)\} \tag{2}$$

where the function F is defined as the vector of the objective functions

$$F : Q \subset \mathbb{R}^n \to \mathbb{R}^k \qquad F(x) = (f_1(x), \dots, f_k(x))$$

and where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is continuous. In the next definition, we state the classical concept of optimality for MOPs.

Definition 1:

a) Let $v = (v_1, \dots, v_k)$, $w = (w_1, \dots, w_k) \in \mathbb{R}^k$. Then the vector v is less than w ($v <_p w$), if $v_i < w_i$ for all $i \in \{1, \dots, k\}$. The relation $\leq_p$ is defined analogously.
b) A vector $y \in \mathbb{R}^n$ is *dominated* by a vector $x \in \mathbb{R}^n$ (in short $x \prec y$) with respect to (2) if $F(x) \leq_p F(y)$ and $F(x) \neq F(y)$ (i.e., there exists a $j \in \{1, \dots, k\}$ such that $f_j(x) < f_j(y)$).
c) A point $x \in \mathbb{R}^n$ is called Pareto optimal or a Pareto point if there is no $y \in \mathbb{R}^n$ which dominates x.

Denote by P_Q the Pareto set of (2) and its image $F(P_Q)$ the Pareto front. In the following, we will assume that P_Q is compact. This is, for instance, always given if the domain Q is compact which is in turn typically given if Q is defined by inequality and equality constraints. As one example, which is also the most common one considered in EMO literature, assume the domain is given by box-constraints, that is

$$Q = B_{l,u} := \{x \in \mathbb{R}^n \ : \ l_i \leq x_i \leq u_i, \ i = 1, \dots, n\} \tag{3}$$

where $l = (l_1, \dots, l_n)$, $r = (r_1, \dots, r_n) \in \mathbb{R}^n$ with $l_i \leq u_i$, $i = 1, \dots, n$.

In the following, we define metrics and related functions [23].

Definition 2: Suppose X is a set and d is a real function defined on the Cartesian product $X \times X$. Then d is called a metric on X if, and only if, for each $a, b, c \in X$:

a) (positive property) $d(a, b) \geq 0$ with equality if, and only if, $a = b$;
b) (symmetric property) $d(a, b) = d(b, a)$;
c) (triangle inequality) $d(a, c) \leq d(a, b) + d(b, c)$.

d is called a *semi-metric*, if properties (a) and (b) are satisfied. If a semi-metric satisfies the relaxed triangle inequality

$$d(a, c) \leq \sigma(d(a, b) + d(b, c)), \forall\, a, b, c \in X \tag{4}$$

for a value $\sigma \geq 1$, d is called a pseudo-metric. In the following, we will consider X as the set of compact subsets of the $\mathbb{R}^k$. A well-known metric on X is the Hausdorff distance d_H.

Definition 3: Let $u, v \in \mathbb{R}^n$, $A, B \subset \mathbb{R}^n$, and $\|\cdot\|$ be a vector norm. The Hausdorff distance $d_H(\cdot, \cdot)$ is defined as follows:

a) $\mathrm{dist}(u, A) := \inf_{v\in A} \|u - v\|$;

Algorithm 1 Generic Stochastic Search Algorithm

```
P_0 ⊂ Q drawn at random
A_0 = ArchiveUpdate(P_0, ∅)
for j = 0, 1, 2, ... do
    P_{j+1} = Generate(P_j)
    A_{j+1} = ArchiveUpdate(P_{j+1}, A_j)
end for
```

b) $\mathrm{dist}(B, A) := \sup_{u \in B} \mathrm{dist}(u, A)$;

c) $d_H(A, B) := \max(\mathrm{dist}(A, B), \mathrm{dist}(B, A))$.

Given a candidate set $A = \{a_1, \ldots, a_N\}$ (in objective space) and a Pareto front $F(P_Q) = \{y_1, \ldots, y_M\}$, the generational distance (GD, see [49]) and the inverted generational distance (IGD, see [5]) are defined as follows:

$$GD(A) := \frac{1}{N} \left(\sum_{i=1}^{N} d_i^p \right)^{1/p} \tag{5}$$

where d_i denotes the minimal Euclidean distance from a_i to $F(P_Q)$ (though in principle any other norm can be chosen depending on the user's preferences), and

$$IGD(A) := \frac{1}{M} \left(\sum_{i=1}^{M} \tilde{d}_i^p \right)^{1/p} \tag{6}$$

where $\tilde{d}_i$ denotes the minimal Euclidean distance from y_i to A.

There exist next to GD and IGD quite a few performance indicators for the evaluation of MOEAs. The most prominent ones are the $\mathcal{S}$ metric (or hypervolume indicator, see [3], [51]), the ϵ-Indicator [51], the error ratio [49], and Schott's spacing metric [38]. A discussion of these and further indicators can be found in [25] and [54]. However, it has to be noted that none of them are related to Hausdorff approximations of the set of interest.

Within this paper we will concentrate on the evaluation of the outcome sets of stochastic search algorithms such as multiobjective evolutionary algorithms (MOEAs). Most of such procedures consist basically of two operators: a generator and an archiver which are applied in a loop, see Algorithm 1. The task of the generator is to generate a new set of candidate solutions P_{j+1} from a given set (or population) P_j, where j denotes the current iteration step. The task of the archiver is to store and update the sequence of archives A_j by the data coming from the generator. In the following, we will refer to the archive as the candidate set obtained by the MOEA. (Alternatively, the word *population* could be used. In this paper, we will not distinguish between these two notations.)

III. Investigating the Indicators

Here, we discuss GD and IGD with respect to their ability to measure the distance between a candidate set and the Pareto front. We argue that a slight change in both definitions makes both indicators more "fair," in particular when comparing sets with different magnitudes. Out of these two modifications (GD_p and IGD_p), we will derive a "new" indicator, Δ_p. We will investigate all three indicators with respect to their metric properties, their relation to other distance measurements used in EMO literature, and their compliance to "Pareto optimality" (i.e., the compliance of the indicators with the dominance relation or, more general, with modern Pareto-based MOEAs).

In the following, we will assume that the Pareto front is discrete or discretized; extensions to continuous models will be studied in Section IV.

A. GD

1) *Discussion of the Original Indicator:* Given two finite sets $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_M\}$, and using *dist*, the indicator GD as proposed in [49] can be written as follows:

$$GD(X, Y) := \frac{1}{N} \left(\sum_{i=1}^{N} dist(x_i, Y)^p \right)^{1/p} = \frac{\|d_{XY}\|_p}{N} \tag{7}$$

where $d_{XY} \in \mathbb{R}^N$ is the associated vector of distances, i.e., the ith component is given by $(d_{XY})_i = dist(x_i, Y)$. If not explicitly stated otherwise, we will use the 2-norm for *dist*. However, in general, the q-norm can be taken, that is

$$dist_q(x_i, Y) = \inf_{y \in Y} \|x_i - y\|_q. \tag{8}$$

Though used in many studies, GD is not accepted by all researchers in the EMO community. We conjecture that the main reason for this, at least in the context of distance assignment, is its averaging strategy as the following example demonstrates: assume we are given one (arbitrary) point $a \in Q$, and without loss of generality let the distance of the image $F(a)$ toward the Pareto front be 1. Now, define the archive A_n as the multiset which is given by n copies of a, i.e., $A = \{a, \ldots, a\}$. Then, for the "averaged" distance of $F(A)$ toward the Pareto front it holds

$$GD(F(A_n), F(P_Q)) = \frac{\|(1, \ldots, 1)^T\|_p}{n} = \frac{\sqrt[p]{n}}{n}. \tag{9}$$

We see that: 1) with increasing number n, the approximation quality gets "better" though the approximation has apparently not changed, and 2) the sequence of archives A_n converges even to a "perfect" approximation, i.e., it is

$$\lim_{n \to \infty} GD(F(A_n), F(P_Q)) = 0. \tag{10}$$

The result in (10) can be generalized: instead of multisets, one can for instance consider small perturbations of a. Or, if the image $F(A)$ is bounded, even *any* sequence of archives A_n with $|A_n| = n$ can be chosen, regardless if the entries a of A_n are dominated or not, nor how far $F(a)$ is away from the Pareto front. Hence, in the context of EMO, it is advantageous from this point of view to "fill" the archive with further, even dominated, solutions since typically larger sets yield better GD values. In the community, it has been established to fix the population size in order to allow a comparison of different algorithms (say, $N_{\mathrm{pop}} = 100$). However, this leads to trouble for MOEAs which are based on archives that are not bounded by an *a priori* defined value (but rather indirectly, e.g., by the use of ϵ-dominance as in [8], [29], [43], and [44]). A "perfect" archiver (with respect to GD) is, hence, the one that accepts *all* (or at least as many as possible) candidate solutions. An effect which is certainly not desired.

2) *An Alternative Version of GD:* To avoid the effect discussed above, we propose a nearby modification of the indicator, namely to use the power mean[1] to average the distances $dist(x_i, Y)$, that is

$$GD_p(X, Y) := \left(\frac{1}{N}\sum_{i=1}^{N} dist(x_i, Y)^p\right)^{1/p} = \frac{\|d_{XY}\|_p}{\sqrt[p]{N}}. \quad (11)$$

We name the new indicator here GD_p (i.e., with the index p) only to distinguish between the classical version which is needed for further comparison in this paper. The "new" indicator does not have the unwanted characteristic as discussed above and seems hence to be more fair for a comparison of sets with different magnitudes. In particular, large candidate sets do not have to be "good" any more. For instance, for the above example we have $GD(F(A_n), F(P_Q)) = 1$ for all numbers $n \in \mathbb{N}$. The discussion in the next subsection shows that the Pareto compliance gets improved significantly by the modification.

3) *Compliance to Pareto Optimality:* Here, we investigate the compliance of GD_p with Pareto-based MOEAs. Apparently, the question cannot be answered right away since the Pareto front of the given MOP is *a priori* not known. However, the answer can be given at least indirectly: according to [4], state-of-the-art MOEAs share three characteristics which are crucial in the present context:

a) they incorporate a selection mechanism based on Pareto optimality (i.e., based on the dominance relation defined in Definition 1);
b) they adopt a diversity preservation mechanism that avoids that the entire population converges to a single solution;
c) they incorporate elitism.

Interesting in the current context are points a and c: the following results show that dominance replacements lead in certain situations to better GD_p values of the archive, which shows a certain compliance of the indicator with state-of-the-art Pareto-based MOEAs. To investigate this compliance, we will first address *dist* for single solutions, and will proceed with a consideration of GD_p on sets.

The following proposition states that if two objectives are considered and the Pareto front is connected, then dominating solutions always offer better *dist* values than the dominated ones.

Proposition 1: Let $k = 2$ and $F(P_Q)$ be connected. Then, for $a, b \in Q$ it holds

$$a \prec b \Rightarrow dist(F(a), F(P_Q)) < dist(F(b), F(P_Q)). \quad (12)$$

Proof: Let $a, b \in Q$ with $a \prec b$. Since P_Q is compact, there exists a point $p_b \in P_Q$ such that

$$dist(F(b), F(P_Q)) = \|F(b) - F(p_b)\| > 0 \quad (13)$$

(the positivity follows since a dominates b). If $a \in P_Q$, the claim follows since then $dist(F(a), F(P_Q)) = 0$, hence, we can assume in the following that $a \notin P_Q$.

If $p_b \prec a$, then we have since $a \prec b$

$$\begin{aligned} &dist(F(a), F(P_Q)) \le \|F(a) - F(p_b)\| < \|F(b) - F(p_b)\| \\ &= dist(F(b), F(P_Q)) \end{aligned} \quad (14)$$

and we are done. Now assume that $p_b \nprec a$, i.e., p_b and a are mutually non-dominating. That is, there exist two indexes $i, j \in \{1, 2\}$, $i \neq j$, such that

$$f_i(p_b) < f_i(a) \quad \text{and} \quad f_j(p_b) > f_j(a). \quad (15)$$

Since $a \notin P_Q$ there exists a point $p_a \in P_Q$ such that $p_a \prec a$. Further, since $F(P_Q)$ is connected there exists a path from $F(p_a)$ to $F(p_b)$ along the Pareto front. Hence, there exists a point $\bar{p} \in P_Q$ such that $f_j(\bar{p}) = f_j(a)$, and since $\bar{p}$ and p_b are mutually non-dominating we obtain

$$\begin{aligned} dist(F(a), F(P_Q)) &\le \|F(a) - F(\bar{p})\| = |f_i(a) - f_i(\bar{p})| \\ &< |f_i(b) - f_i(p_b)| \le \|F(b) - F(p_b)\| \\ &= dist(F(b), F(P_Q)) \end{aligned} \quad (16)$$

and the proof is complete. ■

One interesting question is certainly what happens if more than two objectives are involved in a MOP. However, we have to leave this for future investigation.

The above result does not hold when the Pareto front is disconnected. However, this "monotonic behavior" again holds if an element is close enough to the Pareto set. The following example and proposition give the counterexample and the proof, respectively.

