

On the Pareto Compliance of the Averaged Hausdorff Distance as a Performance Indicator

Andrés Vargas^{1,*}

Edited by

Juan Carlos Salcedo-Reyes
(salcedo.juan@javeriana.edu.co)

1. Departamento de Matemáticas,
Pontificia Universidad Javeriana,
Bogotá, Colombia.

* a.vargasd@javeriana.edu.co

Received: 20-08-2018

Accepted: 27-09-2018

Published on line: 28-09-2018

Citation: Vargas A. On the Pareto Compliance of the Averaged Hausdorff Distance as a Performance Indicator, *Universitas Scientiarum*, 23 (3): 333-354, 2018. doi: 10.11144/Javeriana.SC23-3.otpc

Funding:

N.A.

Electronic supplementary material:
N.A.



Abstract

The averaged Hausdorff distance Δ_p is an inframetric, recently introduced in evolutionary multiobjective optimization (EMO) as a tool to measure the optimality of finite size approximations to the Pareto front associated to a multiobjective optimization problem (MOP). Tools of this kind are called performance indicators, and their quality depends on the useful criteria they provide to evaluate the suitability of different candidate solutions to a given MOP.

We present here a purely theoretical study of the compliance of the Δ_p -indicator to the notion of Pareto optimality. Since Δ_p is defined in terms of a modified version of other well-known indicators, namely the generational distance GD_p , and the inverted generational distance IGD_p , specific criteria for the Pareto compliance of each one of them is discussed in detail. In doing so, we review some previously available knowledge on the behavior of these indicators, correcting inaccuracies found in the literature, and establish new and more general results, including detailed proofs and examples of illustrative situations.

Keywords: averaged Hausdorff distance; generational distance; inverted generational distance; multiobjective optimization; Pareto optimality; performance indicator.

Introduction

A fundamental task in evolutionary multiobjective optimization (EMO) consists in the explicit computation of the set of solutions (known as the *Pareto set*) and their images (the *Pareto front*) corresponding to the problem of simultaneous optimization of multiple objective functions, or multiobjective optimization problem (MOP), for short. It is an important fact (see, e.g. [7]) that every non-trivial MOP admits more than one solution, i.e., there is no single point that simultaneously optimizes all the objective functions. A solution is called *Pareto optimal* if there is no objective function that can be improved without degrading the rest. Even though the Pareto set P of a MOP,

consisting of the set of all optimal solutions, turns out to be a compact subset of \mathbb{R}^n in common situations, generally it cannot be calculated in a purely analytical way, and the use of numerical algorithms becomes essential.

It is often desirable (and even necessary) to approximate P with a subset $A \subset \mathbb{R}^n$, called *archive*, that resembles P and its properties as closely as possible. Archives are usually assumed to consist of a finite number of points that can be numerically found, and EMO algorithms are an important tool employed for achieving that aim. To measure their accuracy, the distance between an outcome archive A and the original Pareto set P should be defined in an appropriate sense, but several inequivalent notions of distance can be considered, and their values are not necessarily attained in a unique manner. This means that the set of candidate approximations is usually not unique.

A readily available notion of distance between sets that can be used in this setting is the Hausdorff distance (see, e.g. [4]), but due to its definition, it allows for undesirable ambiguities, and heavily punishes single outlier solutions. Alternative performance indicators have been introduced in the literature (see, e.g. [10]) and among them, the averaged Hausdorff distance Δ_p was recently proposed in [9] by modifying the well-known generational (GD) and inverted generational (IGD) indicators, in such a way that their values correspond to useful averages. As a result, Δ_p does not punish individual but collective behavior, fixing some drawbacks of the standard Hausdorff distance. Other properties of Δ_p have been investigated in the literature, for example, from the theoretical side that we are interested here, explicit analytical calculations of optimal archives with Δ_p for particular Pareto fronts have been obtained in [8].

In this work, we establish conditions ensuring the Pareto compliance of the GD_p and IGD_p indicators by means of mathematical criteria involving the behavior of candidate solutions, and summarize at the end the consequences for the compliance of the averaged Hausdorff distance. The proofs require only simple properties of Δ_p (or the intermediate GD_p and IGD_p) derived from their definitions, which will be recalled in the section on preliminaries.

It is expected that these criteria for compliance with Pareto optimality can help to elucidate advantages and possible drawbacks of Δ_p as a performance indicator for the evaluation of MOEAs. This is a necessary step in view of the possibility to modify and generalize the averaged Hausdorff distance with the purpose of enhancing its usefulness for applications. In fact, promising generalizations are already appearing in the literature for both, the cases of finite and continuous approximations (see [12] and [1], respectively). Further

work in this direction requires a detailed understanding of the behavior of the original p -indicators, and the treatment presented here reveals also relevant aspects that should be considered.

This paper is organized as follows: In Section 2, we briefly present the background and notation required for the understanding of the rest of the manuscript. The core of this work appears in Section 3, where different criteria for the Pareto compliance of all the indicators, GD_p , IGD_p and Δ_p are provided, including complete proofs, particular examples and important observations. Finally, conclusions and perspectives for future research are pointed out in Section 4.

Preliminaries

Throughout the document we will employ the abbreviations $\mathbb{R}^* := \mathbb{R} \setminus \{0\}$ and $\mathbb{R}_+ := [0, \infty)$ whenever necessary.

Multiobjective Optimization Given a *decision space* $X \subset \mathbb{R}^n$ and a vector valued function $F: X \subset \mathbb{R}^n \rightarrow \mathbb{R}^k$ defined on it, a multiobjective optimization problem consists in the simultaneous minimization of its k components f_1, \dots, f_k . A solution is called *Pareto-optimal* when the elements of the image, or *objective space*, $Y = F(X)$ are *non-dominated* in the sense of Pareto [7]. This notion is defined in terms of a partial order relation that in this context and for our purposes can be introduced directly on $X \subset \mathbb{R}^n$, as follows.

Definition 2.1. For $x, x' \in X$ the *Pareto partial order* \preceq associated to F is defined by

$$x \preceq x' \text{ if and only if } f_i(x) \leq f_i(x') \text{ for all } i = 1, \dots, k.$$

Additionally:

1. An element $x \in X$ is said to be *dominated by* $x' \in X$ and denoted $x' \prec x$, if $x' \preceq x$ and $f(x) \neq f(x')$.
2. An element $x \in X$ is *dominated by* $A \subset X$, written $A \preceq x$, if there exists some $a \in A$ such that $a \preceq x$, otherwise it is said to be *non-dominated* by it, $A \not\preceq x$.
3. A subset $A \subset X$ is *dominated* by a subset $B \subset X$, and written $B \preceq A$, if for every $a \in A$ there exist some $b \in B$ such that $b \preceq a$. If this is not the case A is said to be *non-dominated* by B and denoted $B \not\preceq A$.
4. An element $x \in X$ is called *Pareto-optimal* if it is non-dominated, i.e., $\nexists x' \in X: x' \prec x$, and the set P , of all Pareto-optimal points is called the *Pareto set*.

5. The *Pareto front* is defined as the image $F(P)$ of the Pareto set $P \subset X$.

Throughout the rest of this document we make use of the following handy abbreviations

Notation 2.2. For $A, B \subset X$ and an arbitrary $x \in X$ we define the subsets:

- $A_{\preceq x} := \{a \in A \mid a \preceq x\}$, and $A_{\not\preceq x} := \{a \in A \mid a \not\preceq x\}$.
- $A_{\preceq B} := \{a \in A \mid \exists b \in B : a \preceq b\}$, and $A_{\not\preceq B} := \{a \in A \mid \nexists b \in B : a \preceq b\}$.

Clearly for any $x \in X$, or any $B \subset X$ we have that: $A = A_{\preceq x} \sqcup A_{\not\preceq x}$, and $A = A_{\preceq B} \sqcup A_{\not\preceq B}$, where \sqcup denotes a disjoint union. Similar notations can be used in terms of \prec , \succeq , and \succ , according to the needs of the situation.

Typically, we want to approximate P with a set of the form $F(A)$ where $A \subset X$ is an appropriate subset called *archive*, which is often assumed to be finite. More precisely,

Definition 2.3. A subset $A \subset X$ will be called an *approximation set* or *archive* if it consists only of mutually non-dominated points. Equivalently, $\forall a \in A, \forall a' \in A: a \preceq a' \implies a = a'$.