Example 1: Let the Pareto front be given by

$$F(P_Q) = \{(10, 0)^T, (0, 1)^T\} \quad (17)$$

and further the points a, b with $F(a) = (5, 2)^T$ and $F(b) = (11, 3)^T$ (see Fig. 3). Then, it is $a \prec b$, but

$$\begin{aligned} &dist(F(b), F(P_Q)) = \sqrt{1^2 + 3^2} = \sqrt{10} < \sqrt{29} = \sqrt{5^2 + 2^2} \\ &= dist(F(a), F(P_Q)) \end{aligned} \quad (18)$$

i.e., the distance of F(a) toward the Pareto front is larger than the distance from $F(a)$.

Proposition 2: Let $a, b \in Q$ such that $a \prec b$ and

$$\begin{aligned} &\forall i = 1, \dots, k;\ \exists y(a, i) \in F(P_Q) \text{ s.t. } f_j(a) \\ &= y(a, i)_j\ \forall j \in \{1, \dots, k\} \backslash \{i\}. \end{aligned} \quad (19)$$

Then

$$dist(F(a), F(P_Q)) < dist(F(b), F(P_Q)). \quad (20)$$

Proof: Since P_Q is compact, there exists a point $p_b \in P_Q$ such that

$$dist(F(b), F(P_Q)) = \|F(b) - F(P_b)\|. \quad (21)$$

First, let us assume that $p_b \prec a$, then since $a \prec b$ we have

$$dist(F(a), F(P_Q)) \le \|F(a) - F(P_b)\| < \|F(b) - F(p_b)\| \quad (22)$$

[1] Also known as generalized mean or Hölder mean.

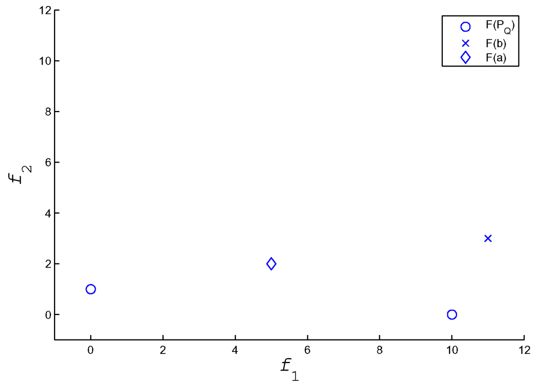


Fig. 3. If $a \prec b$, it does not follow that $dist(F(a), R(P_Q)) < dist(F(b), F(P_Q))$ (compare to Example 1).

and the claim follows. Second, assume that $p_b \nprec a$. Then there exists an index $i \in \{1, \dots, k\}$ such that $f_i(p_b) < y(a, i)_i$, and we obtain

$$\begin{aligned} dist(F(a), F(P_Q)) &\le \|F(a) - y(a, i)\| = f_i(a) - y(a, i)_i \\ &< f_i(b) - f_i(p_b) \\ &\le \|F(b) - F(P_b)\| = dist(F(b), F(P_Q)). \end{aligned} \tag{23}$$

■

Crucial for this result is the existence of the projections $y(a, i)$. This is given if $F(a)$ is close enough to the Pareto front (compare to Fig. 4), and in this case connectedness of $F(P_Q)$ is not required.

To summarize, dominating solutions a yield better *dist* values than its dominated points b in case the Pareto front is connected (at least for $k = 2$). Further, this holds when $F(a)$ is either "sufficiently far away" from the Pareto front [in this case, the claim follows with (13) since then p_b has to dominate a] or sufficiently close to it (Proposition 2).

These results can, in light of GD_p, be interpreted as follows: if the new archive results from the former one by replacement of one dominated solution by a dominating one, the GD_p value decreases. That is, for $A_1 = \{b, x_2, \dots, x_n\}$ and $A_2 = \{a, x_2, \dots, x_n\}$, where a and b are as above, it is

$$GD_p(F(A_2), F(P_Q)) < GD_p(F(A_1), F(P_Q)). \tag{24}$$

The following result is more general, however, and requires further assumptions.

Proposition 3: Let $A, B \subset \mathbb{R}^n$ be finite sets such that:

1) $\forall a \in A\ \exists b \in B: \quad F(b) \le_p F(a)$;
2) $\forall b \in B\ \exists a \in A: \quad F(b) \le_p F(a)$;
3) $\exists b \in B \backslash A, \exists a \in A \backslash B: \quad b \prec a$;
4) $\forall a \in A, \forall b \in B: \quad$ if $a \prec b \Rightarrow dist(F(a), F(P_Q)) < dist(F(b), F(P_Q))$.

Then

$$GD_p(F(B), F(P_Q)) < GD_p(F(A), F(P_Q)). \tag{25}$$

Proof: Let $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_m\}$. Now we rearrange B as follows: choose $B_1 \subset B$ as the set of elements of B whose images are partially less than $F(a_1)$, that is

$$B_1 := \{b \in B \mid F(b) \le_p F(a_1)\}. \tag{26}$$

By assumption 1, it is $m_1 \ge 1$. If $B_1 \ne B$, proceed with B_2 as the subset of $B \backslash B_1$ those images are partially less than $F(a_2)$, and so on. This leads to a sequence $B_1, \dots, B_\nu$, $\nu \le n$. By assumption 2, it follows that $B = B_1 \cup \dots \cup B_\nu$, where $|B_i| = m_i \ge 1$, $i = 1, \dots, \nu$ and $\sum_{i=1}^{\nu} m_i = m$. Using the B_i, we can write

$$\begin{aligned} &GD_p(F(B), F(P_Q)) \\ &\quad = \left(\frac{1}{m} \sum_{b \in B_1} dist(F(b), F(P_Q))^p \right. \\ &\qquad \left. + \dots + \sum_{b \in B_\nu} dist(F(b), F(P_Q))^p \right)^{1/p}. \end{aligned} \tag{27}$$

By assumptions 3 and 4, and using (27) it follows that

$$\begin{aligned} GD_p(F(B), F(P_Q))^p &< \frac{1}{m}(m_1 dist(F(a_1), F(P_Q))^p \\ &\qquad + \dots + m_\nu dist(F(a_\nu), F(P_Q))^p) \\ &\le \frac{m_1}{m} dist(F(a_1), F(P_Q))^p + \dots + \frac{m_\nu}{m} dist(F(a_\nu), F(P_Q))^p \\ &\quad + dist(F(a_{\nu+1}), F(P_Q))^p + \dots + dist(F(a_n), F(P_Q))^p \\ &\le GD_p(F(A), F(P_Q))^p. \end{aligned} \tag{28}$$

■

Assumptions 1 to 3 say, roughly speaking, that B "evolves" out of A by dominance replacement, but B does not contain any point outside the region of dominance of A. In the EMO literature, other (more intuitive) dominance relations between sets have been introduced, which can, however, not be taken in our setting. For instance, Hansen and Jaskiewicz [21] defined complete outperformance of sets as follows: B completely outperforms A (in short $B \prec_c A$), if for every solution $a \in A$ there exists a solution $b \in B$ such that $b \prec a$. Note that if $B \prec_c A$, then also $B \cup C \prec_c A$ for *every* set C, and its members can be either "far away" from the Pareto set, or be contained outside the region of dominance of A. In both cases, no prediction can be made on averaged distance to the Pareto front, i.e., on the GD_p value.

Note that the scenario described by the assumptions 1 to 3 involves the situations shown in Fig. 5. One important implication is that the result is independent of the magnitudes of A and B (which is in contrast to the classical version of GD).

4) *Metric Properties:* Here, we discuss the metric properties of GD_p (see Definition 1) which are the same as for the classical variant GD.

Due to the non-negativity of norms, also GD_p is non-negative, i.e., it is $GD_p(X, Y) \ge 0$ for all finite sets X and Y. However, it is

$$GD_p(X, Y) = 0 \ \Leftrightarrow\ X \subset Y \tag{29}$$

and hence, the positive property is not satisfied since X can be a proper subset of Y.

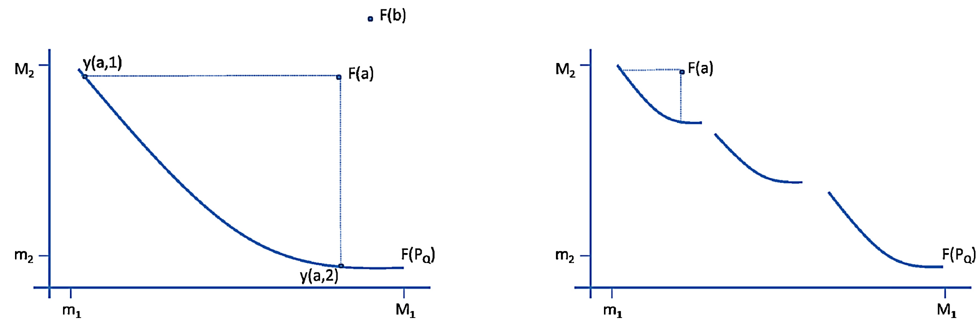


Fig. 4. Examples where dominance replacement leads to better *dist* values (compare to Proposition 2).

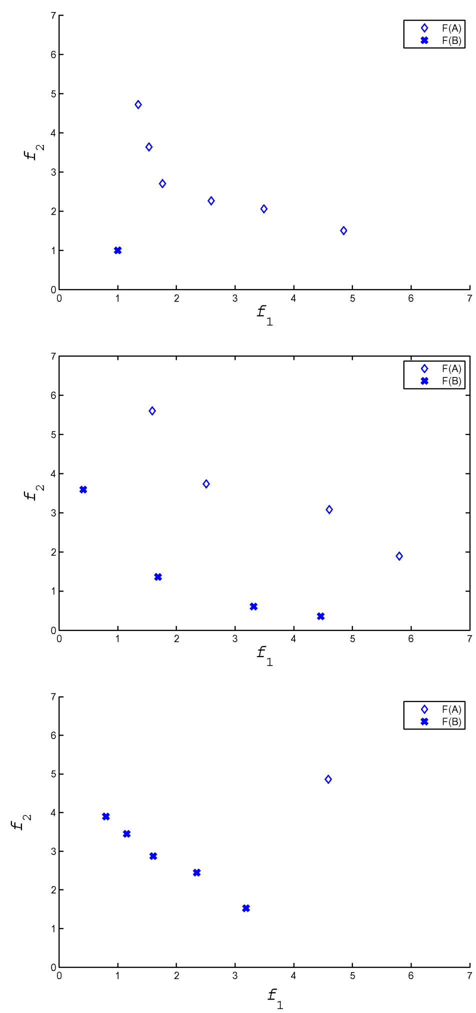


Fig. 5. Three different scenarios where the GD_p value of archive B is better than the GD_p value of archive A (under the additional assumptions made in Proposition 3).

Further, GD_p is not symmetric. For this, consider two sets X, Y such that X is a proper subset of Y. Then, it is $GD_p(X, Y) = 0$ and $GD_p(Y, X) > 0$.

Finally, also the triangle inequality does not hold. As an example, consider $A = \{(2, 3)^T, (4, 5)^T\}$, $B = \{(9, 3)^T, (5, 4)^T\}$, and $C = \{(7, 10)^T, (9, 6)^T\}$. The triangle equality is violated for $p \leq 3$, i.e., $GD_p(A, C) > GD_p(A, B) + GD_p(B, C)$ for $p = 1, 2, 3$.

The next example shows that GD_p does not satisfy a relaxed triangle inequality of the form (4) for $p < \infty$ if the number of elements in the sets are not bounded. For this, consider any two sets X and Z such that $GD(X, Z) > 0$. Given these two sets, the right-hand side of the triangle equation reads as

$$rhs(Y) := GD_p(X, Y) + GD_p(Y, Z) = \frac{\|d_{XY}\|_p}{\sqrt[p]{|X|}} + \frac{\|d_{YZ}\|_p}{\sqrt[p]{|Y|}}. \tag{30}$$

Now, choose the set Y_n as follows:

$$Y_n := X \cup \{y_1, y_2, \ldots, y_n\} \tag{31}$$

such that the values $\delta_i := dist(y_i, Z)$ are monotonically decreasing with $\delta_i \to 0$ for $i \to \infty$ and $\sum_{i=1}^{\infty} \delta_i^p < \infty$. By construction, we have $\|d_{XY}\|_p = 0$ and $\|d_{YZ}\|_p / \sqrt[p]{|Y_n|} \to 0$, i.e., $rhs(Y_n) \to 0$ for $n \to \infty$. That is, there is no value σ such that (4) is fulfilled for all such sets Y_n.

5) *Relation to Other Distance Measurements:* Apparently, GD_p has a relation to *dist*, that is

$$GD_\infty(A, B) = dist(A, B). \tag{32}$$

That is, for $p < \infty$, GD_p can be viewed as an "averaged" version of *dist*.