In this work we are interested only in finite archives. Those are the ones for which the definition of the (modified) Generational Distance GD_p defined below, makes sense.

Now, we introduce a set of minimal conditions that will be assumed to hold in all the forthcoming results. They are part of the criteria required to ensure the compliance of all indicators with Pareto optimality. Given their omnipresence, they will be collectively employed to define a notion that we call *well-dominance*.

Definition 2.4. For two archives $A, B \subset X$ we say that A is *well-dominated* by B if

1. A is dominated by B , written $B \preceq A$, i.e., $\forall a \in A, \exists b \in B$ s.t. $b \preceq a$, and
2. B contains only dominating points, i.e., $\forall b \in B, \exists a \in A$ s.t. $b \preceq a$.

If, in addition,

3. $\exists b \in B \setminus A, \exists a \in A \setminus B$ such that $b \prec a$,

we will say that A is *strictly well-dominated* by B .

The Averaged Hausdorff Distance Let (X, d) denote a general metric space X carrying a distance function, or metric, $d : X \times X \rightarrow \mathbb{R}_+$, satisfying the standard properties of non-negativity with identity of indiscernibles, symmetry, and the triangle inequality.

Definition 2.5. For $x \in X$ and arbitrary $A, B \subset X$, the *Hausdorff distance* $d_H(\cdot, \cdot)$ is defined by extending d to subsets of X through the following steps:

1. A distance between points and sets: $d(x, A) := \inf \{d(x, a) \mid a \in A\}$.
2. A pre-distance between sets: $d(B, A) := \sup \{d(b, A) \mid b \in B\}$.
3. The Hausdorff distance between sets: $d_H(A, B) := \max \{d(A, B), d(B, A)\}$.

It is well-known that endowed with d_H , the set $\mathcal{K}(X) \subset \mathcal{P}(X)$ of all non-empty compact subsets of X turns into a metric space itself. Moreover, $(\mathcal{K}(M), d_H)$ is a complete metric space if (X, d) is also complete (cf. [4]).

In our context, the metric space under consideration will always be a subset $X \subset \mathbb{R}^n$ (or \mathbb{R}^k) endowed with the Euclidean distance $d(x, x') := \|x - x'\|$ induced by its standard 2-norm.

Definition 2.6. For $p \in \mathbb{N}$, and finite subsets $A, B \subset \mathbb{R}^n$ the (modified) *Generational Distance* between them is given by

$$\text{GD}_p(A, B) := \left(\frac{1}{|A|} \sum_{a \in A} d(a, B)^p \right)^{\frac{1}{p}},$$

while the (modified) *Inverted Generational Distance* is

$$\text{IGD}_p(A, B) := \left(\frac{1}{|B|} \sum_{b \in B} d(b, A)^p \right)^{\frac{1}{p}}.$$

Finally, their *Averaged Hausdorff Distance* is defined as

$$\Delta_p(A, B) := \max \{ \text{GD}_p(A, B), \text{IGD}_p(A, B) \}.$$

The indicator Δ_p can be viewed as a composition of slight variations of the Generational Distance (GD, see [13]) and the Inverted Generational Distance (IGD, see [2]). From Δ_p the standard Hausdorff distance can be obtained as $\lim_{p \rightarrow \infty} \Delta_p = d_H$, but for finite values of p the indicator Δ_p returns a p -power mean or p -average of the distances considered for d_H .

Performance Indicators Let $Y = F(X) \subset \mathbb{R}^k$ be the objective space of a MOP. In this context, a *performance indicator* is a function $\mathcal{I} : \mathcal{P}(Y) \rightarrow \mathbb{R}_+$ used to measure the suitability of solution sets. In standard terminology, such an indicator is said to be *Pareto-compliant* if for subsets $F(A), F(B) \subset Y$

the strict dominance condition $A \preceq B$ and $B \not\preceq A$ implies the inequality $\mathcal{J}(F(A)) \leq \mathcal{J}(F(B))$, or a stronger sense when it implies $\mathcal{J}(F(A)) < \mathcal{J}(F(B))$. Additional details can be found, e.g., in [15] (see also [3] and [14]).

If $A \subset X$ denotes a candidate archive, the explicit values of the GD_p , IGD_p , and Δ_p -performance indicators assigned to A are given, respectively, by

$$\begin{aligned} \mathcal{J}_p^{\text{GD}}(A) &:= \text{GD}_p(F(A), F(P)), \quad \mathcal{J}_p^{\text{IGD}}(A) := \text{IGD}_p(F(A), F(P)), \quad \text{and} \\ \mathcal{J}_p^{\Delta}(A) &:= \Delta_p(F(A), F(P)). \end{aligned}$$

Remark. We will calculate the GD_p and IGD_p distances between subsets of the objective space $Y = F(X)$ that are images through F of subsets $A \subset X$ which, by definition, are assumed to consist of mutually non-dominated points (like archives or the Pareto set itself). This implies that the restriction $F|_A: A \rightarrow F(A)$ is injective, thus a bijection onto its image. Since $|A| = |F(A)|$, in the finite case the elements of $F(A)$ can be labelled by those of its preimage A in decision space $X \subset \mathbb{R}^n$, and all sums will appear over subsets of X .

In practice, additional assumptions on the sets of comparable archives may be necessary for the required inequalities of Pareto-compliance to hold, and even knowledge on the behavior of the underlying Pareto set P , or its front $F(P)$, can be necessary. Although, in real applications, information on the Pareto set is not usually obtainable, from a purely theoretic point of view those assumptions can help to determine the usefulness of an indicator by elucidating its behavior, and by comparison with known examples.

Compliance to Pareto Optimality

For the Generational Distance We begin by recalling a useful property of GD_p -distances proved in [9], and valid for pointwise solutions under simple hypotheses on the Pareto front. Variations of this property will appear as assumptions in the statement of further results on GD_p -values for arbitrary finite sets. For calculating the GD_p -indicator it is not necessary to assume that P is finite, it can be infinite, continuous, or even piecewise continuous. If necessary, it will be assumed to be compact where explicitly stated.

Lemma 3.1. *Let $a, b \in X \subset \mathbb{R}^n$ and suppose that the Pareto front $F(P) \subset \mathbb{R}^k$ is connected, or that for each $i = 1, \dots, k$ there is an element $y^{(i)} \in F(P)$ such that the components $y_j^{(i)} = f_j(b)$, for all $j \neq i$. Then:*

$$\forall a \in A, \forall b \in B: b \prec a \implies \text{GD}_p(F(b), F(P)) < \text{GD}_p(F(a), F(P)).$$

Proof. The proofs are direct consequences of [9, Prop. 1, Prop. 2]. \square

Briefly speaking, the property established by Lemma 3.1 says that if an archive B can be obtained from another archive A by only replacing one dominated solution $a \in A$ with a dominating one $b \in B$, the value of the GD_p indicator decreases. Namely, if for archives $B := \{b, x_2, \dots, x_n\}$ and $A := \{a, x_2, \dots, x_n\}$, we have $b \prec a$, then $GD_p(F(A), F(P)) < GD_p(F(B), F(P))$, i.e., the Pareto compliance holds when comparing individual solutions.

In [9, Prop. 3] the Pareto compliance of the Generational Distance GD_p is also considered. In particular, it is stated that for finite archives $A, B \subset \mathbb{R}^n$, where A is well-dominated by B , the extra assumption

$$\forall a \in A, \forall b \in B: b \prec a \implies d(F(b), F(P)) < d(F(a), F(P)),$$

suffices to ensure that $GD_p(F(B), F(P)) < GD_p(F(A), F(P))$. Unfortunately, the last step of the proof is not correct under the generality of the original assumptions. In fact, scenarios shown in Example 3.2 and Figure 1 indicate that the conclusion does not always hold. This is because the behavior of the p -average distance for a dominating archive B , could be worst than the corresponding p -average for a dominated set A having more points in its image close to $F(P)$ than those of B . A remedy to those situations is introduced in Theorem 3.4 by imposing assumptions on the sizes of some dominated/dominating sets of points.