B. *IGD*

Here, we proceed with a discussion of the IGD indicator analog to GD. We will propose the same modification, IGD_p, which has the same (poor) metric properties as GD_p. Surprisingly, the new indicator is related to many distance measurements used in the EMO literature.

1) *Discussion of the Original Indicator:* Analog to GD, the indicator IGD as proposed in [5] can be written as follows:

$$IGD(X, Y) := \frac{1}{M}\left(\sum_{i=1}^{M} dist(y_i, X)^p\right)^{1/p} = \frac{\|d_{YX}\|_p}{M} \tag{33}$$

where $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_M\}$. Apparently, it is

$$IGD(X, Y) = GD(Y, X) \tag{34}$$

for all finite sets X and Y. Hence, in principle the same argumentation can be applied to justify a modification of

TABLE I
NUMERICAL VALUES FOR THE IGD AND IGD_p INDICATORS FROM EXAMPLE 2

p	Indicator	Y_1	Y_2
$p = 1$	IGD	0.3857	0.3571
	IGD_1	0.3857	0.3571
$p = 2$	IGD	0.1348	0.0410
	IGD_2	0.4472	0.4123
$p = \infty$	IGD	0.0643	0.0070
	IGD_∞	0.7071	0.7071

the operator. In the context of multiobjective optimization, a (suitable) discretization Y of the Pareto front has to be chosen. Analog to the discussion for GD, the IGD value gets better when choosing a finer discretization of the Pareto front: assume we are given an archive A, and two discretizations Y_1 and Y_2 of the Pareto front, where Y_2 is finer than Y_1 (i.e., better in the Hausdorff sense and contains more elements). Then, it is $IGD(Y_2, F(A)) < IGD(Y_1, F(A))$ (see also Example 2). Though this problem can in principle be avoided by fixing a discretization of the Pareto front, this is also an unwanted effect. Also, as we will see later on, the classical IGD indicator allows no extension to continuous models.

2) *Alternative Version of IGD:* Motivated by the above discussion we propose to use the power mean as for GD_p, i.e., to use

$$IGD_p(X, Y) := \left(\frac{1}{M}\sum_{i=1}^{M} dist(y_i, X)^p\right)^{1/p} = \frac{\|d_{YX}\|_p}{\sqrt[p]{M}}. \quad (35)$$

Example 2: We consider the Pareto front of a hypothetical MOP that is the line segment between the points $y_1 = (0, 1)^T \in \mathbb{R}^2$ and $y_2 = (1, 0)^T \in \mathbb{R}^2$, that is

$$F(P_Q) = \{\lambda y_1 + (1-\lambda)y_2 \ : \ \lambda \in [0, 1]\}. \quad (36)$$

Further, we consider the two following discretizations Y_1 and Y_2:

$$\begin{aligned} Y_1 &= \{(i * 0.1, 1 - i * 0.1)^T \ : \ i \in \{0, \dots, 10\}\} \\ Y_2 &= \{(i * 0.01, 1 - i * 0.01)^T \ : \ i \in \{0, \dots, 100\}\} \end{aligned} \quad (37)$$

i.e., we have $|Y_1| = 11$ and $|Y_2| = 101$. We assume, for simplicity, that the archive consists only of one point, $A = \{a\}$, with $F(A) = (0.5, 0.5)^T$. Different IGD and IGD_p values are shown in Table I. The following observations can be made: the IGD values get lower for the finer approximation Y_2. This is in accord to the related discussion on GD: a given approximation A can be made "better" (measured by IGD) simply by refining the approximation of the Pareto front and without changing A which is against the intuition of an approximation quality indicator. On the other hand, such a decay cannot be observed in the IGD_p values. To get a better understanding of the difference in these values we refer to Example 4.

3) *Metric Properties:* Due to (34), IGD_p has the same metric properties as GD_p, i.e., GD_p is merely non-negative. In particular, it is

$$IGD_p(X, Y) = 0 \ \Leftrightarrow \ Y \subset X. \quad (38)$$

In the context of multiobjective optimization, this means that whenever the Pareto front is contained in the image of the archive $F(A)$, then the IGD_p value is zero.

4) *Compliance to Pareto Optimality:* As for GD_p, the compliance of IGD_p with state-of-the-art MOEAs cannot be answered directly, but rather indirectly. In this context, a combination of the characteristic (2) and (3) of modern MOEAs (see related discussion on GD_p) is most influential: a higher diversity among the archive entries can lead to a better IGD_p as it may reduce the maximal gap in the distance from the Pareto front to the image of the outcome set. Up to date, there exist quite a few diversity preservation mechanisms. There are, for instance, fitness sharing schemes [10], [19], clustering [48], the adaptive grid [26], [27], the crowded-comparison operator [11], and entropy [6], [16], [17], among others. Further, there exist algorithms that are specialized on a movement *along* the Pareto set such as multiobjective continuation methods [22], [41], [45] or the mutation operator HCS [28] which may be helpful to increase the diversity among the archive entries. Future studies have to show if and to which extent these operators indeed help to improve the IGD_p value.

5) *Relation to Other Distance Measurements:* Before we can discuss the relations to other measurements, we have to state the following definitions.

Definition 4 [32]: Let $\epsilon \in \mathbb{R}^k_+$ and $x, y \in \mathbb{R}^n$. x is said to ϵ-dominate y (short $x \prec_\epsilon y$), with respect to (MOP) if $F(x) - \epsilon \leq_p F(y)$ and $F(x) - \epsilon \neq F(y)$.

Definition 5 [29]: Let $\epsilon \in \mathbb{R}^k_+$. A set $A \subset \mathbb{R}^n$ is called an ϵ-approximate Pareto set of (MOP) if for all $x \in \mathbb{R}^n$ there exists an $a \in A$ such that $a \prec_\epsilon x$.

In the following, we use the notation $1\epsilon := (\epsilon, \dots, \epsilon) \in \mathbb{R}^k_+$ for $\epsilon \in \mathbb{R}_+$.

Definition 6 [53]: Let $A, B \subset \mathbb{R}^n$. The ϵ-Indicator of A and B is defined as

$$I_{\epsilon^+}(A, B) := \min_{\epsilon} \in \mathbb{R}_+\{\forall b \in B \ \exists a \in A : \ F(a) - 1\epsilon \leq_p F(b)\}. \quad (39)$$

Definition 7 [36]: Let d be a metric, $\delta > 0$, and $D \subset Z$ be a discrete set. D is called a d_δ representation of Z if for any $z \in Z$ there exists an element $y \in D$ such that $d(z, y) \leq \delta$.

Now we are in the position to state the relations of IGD_p to the different distance measurements.

First, there is the relation to *dist*. Analog to (32) we have

$$IGD_\infty(X, Y) = dist(Y, X). \quad (40)$$

The next proposition gives the relation of IGD_∞ to the measurements based on ϵ-dominance. Hereby, we use IGD^q_∞ to indicate that the q-norm is used for $d(a, b)$ [see (8)].

Proposition 4: Let $A \subset \mathbb{R}^n$ be given.

a) A is a $1c$-approximate Pareto set of the MOP, where $c := IGD^q_\infty(F(A), F(P_Q))$.

b) $I_{\epsilon^+}(A, P_Q) = IGD^\infty_\infty(F(A), F(P_Q))$.

Proof:

Ad a): it is

$$IGD^q_\infty(F(A), F(P_Q)) = \max_{p \in P_Q} \min_{a \in A} \|F(p) - F(a)\|_q. \quad (41)$$

That is, for all $p \in P_Q$ there exists an $a \in A$ such that $\|F(p) - F(a)\|_q \leq c$. Since in particular $|f_i(p) - f_i(a)| \leq c$ for all $i = 1, \ldots, k$ it is also $a \prec_{1c} p$, and the claim follows.

Ad b): it is

$$\begin{aligned} I_{\epsilon^+}(A, P_Q) &= \min_{\epsilon \in \mathbb{R}_+}\{\forall p \in P_Q \exists a \in A \;:\; F(a) - 1\epsilon \leq_p F(p)\} \\ IGD_\infty^\infty(F(A), F(P_Q)) &= \max_{p \in P_Q} \min_{a \in A} \|F(p) - F(a)\|_\infty =: c \end{aligned} \tag{42}$$

and there exist $\bar{p} \in P_Q$, $\bar{a} \in A$ such that

$$\|F(\bar{p}) - F(\bar{a})\|_\infty = c. \tag{43}$$

That is, for all $p \in P_Q$ there exists an $a \in A$ such that $\|F(p) - F(a)\|_\infty \leq c$. Since also here $|f_i(p) - f_i(a)| \leq c$ for all $i = 1, \ldots, k$ it follows that $F(a) - 1\epsilon \leq_p F(p)$. This together with (43) completes the proof. ∎

Finally, there is a relation to the measurement in Definition III-B5, but for this we need in addition GD_∞. Let $A \in \mathbb{R}^n$. Then its image, $F(A)$, is a d_δ representation of the Pareto front iff

$$\begin{aligned} GD_\infty(F(A), F(P_Q)) = 0 \quad &\text{and} \\ IGD_\infty(F(A), F(P_Q)) \leq \delta& \end{aligned} \tag{44}$$

where d is the metric induced by the 2-norm (or more general, the q-norm).

C. "New" Indicator to Measure the Hausdorff Distance to the Pareto Front

Here, we combine GD_p and IGD_p to the "new" indicator Δ_p, which can be viewed as an "averaged Hausdorff distance."

1) *Indicator:* Inspired by the relation of GD_p and IGD_p to $dist$, we define the new indicator Δ_p as follows.

Definition 8: Let $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_m\} \subset \mathbb{R}^k$ be finite and non-empty sets. Then we define $\Delta_p(X, Y)$ by

$$\begin{aligned} \Delta_p(X, Y) :&= \max(GD_p(X, Y), IGD_p(X, Y)) \\ &= \max\left(\left(\frac{1}{N}\sum_{i=1}^{N} dist(x_i, Y)^p\right)^{1/p}, \right. \\ &\qquad \left. \left(\frac{1}{M}\sum_{i=1}^{M} dist(y_i, X)^p\right)^{1/p}\right). \end{aligned} \tag{45}$$

Example 3: We revisit the two introductory examples from Section I. Table II shows the numerical values of Δ_p for different values of p. For $p = 1, 2, 3$, X_1 is the "better" approximation, and X_2 is "better" for $p > 3$ (compare to Fig. 2). For the two fronts F_i, $i = 1, 2$, (compare to Fig. 1) obtained by NSGA-II, the Δ_p values are "good" for low values of p. That changes, however, for larger values of p since in that case outliers are penalized more.

2) *Metric Properties:* Due to its combination of GD_p and IGD_p, Δ_p has stronger metric properties than the first two indicators. For instance, in case of bounded archive sizes, Δ_p defines a pseudo-metric for all values of p. Further on, we address the "outlier tradeoff" (i.e., penalizing single outliers but having a metric in the mathematical sense versus diminishing the influence of outliers to the indicator value by considering averaged results while losing the advantages of a metric by violating the triangle inequality).

TABLE II

VALUES OF $\Delta_p(P, X_i)$ AND $\Delta_p(F(P_Q), F_i)$, $i = 1, 2$ (COMPARE TO FIGS. 1 AND 2) FOR DIFFERENT VALUES OF p

	$p = 1$	$p = 2$	$p = 3$	$p = 5$	$p = 10$	$p = \infty$
$\Delta_p(P, X_1)$	0.818	2.714	4.047	5.571	7.080	9.000
$\Delta_p(P, X_2)$	4.541	4.550	4.558	4.575	4.616	5.000
$\Delta_p(F(P_Q), F_1)$	0.159	0.566	1.059	1.939	3.289	5.885
$\Delta_p(F(P_Q), F_2)$	0.037	0.010	0.177	0.283	0.403	0.599

The higher the value of p, the more outliers are penalized by Δ_p.

Proposition 5: Δ_p is a semi-metric for $1 \leq p < \infty$ and a metric for $p = \infty$.

Proof: The positive property follows directly by the non-negativity of the norm, and (29) and (38). The symmetry follows by the construction of Δ_p. Hence, Δ_p is a semi-metric.

Let $p = \infty$, then

$$\begin{aligned} \Delta_\infty(X, Y) &= \max\left(\max_{i=1,\ldots,|X|}(dist(x_i, Y)), \max_{i=1,\ldots,|Y|}(dist(y_i, X))\right) \\ &= \max(dist(X, Y), dist(Y, X)) = d_H(X, Y) \end{aligned} \tag{46}$$

i.e., for $p = \infty$ the indicator Δ_p coincides with the Hausdorff distance. ∎

Δ_p does not satisfy the triangle inequality for $p < \infty$ which is caused by the averaging of the distances. Assume, for instance, $X = \{(7, 1)^T, (5, 3)^T\}$, $Y = \{(5, 4)^T, (3, 6)^T\}$, and $Z = \{(1, 9)^T, (3, 7)^T\}$. Then, it is $\Delta_1(X, Z) > \Delta_1(X, Y) + \Delta_1(Y, Z)$. However, in the case the magnitudes of the sets are bounded—and this is typically the case for most MOEAs—it follows that Δ_p is a pseudo-metric in the sense of (4).

Proposition 6: Let $X, Y, Z \subset \mathbb{R}^k$ be non-empty with $|X|, |Y|, |Z| \leq N$, then

$$\Delta_p(X, Z) \leq \sqrt[p]{N}(\Delta_p(X, Y) + \Delta_p(Y, Z)). \tag{47}$$

Proof:

$$\begin{aligned} \Delta_p(X, Z) &= \max\left(\frac{\|d_{XZ}\|_p}{\sqrt[p]{|X|}}, \frac{\|d_{ZX}\|_p}{\sqrt[p]{|Z|}}\right) \\ &\leq \max\left(\frac{\sqrt[p]{|X|}\|d_{XZ}\|_\infty}{\sqrt[p]{|X|}}, \frac{\sqrt[p]{|Z|}\|d_{ZX}\|_\infty}{\sqrt[p]{|Z|}}\right) \\ &= d_H(X, Z) \leq d_H(X, Y) + d_H(Y, Z) \\ &= \max(\|d_{XY}\|_\infty, \|d_{YX}\|_\infty) + \max(\|d_{YZ}\|_\infty, \|d_{ZY}\|_\infty) \\ &\leq \max\left(\sqrt[p]{|X|}, \sqrt[p]{|Y|}\right)\Delta_p(X, Y) \\ &\quad + \max\left(\sqrt[p]{|Y|}, \sqrt[p]{|Z|}\right)\Delta_p(Y, Z) \\ &\leq \sqrt[p]{N}(\Delta_p(X, Y) + \Delta_p(Y, Z)). \end{aligned} \tag{48}$$

∎

Apparently, the choice of the p-norm in (45) is the key to handle the "outlier tradeoff:" the smaller p, the higher the averaging effect and the lower the influence of single outliers. If, on the other hand, p is increased, the more the largest

TABLE III
PERCENTAGE OF THE TRIANGLE VIOLATIONS FOR DIFFERENT VALUES OF p

	$p = 1$	$p = 2$	$p = 5$	$p = 10$	$p = 20$	$p = \infty$
N = 2	0.541	0.15	0.026	0.008	0.005	0.006
N = 4	0.249	0.06	0.019	0.009	0.005	0.002
N = 6	0.105	0.033	0.008	0.003	0.001	0
N = 10	0.02	0.002	0.004	0.001	0	0
N = 100	0	0	0	0	0	0

Here 100 000 different sets X, Y, and Z with magnitude N = 2, 4, 6, 10, and 100 have been chosen, and each entry of each set has been chosen randomly from $[0, 10]^2$.

distances in $GD(X, Y)$ get dominant, and hence, outliers influence the value of $\Delta_p(X, Y)$ (recall that $\lim_{p\to\infty} \|x\|_p = \|x\|_\infty$). In the extreme case, $p = \infty$, only the farthest distances are considered (i.e., the value of the distance is determined entirely by the largest outlier), but in turn Δ_p defines a metric on the set of discrete sets.

This is reflected in Table III; it shows the percentage of the triangle inequality violations for different values of p for a sequence of randomly chosen sets X, Y, and Z with different magnitudes. The larger p, the fewer triangle inequality violations are observed, and hence, the "nearer" Δ_p is to a metric (measured empirically by the probability of a triangle inequality violation). Note that the triangle inequality violations decrease both with increasing value of p as well as with increasing number of elements considered in the sets. This might be a reason that the violation of the triangle inequality has never been observed in the literature. For practical use (i.e., assuming the magnitude of both the candidate set and the Pareto front approximation to be at least 100, and $p \geq 2$), it seems that Δ_p might be already quite "close" to a metric.

3) *Compliance to Pareto Optimality:* This follows directly by the related discussions for GD_p and IGD_p. Note that in particular all the three characteristics of a state-of-the-art MOEA (see related discussion for GD_p) are indeed helpful to decrease the Δ_p value. Hence, one can say that state-of-the-art MOEAs are in principle compliant with the new indicator Δ_p. It remains, however, to detect to what extent Pareto-based MOEAs can be evaluated by Δ_p.

4) *Relation to Other Distance Measurements:* By construction, there is a strong relation to the Hausdorff distance, i.e., it is

$$\Delta_\infty(X, Y) = d_H(X, Y). \tag{49}$$

Hence, Δ_p can be viewed as an "averaged Hausdorff distance" for $p < \infty$.

It is important to note that Δ_p is *not* compliant with the dominance relations defined by Hansen and Jaskiewicz [21] such as the complete outperformance (see also the related discussion in Section III-A on GD_p). For this, consider the third example introduced in Section I (see also Fig. 2), and consider X_2 is obtained via the translation $x_{2,i} = p_i + (2\epsilon, 5)^T$, $i = 1, \ldots, 11$, while P and X_1 are unchanged. Then, it is $X_1 \prec_c X_2$ regardless of the outlier in X_1, but one obtains very similar values as the one shown in Table II (in particular, the Δ_p value is only better for X_1 for low values of p). This is due to the fact that Δ_p considers all *dist* values of the two given sets in order to compute the "distance" between them.

IV. Extension to Continuous Models

So far, we have assumed that P_Q, and hence also $F(P_Q)$, was finite. Since the Pareto set of a continuous MOP typically forms a $(k-1)$-dimensional set, a natural question arises—at least from the theoretical point of view—how to extend the indicators to such problems which we address here. Though the "extended indicators" can hardly be solved for a general model, their definition allows us to address the (practically relevant) question of the discretization error when discretizing $F(P_Q)$ (see Proposition 7 for such a result).

A. Extension of the Indicators

In the following, we investigate how GD_p and IGD_p (and hence, also Δ_p) can be extended for the case where all objectives are continuous. Hereby, we consider the sets $A, P_Q \subset \mathbb{R}^n$, where $A = \{a_1, \ldots, a_{|A|}\}$ (i.e., the archive) is finite.

1) *GD_p for Continuous Models:* It is

$$GD_p(F(A), F(P_Q)) = \left(\frac{1}{|A|}\sum_{i=1}^{|A|} dist(F(a_i), F(P_Q))^p\right)^{1/p}. \tag{50}$$

Since P_Q is compact and F is assumed to be continuous it is

$$dist(F(a_i), F(P_Q)) = \min_{p\in P_Q} \|F(a_i) - F(p)\|. \tag{51}$$

That is, the form of GD_p does not change, but it turns from a discrete optimization problem (to be more precise, an enumeration problem) into a continuous optimization problem.

2) *IGD_p for Continuous Models:* The extension of IGD_p requires the integration over the Pareto front (see the Appendix for a derivation). Assume for the sake of a better understanding first the bi-objective case (i.e., $k = 2$) and that the Pareto front is connected. In that case, the Pareto front can be expressed as a curve $\gamma : [m_1, M_1] \subset \mathbb{R} \to \mathbb{R}^2$, where $m_1 := \min_{p\in P_Q} f_1(p)$ and $M_1 := \max_{p\in P_Q} f_1(p)$, and IGD_p reads as follows:

$$IGD_p(F(A), F(P_Q)) = \left(\frac{1}{M_1 - m_1}\int_{m_1}^{M_1} dist(\gamma(t), F(A))^p dt\right)^{1/p}. \tag{52}$$

In case $F(P_Q)$ consists of l connected components one can define in an analog way l such curves $\gamma_i : [m_{i,1}, M_{i,1}] \to \mathbb{R}^2$ such that the union of these curves is equal to the Pareto front. In that case one obtains

$$IGD_p(F(A), F(P_Q)) = \sum_{i=1}^{l}\left(\frac{1}{M_{i,1} - m_{i,1}}\int_{m_{i,1}}^{M_{i,1}} dist(\gamma_i(t), F(A))^p dt\right)^{1/p}. \tag{53}$$

Finally, we consider the general case: assume we are given a MOP with k objectives where the Pareto front consists of l connected components. Then there exist l mappings $\Phi_i : D_i \to$

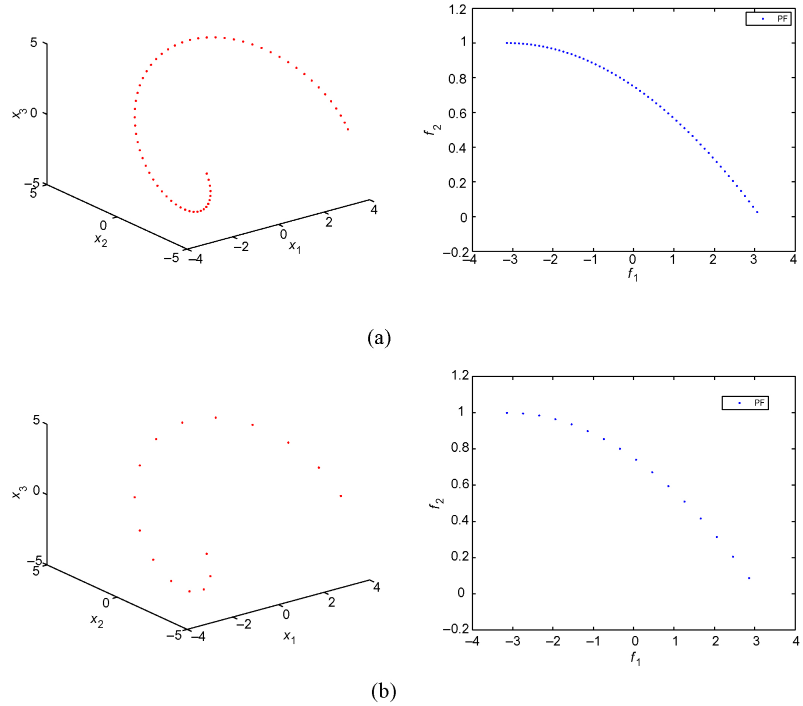


Fig. 6. Discretizations of the Pareto front of model OKA2 ([34]) using a continuation method together with the step size control described in Section IV-B. (a) $\delta = 0.1$. (b) $\delta = 0.4$.

$\mathbb{R}$, $i = 1, \ldots, l$, where $D_i \subset [m_1, M_1] \times \ldots \times [m_{k-1}, M_{k-1}]$ (where m_i and M_i are defined analogously for $i = 2, \ldots, k-1$), such that the union of the graphs of the Φ's is equal to $F(P_Q)$. Define $\Psi_i : D_i \to \mathbb{R}^k$ as $\Psi_i(x) = (x, \Phi_i(x))^T$. Then, we obtain

$$IGD_p(F(A), F(P_Q)) = \sum_{i=1}^{l} \left(\frac{1}{vol(D_i)} \int_{D_i} dist(\Psi_i(x), F(A))^p dx \right)^{1/p} \tag{54}$$

where $vol(D_i)$ is the $(k-1)$-dimensional volume of D_i, $i = 1, \ldots, l$.

B. Discretization of the Pareto Front

Though in principle the Pareto fronts of all commonly used benchmark models are given analytically, and there exist attempts to express $F(P_Q)$ analytically for a given model (e.g., [1]), the indicator values are—at least for $k > 2$—in general not easy to calculate, or relatively expensive to approximate numerically in terms of function calls. Since we assume that $F(P_Q)$ is given the question arises if it is not advantageous to use a discretization of the Pareto front (as done so far in the literature). In the following we analyze this.

Since $F(P_Q)$ is given, we can assume that we are given a finite approximation $Y \subset \mathbb{R}^k$ of the Pareto front with $d_H(Y, F(P_Q)) \leq \delta$ (i.e., Y contains no outliers, see below for one possible heuristic for the generation of Y for bi-objective problems). The natural question that arises in this context is the resulting discretization error that has to be considered when comparing different indicator values. Here, we define the approximation error in a straightforward way: given an archive A, the Pareto front $F(P_Q)$ and its discretization Y, we define the error, e.g., for GD_p as $|GD_p(F(A), F(P_Q)) - GD_p(F(A), Y))|$, and analog for the other indicators.

The following result shows that the discretization error for the three indicators under investigation is equal to the approximation quality of Y.

Proposition 7: Let $A \subset \mathbb{R}^n$ be finite, $F(P_Q)$ is given and can be expressed as in (54), and let $Y \subset \mathbb{R}^k$ be finite such that $d_H(F(P_Q), Y) \leq \delta$. Then:

a) $|GD_p(F(A), F(P_Q)) - GD_p(F(A), Y))| \leq \delta$;
b) $|IGD_p(F(A), F(P_Q)) - IGD_p(F(A), Y))| \leq \delta$;
c) $|\Delta_p(F(A), F(P_Q)) - \Delta_p(F(A), Y))| \leq \delta$.