Example 3.2. A simple case violating [9, Prop. 3]:

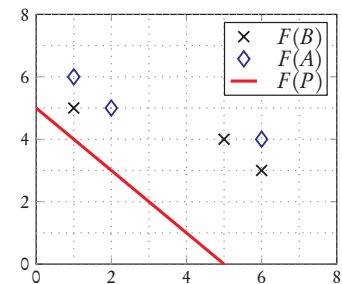
Pareto Front $F(P) :=$ segment from $(0, 5)$ to $(5, 0)$,

Dominated Archive $F(A) := \{(1, 6), (2, 5), (6, 4)\}$,

Dominating Archive $F(B) := \{(1, 5), (5, 4), (6, 3)\}$,

Smaller Indicator $GD_2(F(A), F(P)) = \sqrt{19.3}$,

Larger Indicator $GD_2(F(B), F(P)) = \sqrt{22}$.



Nevertheless, a slight modification (regarding the uniqueness of the elements whose existence is ensured by the hypotheses) already makes the conclusion of the original proposition true. The complete corrected statement is a consequence of the more general result stated in Theorem 3.4 and reads as follows

Proposition 3.3. *Let $A, B \subset \mathbb{R}^n$ be finite archives such that:*

1. $\forall a \in A, \exists! b \in B$ such that $b \preceq a$;
2. $\forall b \in B, \exists! a \in A$ such that $b \preceq a$;
3. $\exists b \in B \setminus A, \exists a \in A \setminus B$ such that $b \prec a$;
4. $\forall a \in A, \forall b \in B: b \prec a \implies d(F(b), F(P)) < d(F(a), F(P))$.

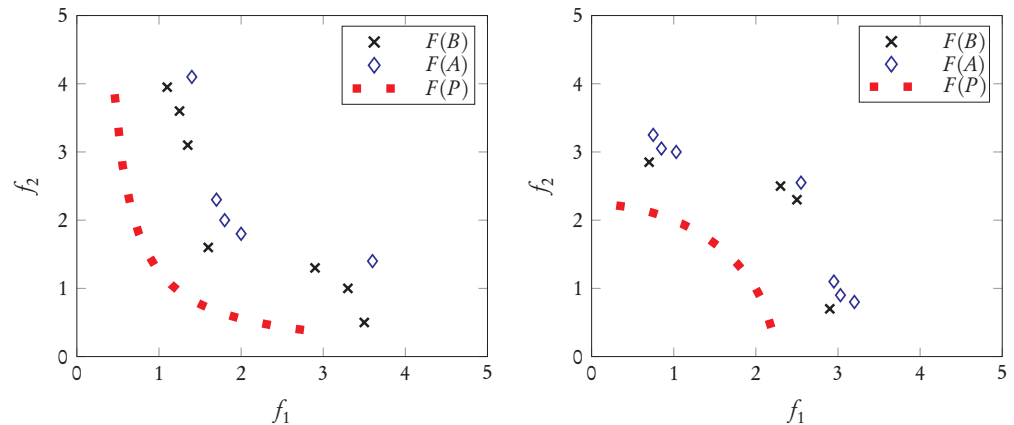


Figure 1. Two different situations where the GD_p value of archive A turns out to be better (smaller) than the GD_p value of archive B , even though $B \preceq A$ and the original assumptions of Prop. 3 of [9] are satisfied. Notice the number and distribution of dominated/dominating points in relation to $F(P)$.

Then, $GD_p(F(B), F(P)) < GD_p(F(A), F(P))$.

Proof. It is a simple consequence of Theorem 3.4 below. \square

Figure 2 illustrates two examples where the conditions of Proposition 3.3 hold, and therefore, the Pareto-compliance conclusion applies. In both cases, each element of B dominates only one element of A , and each element of A is dominated by only one of B , so that the archives are equal sized. In these examples the assumptions of the original statement in [9, Prop. 3] hold.

Notice that the only difference between Proposition 3.3 and [9, Prop. 3] is that conditions 1 and 2 ask for the existence of unique dominating ($b \in B$) and dominated ($a \in A$) elements, respectively. Without these clarifications the conclusion can easily become false, but with them there is still the problem that the assumptions turn out to be unnecessarily restrictive. In fact, they imply that the archives A and B have the same size. However, simple cases already presented in [9] and shown in Figure 3 suggest that it should be possible to keep the conclusion under milder conditions.