Proof: Since $d_H(F(P_Q), Y) \leq \delta$ it holds

$$\forall p \in P_Q : \quad dist(F(p), Y) \leq \delta, \quad \text{and} \tag{55}$$
$$\forall y \in Y : \quad dist(y, F(P_Q)) \leq \delta. \tag{56}$$

Ad a): this follows by the reverse triangle inequality (RTI) and (55)

$$\begin{aligned} &|GD_p(F(A), F(P_Q)) - GD_p(F(A), Y))| \\ &= \left| \frac{1}{\sqrt[p]{|A|}} \|d_{F(A)F(P_Q)}\|_p - \frac{1}{\sqrt[p]{|A|}} \|d_{F(A)Y}\|_p \right| \\ &= \frac{1}{\sqrt[p]{|A|}} \left| \|d_{F(A)F(P_Q)}\|_p - \|d_{F(A)Y}\|_p \right| \\ &\overset{(RTI)}{\leq} \frac{1}{\sqrt[p]{|A|}} \|d_{F(A)F(P_Q)} - d_{F(A)Y}\|_p \overset{(55)}{\leq} \|(\delta, \ldots, \delta)^T\|_p \leq \delta. \end{aligned} \tag{57}$$

Ad b): here we give the proof for $k = 2$ and $l = 1$, all the other cases are analog. By assumption there exists a curve $\gamma : [m_1, M_1] \to \mathbb{R}^2$ such that

$$IGD_p(F(A), F(P_Q)) = \left(\frac{1}{M_1 - m_1} \int_{m_1}^{M_1} dist(\gamma(t), F(A))^p dt \right)^{1/p}. \quad (58)$$

In the following, we show that $IGD_p(F(A), Y)$ can be estimated above by an upper Riemann sum which leads to the desired result. Denote

$$I(i, |Y|) := \left[m_1 + \frac{i-1}{|Y|}(M_1 - m_1), m_1 + \frac{i}{|Y|}(M_1 - m_1) \right] \quad i = 1, \ldots, |Y| \quad (59)$$

i.e., the union of these intervals forms a uniform partition of the interval $[m_1, M_1]$. Define $\Delta t := (M_1 - m_1)/|Y|$, and choose a value t_i in each interval $I(i, |Y|)$. Since $d_H(Y, F(P_Q)) \le \delta$ there exists for every t_i an element $y_i \in Y$ such that $\|\gamma(t_i) - y_i\| \le \delta$, $i = 1, \ldots, |Y|$. We thus have

$$\begin{aligned} IGD_p(F(A), Y) &= \left(\frac{1}{|Y|} \sum_{i=1}^{|Y|} dist(y_i, F(A))^p \right)^{1/p} \\ &\le \left(\frac{1}{|Y|} \sum_{i=1}^{|Y|} (dist(\gamma(t_i), F(A)) + \delta)^p \right)^{1/p} \\ &= \left(\frac{1}{M_1 - m_1} \sum_{i=1}^{|Y|} (dist(\gamma(t_i), F(A)) + \delta)^p \ \Delta t \right)^{1/p} \end{aligned} \quad (60)$$

i.e., an upper Riemann sum of (58). The maximal error is hence given by

$$\begin{aligned} &\left| \left(\frac{1}{M_1 - m_1} \sum_{i=1}^{|Y|} (dist(\gamma(t_i), F(A)) + \delta)^p \ \Delta t \right)^{1/p} \right. \\ &\left. - \left(\frac{1}{M_1 - m_1} \sum_{i=1}^{|Y|} dist(\gamma(t_i), F(A))^p \ \Delta t \right)^{1/p} \right| \\ &\le \frac{1}{\sqrt[p]{|Y|}} \|(\delta, \ldots, \delta)^T\|_p = \delta \end{aligned} \quad (61)$$

and the claim follows.

Ad c): this follows immediately by a) and b). ■

Example 4: We revisit Example 2. It is $m_1 = 0$ and $M_1 = 1$, and the Pareto front can be expressed by the curve

$$\gamma : [0, 1] \to \mathbb{R}^2, \quad \gamma(t) = \begin{pmatrix} t \\ 1 - t \end{pmatrix}. \quad (62)$$

The IGD_p value for $A = \{a\}$ is given by (see the Appendix for a derivation)

$$IGD_p(F(A), F(P_Q)) = \frac{1}{\sqrt{2}} \sqrt[p]{\frac{1}{p+1}}. \quad (63)$$

In particular, we obtain the following numerical values:

$$\begin{aligned} IGD_1(F(A), F(P_Q)) &\approx 0.3536 \\ IGD_2(F(A), F(P_Q)) &\approx 0.4082 \\ IGD_\infty(F(A), F(P_Q)) &= 1/\sqrt{2}. \end{aligned} \quad (64)$$

It is $d_H(Y_1, F(P_Q)) = 0.1$ and $d_H(Y_2, F(P_Q)) = 0.01$, and the differences of the above results with the IGD_p values in Table I are in accord with the above result.

It remains to obtain such an approximation Y with the desired Hausdorff distance to the Pareto front which is not always an easy task: in almost all benchmark functions, P_Q is given explicitly and in "easy" form, however, this does only in certain cases hold for $F(P_Q)$.

In the following, we present one possible heuristic for the computation of Y for bi-objective optimization problems with differentiable objectives (the latter holds for all continuous benchmark models, even if MOEAs do typically not exploit that information). Assume we are given P_Q analytically (which can consist of one or more connected components), the question is how to get a "suitable" discretization $P = \{p_1, \ldots, p_n\}$, $p_i \in P_Q$, such that $Y := F(P)$ serves as a Pareto front approximation with $d_H(Y, F(P_Q)) \le \delta$, where $\delta \in \mathbb{R}_+$ is given *a priori*. Here, we can lean elements from the step size control for multiobjective continuation (e.g., [44]) since the main difficulty for the problem at hand is to estimate the distance $\|p_{i+1} - p_i\|_\infty$ of two "consecutive" elements p_i and p_{i+1} (for the bi-objective case, P_Q is typically a curve, and hence, the elements p_i can be arranged accordingly): by demand on $F(P)$, the distance of the images of the two consecutive solutions should be

$$\|F(p_{i+1}) - F(p_i)\|_\infty \approx \Theta\delta \quad (65)$$

where $\Theta \in (0, 1)$ is a safety factor. If F is Lipschitz continuous there exists a $L > 0$ such that

$$\|F(x) - F(y)\|_\infty \le L\|x - y\|_\infty \quad \forall x, y \in Q. \quad (66)$$

If x and y are close enough together, then the inequality in (66) turns approximately to the equality when using the local Lipschitz estimate of F around x. The latter can be estimated by

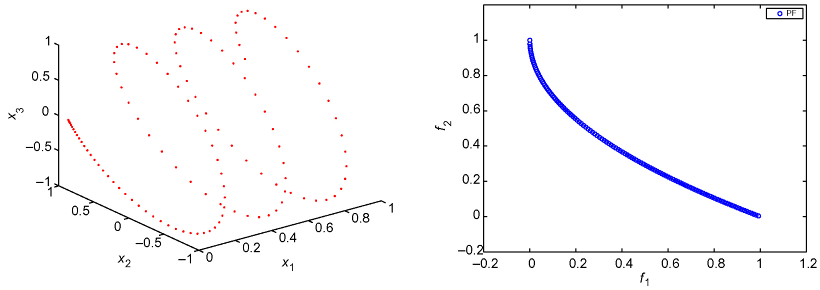
$$L_x := \|DF(x)\|_\infty = \max_{i=1,..,k} \|\nabla f_i(x)\|_1. \quad (67)$$

Putting (65) and (66) together and assuming that δ is "small," we obtain the following estimation for the distance of the two consecutive solutions:

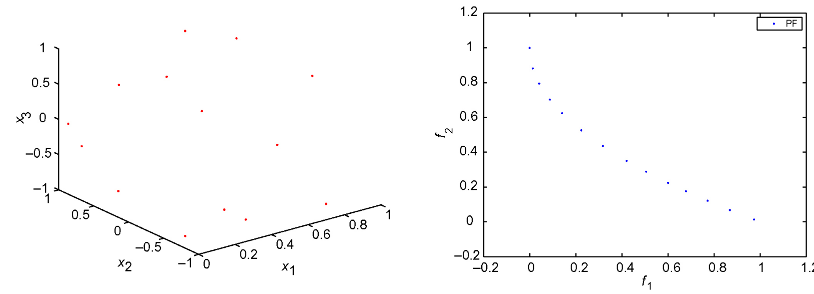
$$\|p_{i+1} - p_i\|_\infty \approx \frac{\Theta\delta}{L_{p_i}}. \quad (68)$$

From this, and the knowledge of P_Q, the next iterate can be computed. In case also P_Q is not easy to track, the above distance can be used as the step size for the predictor within a multiobjective continuation method.

The aim of the method presented above is to generate an equidistant approximation of the Pareto front. Such approximations are considered to be "optimal" by the PL-metric [33].



(a)



(b)

Fig. 7. Discretizations of the Pareto front of model UF1 ([30]) using a continuation method together with the step size control described in Section IV-B. (a) $\delta = 0.01$. (b) $\delta = 0.1$.

V. Numerical Results

Here, we attempt to demonstrate the usefulness of the novel indicator. First, we show some examples of discretizations of the Pareto front as discussed in Section IV-B. Next, we intend to show empirically that modern MOEAs indeed comply (to a certain extent) with Δ_p. For this, we have chosen to apply NSGA-II [11] on a benchmark model. It can be seen that the MOEA indeed generates good (averaged) Hausdorff approximations of the Pareto front. Finally, we want to demonstrate that Δ_p can be used to measure empirically the speed of convergence of certain archive-based MOEAs.

A. Generating Discretizations of the Pareto Front

First, we address the problem of generating a "suitable" discretization of $F(P_Q)$. Here, we have used the multiobjective continuation method proposed in [13] and [42] together with the step size control discussed in Section IV-B. Figs. 6 and 7 show results for different values of δ (in all computations, we have chosen $\Theta = 0.99$) on bi-objective problems, and Fig. 8 shows one result for a 3-objective model (see the Appendix for the definitions of the MOPs under consideration). In all cases, sufficient approximations could be obtained.

B. Measuring the Performance of NSGA-II on DTLZ1

Next, we are interested in measuring the performance of a modern Pareto-based MOEA on a benchmark model. Here, we have decided for the well-known algorithm NSGA-II and the benchmark model DTLZ1 [9] since NSGA-II is a widely accepted state-of-the-art MOEA, and DTLZ1 contains weakly optimal Pareto points which are easily detected—but not easily discarded—by an MOEA.

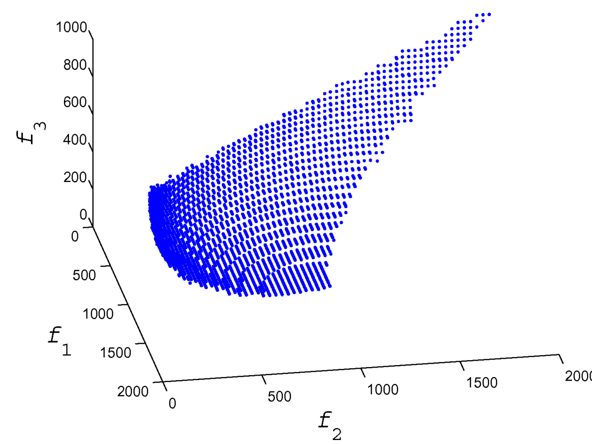


Fig. 8. Discretizations of the Pareto front of model SDD1 ([42]) using a continuation method together with the step size control described in Section IV-B. Here, we have used $\delta = 100$ leading to 3421 solutions.

Fig. 9 and Table IV show the values of GD_p, IGD_p, and Δ_p for the extreme values $p = 1$ and $p = \infty$ for the first 700 generations (averaged over 50 independent runs using population size $N_{\text{pop}} = 60$). In general, a convergent behavior can be observed, which differs, however, for the different values of p: while for $p = 1$ all curves of the indicators values are nearly "smooth," this is not the case for $p = \infty$, where jumps in the indicator values can be observed. The latter is probably due to the (few) outliers NSGA-II has detected time and again (compare to Fig. 1), and/or possibly to the deterioration and cyclic behavior which can occur in the sequence of populations as discussed in [29].

Next, we address the optimality of the NSGA-II approximations. Since the Δ_p value is not known for

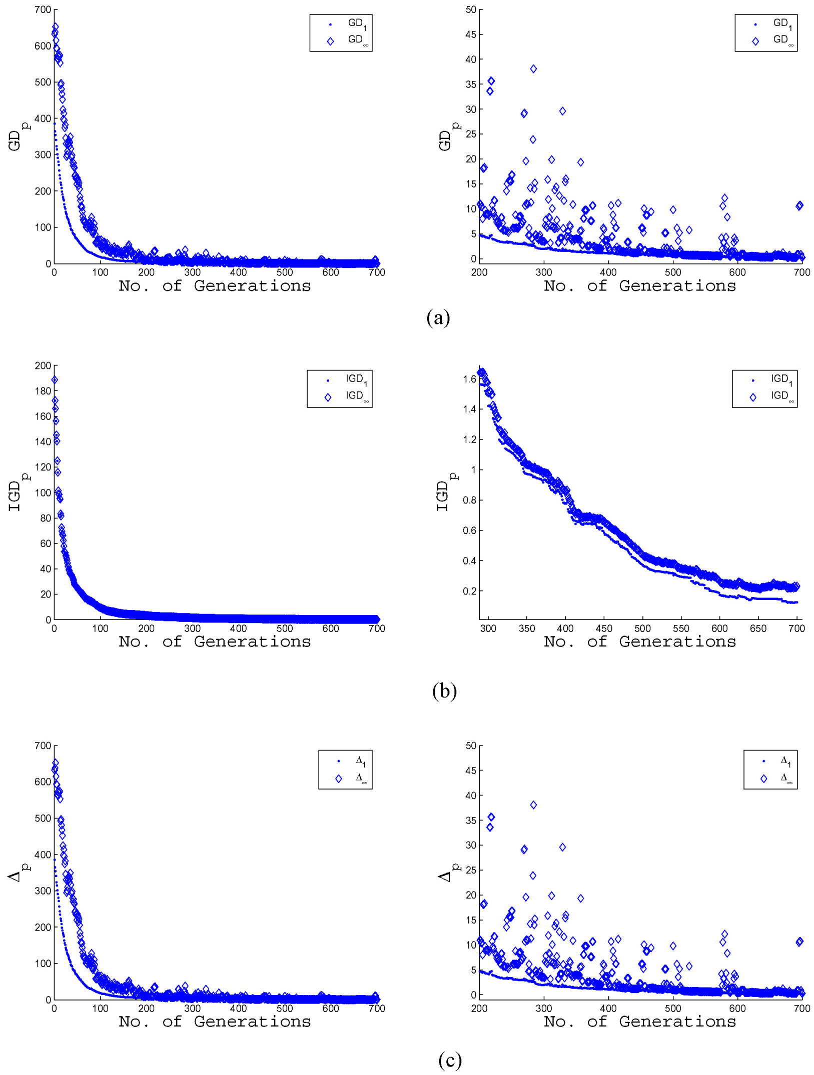


Fig. 9. Numerical results of NSGA-II on the DTLZ1 model, measured by GD_p, IGD_p, and Δ_p for $p = 1$ and $p = \infty$ (compare to Table IV). The results are averaged over 50 independent runs. The left figures show the result of the entire run, and the figures on the right show a zoom. (a) GD_p values. (b) IGD_p values. (c) Δ_p values.

this example (as for Example 2), we have solved numerically the following problem:

$$\min_{x \in \mathbb{R}^{n \times N_{\text{pop}}}} \Delta_p(\{F(x_{(1)}), \ldots, F(x_{(N_{\text{pop}})})\}, F(P_Q)) \tag{69}$$

where $x_{(i)} = (x_{1+(i-1)n}, \ldots, x_{in}) \in \mathbb{R}^n$, $i = 1, \ldots, N_{\text{pop}}$, leading to the approximations of the optimal values

$$\tilde{\Delta}_1 \approx 0.0234, \quad \text{and} \quad \tilde{\Delta}_\infty \approx 0.0514. \tag{70}$$

Hence, the values obtained by NSGA-II are not optimal up to generation 700 (compare to Table IV) which can apart from (70) also be seen that the GD_p values are greater than the IGD_p values. However, since NSGA-II has not been designed to aim for Hausdorff approximations, the algorithm cannot be blamed for that.

It has to be noted that this is just a first attempt to demonstrate the usefulness of the new indicator on a state-of-the-art MOEA. Much further investigation has to be done in this direction.

TABLE IV

NUMERICAL RESULTS OF NSGA-II ON THE DTLZ1 MODEL, MEASURED BY GD_p, IGD_p, AND Δ_p FOR $p = 1$ AND $p = \infty$ (COMPARE TO FIG. 9)

	No. of Generations						
	100	200	300	400	500	600	700
GD_1	18.586	4.682	1.953	1.061	0.670	0.239	0.128
GD_∞	40.051	10.954	3.510	2.205	9791	0.466	0.253
IGD_1	9.327	3.123	1.421	0.778	0.371	0.173	0.124
IGD_∞	9.467	3.217	1.506	0.861	0.438	0.260	0.233
Δ_1	18.586	4.682	1.953	1.061	0.670	0.239	0.128
Δ_∞	40.051	10.954	3.510	2.205	9791	0.466	0.253

C. Evaluation of *ArchiveUpdateTight* Results

In this section, we want to demonstrate that the indicators developed in this paper can be helpful to evaluate the outcome set of evolutionary strategies that are coupled with certain (specialized) archiving strategies. Here, we will investigate the result coming from three different archivers: the archiver that stores all nondominated solutions, and two further ones

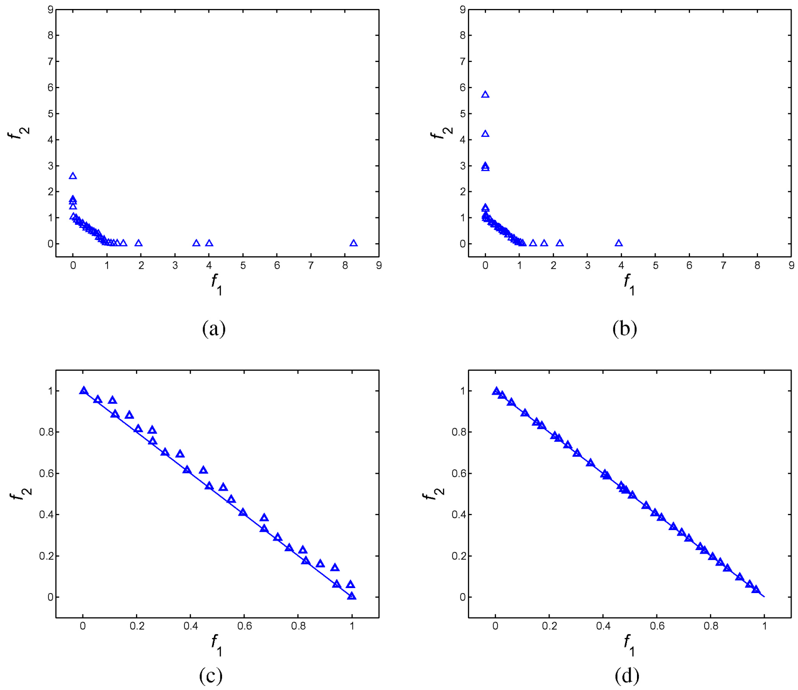


Fig. 10. Numerical results from Tight1 and Tight2 on MOP (75). (a) Tight1, $N = 1e4$. (b) Tight2, $N = 1e4$. (c) Tight1, $N = 1e6$. (d) Tight2, $N = 1e6$.

that aim for particular finite size representations of the Pareto front. We will propose a model where it is likely that an evolutionary strategy traces weakly optimal solutions that are possibly far from the Pareto set, and that the outcome set (i.e., the final archive) can be evaluated more fairly with respect to the occurrence of outliers. We are of the opinion that this can be used in the future to compare the performance of different MOEAs equipped with the same archiver.

The first archiver we consider here, $ArchiveUpdateND$ (short ND), is the one that stores all nondominated solutions obtained by the generation process, that is

$$\begin{aligned} &ArchiveUpdateND(P, A_0) \\ &:= \{x \in P \cup A_0 \ : \ y \nprec x \ \forall y \in P \cup A_0.\} \end{aligned} \tag{71}$$

In [46], it is shown that the archiver generates under certain (mild) assumptions on the generator a sequence of archives A_l, $l \in \mathbb{N}$, such that it holds with probability one

$$\lim_{l \to \infty} d_H(F(A_l), F(P_Q)) = 0. \tag{72}$$

That is, the images of the archives converge to the Pareto front in the Hausdorff sense. The drawback of this archiver—at least for continuous models—is that the magnitudes of the sequence of archives quickly go beyond any given threshold. As a possible remedy, further archives have been proposed that aim for particular finite size representations of the Pareto front, for instance the archivers investigated in [44]. Though the two archivers were developed with different scopes, both can be explained quite well using the distance measurements discussed in this paper.

The first archiver, $ArchiveUpdateTight1$ (short Tight1), is generating a sequence A_l of archives that are aiming to construct a $(\delta, \Theta\epsilon_m)$-tight ϵ-approximate Pareto set, where $\delta \in \mathbb{R}_+$, $\epsilon \in \mathbb{R}^k_+$ are discretization parameters with $\epsilon_m := min_{i=1,\ldots,k}\epsilon_i$, $\epsilon_M := max_{i=1,\ldots,k}\epsilon_i$, and $\Theta \in (0, 1)$ is a safety factor. Though the existence of outliers is not excluded in this set of interest, the underlying idea of such an approximation A_1 is that (at least after removal of the outliers) it holds

$$\begin{aligned} &dist(F(A_1), F(P_Q)) \le \epsilon_M, \quad \text{and} \\ &dist(F(P_Q), F(A_1)) \le \delta. \end{aligned} \tag{73}$$

Since ϵ-approximate solutions are considered to be "good enough" by Tight1, they are not replaced by dominating solutions any more. By this, the uniformity level ϵ_m (i.e., $\|F(a_1) - F(a_2)\|_\infty \ \ge \ \epsilon_m \ \forall a_1, a_2 \in A_1, \ a_1 \neq a_2$) can be guaranteed, but no convergence toward the Pareto front. Hence, the values on the right-hand sides of (73) can be considered to be ideal ones for the resulting archives.

If convergence toward the Pareto front is desired, then the archiver $ArchiveUpdateTight2$ (short Tight2) can be chosen. Tight2 aims for a δ-tight Pareto set, i.e., for an "ideal" approximation A_2 generated by Tight2 it is expected that

$$\begin{aligned} &dist(F(A_2), F(P_Q)) = 0, \quad \text{and} \\ &dist(F(P_Q), F(A_2)) \le \delta. \end{aligned} \tag{74}$$

Hence, unlike the outcome of Tight1, the images of the archive entries have to converge toward the Pareto front (albeit with the price of dropping the uniformity level, see [44] for a thorough discussion).

In [44] and [46], it is shown that all the archivers generate (under certain assumptions on the generator) sequences of archives that converge with probability one to such sets of interest, however, it is not known how fast this happens since this is dependent on the performance of the generation process. To evaluate this, one can in principle use the operators $dist$ and d_H. However, as discussed above, these ones are probably not as "fair" as desired to the occurrence of outliers (this "fairness" is of course depending on the preferences of the

TABLE V
NUMERICAL RESULTS ND

N	Δ_1	GD_1	IGD_1	Δ_2	GD_2	IGD_2	Δ_∞	GD_∞	IGD_∞
1e3	0.0803	0.0791	0.0446	0.1341	0.1331	0.0499	0.4239	0.4194	0.1035
1e4	0.0176	0.0174	0.0140	0.0347	0.0346	0.0158	0.1873	0.1873	0.0361
1e5	0.0047	0.0047	0.0044	0.0071	0.0071	0.0050	0.0581	0.0581	0.0121
1e6	0.0024	0.0024	0.0015	0.0030	0.0030	0.0017	0.0357	0.0357	0.0041

TABLE VI
NUMERICAL RESULTS TIGHT1

N	Δ_1	GD_1	IGD_1	Δ_2	GD_2	IGD_2	Δ_∞	GD_∞	IGD_∞
1e3	0.0973	0.0959	0.0543	0.1582	0.1582	0.0599	0.4957	0.4757	0.1153
1e4	0.0612	0.0605	0.0329	0.1009	0.1005	0.0358	0.3302	0.3292	0.0653
1e5	0.0513	0.0513	0.0284	0.0836	0.0836	0.0312	0.2799	0.2799	0.0562
1e6	0.0508	0.0508	0.0275	0.0798	0.0798	0.0299	0.2156	0.2156	0.0550

TABLE VII
NUMERICAL RESULTS TIGHT2

N	Δ_1	GD_1	IGD_1	Δ_2	GD_2	IGD_2	Δ_∞	GD_∞	IGD_∞
1e3	0.0822	0.0868	0.0448	0.1583	0.1583	0.0495	0.5231	0.5231	0.0986
1e4	0.0243	0.021	0.0223	0.0405	0.0396	0.0249	0.1666	0.1652	0.0543
1e5	0.0150	0.0062	0.0150	0.0191	0.0121	0.0176	0.0569	0.0569	0.0392
1e6	0.0130	0.0024	0.0130	0.0156	0.0029	0.0156	0.0349	0.0088	0.0550

TABLE VIII
MOPS USED IN SECTION V-A

Name	Definition	Constraints
DTLZ1 [9]	$f_1 = 0.5x_1x_2(1+g(x_3))$ $f_2 = 0.5x_1(1-x_2)(1+g(x_3))$ $f_3 = 0.5(1-x_1)(1+g(x_3))$ where $g(x_3) = 100(\lvert x_3\rvert + \sum_{i=1}^{3}(x_i-0.5)^2 - \cos(20\pi(x_i-0.5)))$	$x_i \in [0,1]$
OKA2 [34]	$f_1 = x_1$ $f_2 = 1 - \frac{1}{4\pi^2}(x_1+\pi)^2 + \lvert x_2 - 5\cos(x_1)\rvert^{\frac{1}{3}} + \lvert x_3 - 5\sin(x_1)\rvert^{\frac{1}{3}}$	$x_1 \in [-\pi,\pi]$ $x_2, x_3 \in [-5,5]$
SDD1 [42]	$f_i(x) = \sum_{\substack{j=1\\ j\neq i}}^{n}(x_j - a_j^i)^2 + (x_i - a_i^i)^4,\ i = 1,2,3,$ where $a^1 = (1,1,1,1,\ldots) \in \mathbb{R}^n$ $a^2 = (-1,-1,-1,-1,\ldots) \in \mathbb{R}^n$ $a^3 = (1,-1,1,-1,\ldots) \in \mathbb{R}^n$	None
UF1 [30]	$f_1(x) = x_1 + \frac{2}{J_1}\sum_{j\in J_1}\lvert x_j - sin(6\pi x_1 + j\pi/n)\rvert^2$ $f_2(x) = 1 - \sqrt{x_1} + \frac{2}{J_2}\sum_{j\in J_1}\lvert x_j - sin(6\pi x_1 + j\pi/n)\rvert^2$ where $J_1 = \{j \mid j \text{ is odd and } 2 \leq j \leq n\},\ J_2 = \{j \mid j \text{ is even and } 2 \leq j \leq n\}$	$x_1 \in [0,1]$ $x_i \in [-1,1],\ i = 2,\ldots,n$

algorithm designer and/or the given application). Hence, it might make sense to use the indicators GD_p, IGD_p, and Δ_p instead.

As a test model for the investigation for the determination of the approximation quality we suggest the following MOP:

$$\min_{x\in Q} F(x) = x \tag{75}$$

where $F : \mathbb{R}^k \to \mathbb{R}^k$ and the domain Q is given by

$$Q = \left\{ x \in \mathbb{R}^k \ : \ x_i \in [0,10], i = 1,\ldots,k, \text{ and } \sum_{i=1}^{k} x_i \geq 1 \right\}. \tag{76}$$

Hereby, Pareto set and front are given by the $(k-1)$-standard simplex

$$P_Q = F(P_Q) = \left\{ x \in \mathbb{R}^k : x_i \geq 0, i = 1,\ldots,k, \text{ and } \sum_{k=1}^{k} x_i = 1 \right\}. \tag{77}$$

Though apparently the objectives in MOP (75) are very easy, we have chosen this model for two reasons (that are both induced by the structure of Q): 1) every $x \in \partial Q$ (i.e., the boundary of Q) with $x_i = 0$ for an index $i \in \{1,\ldots,k\}$ is a weak Pareto point, and 2) given $x \in Q$, every vector ν in the non-positive orthant is a descent direction of MOP (75) at x

(i.e., every point $x + tv$ where $t \in \mathbb{R}_+$, dominates x). Hence, it can be expected that weak Pareto points in $\partial Q \backslash P_Q$ will be found easily by a stochastic search algorithm, even if line search methods are involved (e.g., [28]).

In the following, we will use MOP (75) for the bi-objective case (i.e., $k = 2$). Since the aim is to demonstrate the behavior of the indicator values on the sequence of archives and not to compare different algorithms, we have chosen to use random search as the generator: we choose N test points x_i uniformly at random from $[0, 10]^2$ and feed the archiver with the feasible solutions (i.e., if $x_i \in Q$, else x_i is discarded). We have observed that when using random search it is practically impossible to eliminate points that are near to weakly optimal points once they have been detected. Hence, we have chosen to impose the additional constraints to Q in order to reduce (but not eliminate) that problem

$$\begin{aligned} -\alpha + \alpha x_1 - x_2 &\leq 0 \\ -\alpha + \alpha x_2 - x_2 &\leq 0 \end{aligned} \tag{78}$$

where we have chosen $\alpha = 0.01$: the constraints have the effect that the weakly optimal (but not Pareto optimal) points of the original MOP (75) are outside the new domain. A larger value of α leads in general to less outliers in the archive.

VI. Conclusion and Future Work

In this paper, we proposed a new performance indicator, Δ_p, which measures the averaged Hausdorff distance of the image of the outcome set (or final archive) $F(A)$ to the Pareto front $F(P_Q)$ of a given multiobjective optimization problem. Since Δ_p considers the averaged distances between the entries of $F(A)$ and $F(P_Q)$, the novel indicator is in particular interesting for the evaluation of stochastic search algorithms such as multiobjective evolutionary algorithms since such methods tend to generate outliers, and in that case the "classical" Hausdorff distance d_H is entirely determined by the largest outlier (and hence, not always applicable with satisfying results).

To establish Δ_p, we first investigated two widely used indicators in the evolutionary multiobjective optimization community, namely the generational distance and the inverted generational distance. We argued that a slight modification of both operators (i.e., by using the power mean of the considered distances) leads to more "fair" indicators. To be more precise, larger archive sizes (for the modification GD_p of the generational distance), respectively, finer discretizations of the Pareto front (for the modification IGD_p) do not automatically lead to "better" approximations as in their original definitions. This led, in particular, to a better Pareto compliance for GD_p.

Next, we defined Δ_p—analog to d_H—as the maximum of the GD_p and the IGD_p value which defines an averaged Hausdorff distance for $p < \infty$ and coincides with d_H for $p = \infty$. Δ_p offers better metric properties than its components GD_p and IGD_p: it defines a semi-metric for all values of p and is even a pseudo-metric in case the magnitudes of the considered sets are bounded (which is typically the case when considering the outcome sets of evolutionary algorithms). A related topic is the outlier tradeoff which we have addressed next: the lower the value of p, the less single outliers are penalized but the more "far away" Δ_p is to a metric (due to its high probability to violate the triangle inequality). On the other hand, the larger the value of p, the "nearer" Δ_p comes to a metric in the mathematical sense, but, in turn, single outliers get penalized stronger.

Furthermore, we addressed extensions of GD_p, IGD_p, and Δ_p to continuous multiobjective optimization problems. Though the expressions are typically not easy to compute on a general problem, they can be used to bound the discretization error which is certainly of interest when considering discretized Pareto fronts (as usually done in the literature).

Finally, we presented some numerical results that aim to demonstrate the applicability and usefulness of the new indicators.

For the future, there are several aspects worth investigating. For instance, it seems that further theoretical investigations could help for a better understanding of the three indicators such as the influence of the values of p and q. Next, the compliance of Pareto based MOEAs with Δ_p is certainly of major interest. In this paper, we showed that the aim of these algorithms can be described quite well using the Hausdorff distance, however, it is left to investigate: 1) how far this relation goes, and 2) how this can be improved. Finally, it is considerable that the current study can be extended to further sets of interest such as Hausdorff approximations of the Pareto *set* or the family of Pareto sets/fronts of dynamic MOPs.

Appendix A
IGD_p for Continuous Models

To derive (52) for IGD_p, we assume that $k = 2$ and that the Pareto front can be expressed by a curve $\gamma : [m_1, M_1] \subset \mathbb{R} \to \mathbb{R}^2$, where $m_1 := \min_{p \in P_Q} f_1(p)$ and $M_1 := \max_{p \in P_Q} f_1(p)$. Assume first we are given a discretized Pareto front $\tilde{P}_Q = \{\tilde{p}_1, \ldots, \tilde{p}_{|P_Q|}\}$, then

$$IGD_p(F(A), F(\tilde{P}_Q)) = \left(\frac{1}{|\tilde{P}_Q|} \sum_{i=1}^{|\tilde{P}_Q|} dist(F(\tilde{p}_i), F(A))^p \right)^{1/p}. \tag{79}$$

Now we consider (79) using γ: for every point $\tilde{p}_i$ there exists a $\tilde{t}_i \in [m_1, M_1]$ such that $\gamma(\tilde{t}_i) = F(\tilde{p}_i)$, and hence, (79) can be written as

$$IGD_p(F(A), F(\tilde{P}_Q)) = \left(\frac{1}{|\tilde{P}_Q|} \sum_{i=1}^{|\tilde{P}_Q|} dist(\gamma(\tilde{t}_i), F(A)) \right). \tag{80}$$

In the following, we discretize $F(P_Q)$ by choosing samples of the interval $[m_1, M_1]$ which is justified by the above equation. For this, let $[m_1, M_1]$ be subdivided into N subintervals of equal length $\Delta t = (M_1 - m_1)/N$, and choose one t_i in each interval. Then, we obtain for the discretization $P_{Q,N} :=$

$\{\gamma(t_1), \ldots, \gamma(t_N)\}$ the formula

$$
\begin{aligned}
&IGD_p(F(A), F(P_{Q,N})) \\
&\quad = \left(\frac{1}{N} \sum_{i=1}^{N} dist(\gamma(t_i), F(A))^p \right)^{1/p} \\
&\quad = \left(\frac{1}{N \cdot \Delta t} \sum_{i=1}^{N} dist(\gamma(t_i), F(A))^p \cdot \Delta t \right)^{1/p} \\
&\quad = \left(\frac{1}{M_1 - m_1} \sum_{i=1}^{N} dist(\gamma(t_i), F(A)) \cdot \Delta t \right)^{1/p} \qquad (81)
\end{aligned}
$$

i.e., the Riemann sum of $\varphi : [m_1, M_1] \to \mathbb{R}$, $\varphi(t) = dist(\gamma(t), F(A))$, with the given partition. Since we obtain for $N \to \infty$ that $F(P_{Q,N}) \to F(P_Q)$ in the Hausdorff sense and $dist(\cdot, F(A))$ (and hence also φ) is continuous we can define for the limit

$$
\begin{aligned}
&IGD_p(F(A), F(P_Q)) \\
&\quad = \left(\frac{1}{M_1 - m_1} \int_{m_1}^{M_1} dist(\gamma(t), F(A))^p dt \right)^{1/p}. \qquad (82)
\end{aligned}
$$

APPENDIX B
DERIVATION OF (63)

$$
\begin{aligned}
&IGD_p(F(A), F(P_Q)) \\
&= \left(\frac{1}{1-0} \int_0^1 dist\left(\begin{pmatrix} 1 \\ \text{t-1} \end{pmatrix}, \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \right)^p dt \right)^{1/p} \\
&= \left(\int_0^1 \left\| \begin{pmatrix} \text{t-0.5} \\ \text{0.5-t} \end{pmatrix} \right\|_2^p \right)^{1/p} \\
&= \left(\int_0^1 \left(\sqrt{(t-0.5)^2 + (0.5-t)^2} \right)^p dt \right)^{1/p} \\
&= \sqrt{2} \left(\int_0^1 |t - 0.5|^p \right)^{1/p} = \sqrt{2} \left(2 \int_{1/2}^1 (t-0.5)^p \right)^{1/p} \\
&= \sqrt{2}\sqrt[p]{2} \left(\left[\frac{(t-0.5)^{p+1}}{p+1} \right]_{1/2}^1 \right)^{1/p} = \sqrt{2}\sqrt[p]{2} \left(\frac{1}{2} \right)^{\frac{p+1}{p}} \sqrt[p]{\frac{1}{p+1}} \\
&= \frac{1}{\sqrt{2}} \sqrt[p]{\frac{1}{p+1}}. \qquad (83)
\end{aligned}
$$

APPENDIX C
MOPS UNDER CONSIDERATION

In Section V-A, we have used the MOPs which are listed in Table VIII.

REFERENCES

[1] S. S. Askar and A. Tiwari, "Finding exact solutions for multiobjective optimization problems using a symbolic algorithm," in *Proc. 11th Conf. CEC*, 2009, pp. 24–30.

[2] B. Aulbach, M. Rasmussen, and S. Siegmund, "Approximation of attractors of nonautonomous dynamical systems," *Discr. Continuous Dyn. Syst.*, vol. 5, no. 2, pp. 215–238, 2005.

[3] J. Bader, "Hypervolume-based search for multiobjective optimization: Theory and methods," Ph.D. dissertation, Comput. Eng. Netw. Lab., ETH Zurich, Zurich, Switzerland, 2010.

[4] C. A. Coello Coello, "Evolutionary multiobjective optimization: Some current research trends and topics that remain to be explored," *Frontiers Comput. Sci. China*, vol. 3, no. 1, pp. 18–30, 2009.

[5] C. A. Coello Coello and N. Cruz Cortés, "Solving multiobjective optimization problems using an artificial immune system," *Genet. Programming Evolvable Mach.*, vol. 6, no. 2, pp. 163–190, Jun. 2005.

[6] X. Cui, M. Li, and T. Fang, "Study of population diversity of multiobjective evolutionary algorithm based on immune and entropy principles," in *Proc. CEC*, vol. 2. May 2001, pp. 1316–1321.

[7] F. de A. T. de Carvalho, R. M. C. R. de Souza, M. Chavent, and Y. Lechevallier, "Adaptive Hausdorff distances and dynamic clustering of symbolic interval data," *Pattern Recognit. Lett.*, vol. 27, no. 3, pp. 167–179, 2006.

[8] K. Deb, M. Mohan, and S. Mishra, "Evaluating the ϵ-dominated based multiobjective evolutionary algorithm for a quick computation of Pareto-optimal solutions," *Evol. Comput.*, vol. 13, no. 4, pp. 501–525, 2005.

[9] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable test problems for evolutionary multiobjective optimization," in *Evolutionary Multiobjective Optimization. Theoretical Advances and Applications*, A. Abraham, L. Jain, and R. Goldberg, Eds. New York: Springer, 2005, pp. 105–145.

[10] K. Deb and D. E. Goldberg, "An investigation of niche and species formation in genetic function optimization," in *Proc. 3rd Int. Conf. Genet. Algorithms*, Jun. 1989, pp. 42–50.

[11] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA–II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[12] M. Dellnitz and A. Hohmann, "A subdivision algorithm for the computation of unstable manifolds and global attractors," *Numer. Math.*, vol. 75, no. 3, pp. 293–317, 1997.

[13] M. Dellnitz, O. Schütze, and T. Hestermeyer, "Covering Pareto sets by multilevel subdivision techniques," *J. Optimization Theory Applicat.*, vol. 124, no. 1, pp. 113–136, Jan. 2005.

[14] X. Equivel and O. Schuetze, "On the interplay of generator and archiver within archive based multiobjective evolutionary algorithms," in *Proc. Int. Conf. Electr. Eng. Comput. Sci. Autom. Control*, 2010, pp. 375–380.

[15] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 2nd ed. New York: Wiley, 2003.

[16] A. Farhang-Mehr and S. Azarm, "Diversity assessment of Pareto optimal solution sets: An entropy approach," in *Proc. CEC*, vol. 1. May 2002, pp. 723–728.

[17] A. Farhang-Mehr and S. Azarm, "Entropy-based multiobjective genetic algorithm for design optimization," *Struct. Multidisciplinary Optimiz.*, vol. 24, no. 5, pp. 351–361, Nov. 2002.

[18] J. Fliege and B. F. Svaiter, "Steepest descent methods for multicriteria optimization," *Math. Methods Oper. Res.*, vol. 51, no. 3, pp. 479–494, 2000.

[19] D. E. Goldberg and J. Richardson, "Genetic algorithm with sharing for multimodal function optimization," in *Proc. 2nd Int. Conf. Genet. Algorithms Their Applicat.*, 1987, pp. 41–49.

[20] T. Hanne, "Global multiobjective optimization with evolutionary algorithms: Selection mechanisms and mutation control," in *Proc. 1st Int. Conf. Evol. Multi-Criterion Optimiz.*, LNCS 1993. 2001, pp. 197–212.

[21] M. P. Hansen and A. Jaszkiewicz, "Evaluating the quality of approximations to the non-dominated set," Inst. Math. Model., Tech. Univ. Denmark, Copenhagen, Denmark, Tech. Rep. IMM-REP-1998-7, Mar. 1998.

[22] K. Harada, J. Sakuma, S. Kobayashi, and I. Ono, "Uniform sampling of local Pareto-optimal solution curves by Pareto path following and its applications in multiobjective GA," in *Proc. GECCO*, 2007, pp. 813–820.

[23] J. Heinonen, *Lectures on Analysis on Metric Spaces*. New York: Springer, 2001.

[24] D. P. Huttenlocher, G. A. Klanderman, and W. A. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[25] J. Knowles and D. Corne, "On metrics for comparing nondominated sets," in *Proc. CEC*, vol. 1. May 2002, pp. 711–716.

[26] J. D. Knowles and D. W. Corne, "Approximating the nondominated front using the Pareto archived evolution strategy," *Evol. Comput.*, vol. 8, no. 2, pp. 149–172, 2000.

[27] J. D. Knowles and D. W. Corne, "Properties of an adaptive archiving algorithm for storing nondominated vectors," *IEEE Trans. Evol. Comput.*, vol. 7, no. 2, pp. 100–116, Apr. 2003.

[28] A. Lara, G. Sanches, C. A. Coello Coello, and O. Schütze, "HCS: A new local search strategy for memetic multiobjective evolutionary algorithms," *IEEE Trans. Evol. Comput.*, vol. 14, no. 1, pp. 112–132, Feb. 2010.

[29] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multiobjective optimization," *Evol. Comput.*, vol. 10, no. 3, pp. 263–282, 2002.

[30] H. Li and Q. Zhang, "Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 13, no. 2, pp. 284–302, Apr. 2009.

[31] L. Liu, M. Li, and D. Lin, "A novel epsilon-dominance multiobjective evolutionary algorithms for solving DRS multiobjective optimization problems," in *Proc. 3rd ICNC*, Aug. 2007, pp. 96–100.

[32] P. Loridan, "ϵ-solutions in vector minimization problems," *J. Optimization Theory Applicat.*, vol. 43, no. 2, pp. 265–276, 1984.

[33] J. Mehnen, T. Wagner, and G. Rudolph, "Evolutionary optimization of dynamic multiobjective test functions," in *Proc. 2nd Italian Workshop Evol. Comput. (GSICE2)*, 2006.

[34] T. Okabe, Y. Jin, M. Olhofer, and B. Sendhoff, "On test functions for evolutionary multiobjective optimization," in *Proc. PPSN*, 2004, pp. 792–802.

[35] K. Padberg, "Numerical analysis of transport in dynamical systems," Ph.D. dissertation, Fakultät für Elektrotechnik, Informatik und Mathematik, Univ. Paderborn, Paderborn, Germany, 2005.

[36] S. Sayin, "Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming," *Math. Programming*, vol. 87, no. 3, pp. 543–560, 2000.

[37] S. Schäffler, R. Schultz, and K. Weinzierl, "A stochastic method for the solution of unconstrained vector optimization problems," *J. Optimization Theory Applicat.*, vol. 114, no. 1, pp. 209–222, 2002.

[38] J. R. Schott, "Fault tolerant design using single and multicriteria genetic algorithms," Ph.D. dissertation, Dept. Aeronautics Astronautics, Massachusetts Instit. Technol., Cambridge, 1995.

[39] O. Schuetze, X. Equivel, A. Lara, and C. A. Coello Coello, "Some comments on GD and IGD and relations to the Hausdorff distance," in *Proc. 12th Annu. Conf. GECCO*, 2010, pp. 1971–1974.

[40] O. Schütze, "Set oriented methods for global optimization," Ph.D. dissertation, Univ. Paderborn, Paderborn, Germany, 2004 [Online]. Available: http://ubdata.uni-paderborn.de/ediss/17/2004/schuetze

[41] O. Schütze, C. A. Coello Coello, S. Mostaghim, E.-G. Talbi, and M. Dellnitz, "Hybridizing evolutionary strategies with continuation methods for solving multiobjective problems," *Eng. Optimization*, vol. 40, no. 5, pp. 383–402, 2008.

[42] O. Schütze, A. Dell'Aere, and M. Dellnitz, "On continuation methods for the numerical treatment of multiobjective optimization problems," in *Practical Approaches to Multi-Objective Optimization* (Dagstuhl Seminar Proceedings, no. 04461), J. Branke, K. Deb, K. Miettinen, and R. E. Steuer, Eds. Schloss Dagstuhl, Germany: Internationales Begegnungs und Forschungszentrum, 2005 [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2005/349

[43] O. Schütze, M. Laumanns, C. A. Coello Coello, M. Dellnitz, and E.-G. Talbi, "Convergence of stochastic search algorithms to finite size Pareto set approximations," *J. Global Optimization*, vol. 41, no. 4, pp. 559–577, 2008.

[44] O. Schütze, M. Laumanns, E. Tantar, C. A. Coello Coello, and E.-G. Talbi, "Computing gap free Pareto front approximations with stochastic search algorithms," *Evol. Comput.*, vol. 18, no. 1, pp. 65–96, 2010.

[45] O. Schütze, S. Mostaghim, M. Dellnitz, and J. Teich, "Covering Pareto sets by multilevel evolutionary subdivision techniques," in *Evolutionary Multi-Criterion Optimization* (Lecture Notes in Computer Science), C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. Berlin, Germany: Springer, 2003.

[46] O. Schütze, C. A. Coello Coello, S. Mostaghim, E.-G. Talbi, and M. Dellnitz, "Hybridizing evolutionary strategies with continuation methods for solving multiobjective problems," *Eng. Optimization*, vol. 40, no. 5, pp. 383–402, May 2008.

[47] O. Teytaud, "On the hardness of offline multiobjective optimization," *Evol. Comput.*, vol. 15, no. 4, pp. 475–491, 2007.

[48] G. Toscano Pulido and C. A. Coello Coello, "Using clustering techniques to improve the performance of a particle swarm optimizer," in *Proc. GECCO I*, LNCS 3102. Jun. 2004, pp. 225–237.

[49] D. A. Van Veldhuizen, "Multiobjective evolutionary algorithms: Classifications, analyses, and new innovations," Ph.D. dissertation, Dept. Electr. Comput. Eng., Graduate School Eng., Air Force Instit. Technol., Wright-Patterson AFB, OH, May 1999.

[50] X. Yi and O. I. Camps, "Line-based recognition using a multidimensional Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 9, pp. 901–916, Sep. 1999.

[51] E. Zitzler, "Evolutionary algorithms for multiobjective optimization: Methods and applications," Ph.D. dissertation, Comput. Eng. Netw. Lab., ETH Zurich, Zurich, Switzerland, 1999.

[52] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evol. Comput.*, vol. 8, no. 2, pp. 173–195, Jun. 2000.

[53] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach," *IEEE Trans. Evol. Comput.*, vol. 3, no. 4, pp. 257–271, Nov. 1999.

[54] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Trans. Evol. Comput.*, vol. 7, no. 2, pp. 117–132, Apr. 2003.

Oliver Schütze received the Diploma degree in mathematics from the University of Bayreuth, Bayreuth, Germany, and the Ph.D. degree in natural sciences from the University of Paderborn, Paderborn, Germany, in 1999 and 2004, respectively.

He is currently a Professor (CINVESTAV-3A Researcher) with the Department of Computer Science, CINVESTAV-IPN, Mexico City, Mexico.

Xavier Esquivel received the B.Sc. degree in computer science from the National Polytechnic Institute of Mexico, Mexico City, Mexico, in 2005, and the M.Sc. degree in computer science from the Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN), Mexico City, in 2010.

He is currently a Software Developer with the Oracle Mexico Development Center, Zapopan Jalisco, Mexico.

Adriana Lara received the B.Sc. degree in physics and mathematics from the National Polytechnic Institute of Mexico, Mexico City, in 2001, and the M.Sc. degree in electrical and computer engineering from the Centro de Investigación y Estudios Avanzados-Instituto Politécnico Nacional (CINVESTAV-IPN), Mexico City, in 2003. She is currently pursuing the Ph.D. degree in computer sciences with the Department of Computer Science, CINVESTAV-IPN.

She is a Faculty Member with the Mathematics Department of ESFM, National Polytechnic Institute of Mexico. Her current research interests include multiobjective optimization, gradient-based techniques, and hybrid algorithms.

Carlos A. Coello Coello (F'11) received the B.Sc. degree in civil engineering from the Universidad Autónoma de Chiapas, Chiapas, Mexico, and the M.Sc. and Ph.D. degrees in computer science from Tulane University, New Orleans, LA, in 1991, 1993, and 1996, respectively.

He is currently a Professor (CINVESTAV-3F Researcher) and the Chair of the Department of Computer Science, Centro de Investigación y Estudios Avanzados-Instituto Politécnico Nacional (CINVESTAV-IPN), Mexico City, Mexico. He is also with Unité Mixte Internationale LAFMIA 3175 CNRS, CINVESTAV-IPN, Mexico City. He has authored and co-authored over 300 technical papers and several book chapters. He co-authored the book *Evolutionary Algorithms for Solving Multi-Objective Problems* (Berlin, Germany: Springer, 2nd ed., 2007). His publications currently report over 5100 citations. His current research interests include evolutionary multiobjective optimization and constraint-handling techniques for evolutionary algorithms.

Dr. Coello Coello is currently an Associate Editor of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION and serves in the editorial boards of 12 other international journals. He chairs the Evolutionary Computation Technical Committee of the IEEE Computational Intelligence Society. He received the 2007 National Research Award from the Mexican Academy of Sciences in the area of exact sciences and the Medal to the Scientific Merit 2009, granted by Mexico City's congressmen. He is a member of the ACM, Sigma Xi, and the Mexican Academy of Science.