Our next result provides a general statement from which Proposition 3.3 follows as the simplest non-trivial case.

Theorem 3.4. *Let $A, B \subset \mathbb{R}^n$ be finite and strictly well-dominated archives with $B \preceq A$, such that*

1. $\forall a \in A, \forall b \in B: b \prec a \implies d(F(b), F(P)) < d(F(a), F(P))$;
2. $\forall a \in A, \forall b \in B: b \preceq a \implies \frac{|B_{\preceq a}|}{|B|} = \frac{|A_{\succeq b}|}{|A|}$.

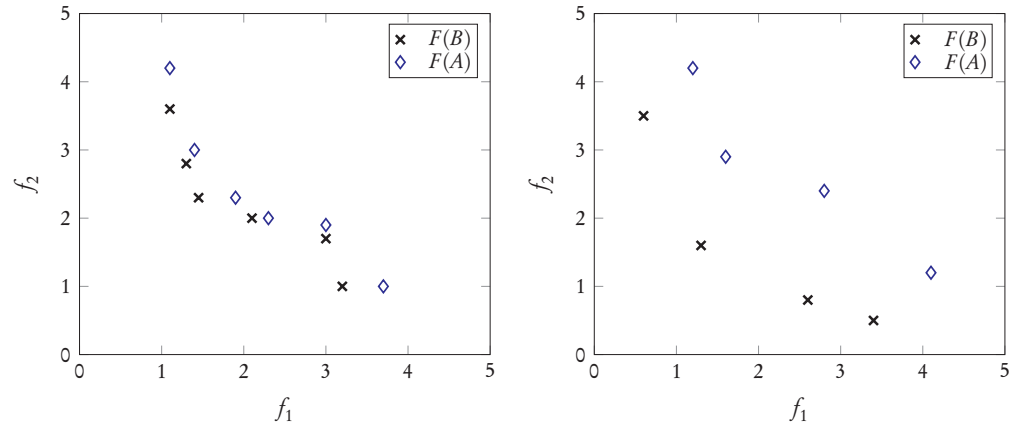


Figure 2. Two different situations where $B \preceq A$ and the adjusted assumptions of Proposition 3.3 are valid. In these cases the GD_p value of archive B is better (smaller) than the GD_p value of archive A .

Then, $GD_p(F(B), F(P)) < GD_p(F(A), F(P))$.

Proof. By condition 1, for arbitrary $a \in A$ and $b \in B_{\preceq a}$ we have the inequality $d(F(b), F(P))^p \leq d(F(a), F(P))^p$. After taking the average over all $b \in B_{\preceq a}$ at the left-hand side, we obtain

$$\frac{1}{|B_{\preceq a}|} \sum_{b \in B_{\preceq a}} d(F(b), F(P))^p \leq d(F(a), F(P))^p,$$

and averaging both sides again over all $a \in A$ yields

$$\frac{1}{|A|} \sum_{a \in A} \left(\frac{1}{|B_{\preceq a}|} \sum_{b \in B_{\preceq a}} d(F(b), F(P))^p \right) \leq \frac{1}{|A|} \sum_{a \in A} d(F(a), F(P))^p. \tag{1}$$

From property 2 we can write the left-hand side as

$$\begin{aligned} \frac{1}{|A|} \sum_{a \in A} \sum_{b \in B_{\preceq a}} \frac{1}{|B_{\preceq a}|} d(F(b), F(P))^p &= \frac{1}{|A|} \sum_{a \in A} \sum_{b \in B_{\preceq a}} \frac{|A|}{|B| |A_{\succeq b}|} d(F(b), F(P))^p \\ &= \frac{1}{|B|} \sum_{a \in A} \sum_{b \in B_{\preceq a}} \frac{1}{|A_{\succeq b}|} d(F(b), F(P))^p. \end{aligned} \tag{2}$$

Now, realizing that each $b \in B$ will appear $|A_{\succeq b}|$ -times in the total it becomes

$$\begin{aligned} &= \frac{1}{|B|} \sum_{b \in B} \frac{|A_{\succeq b}|}{|A_{\succeq b}|} d(F(b), F(P))^p \\ &= \frac{1}{|B|} \sum_{b \in B} d(F(b), F(P))^p. \end{aligned}$$

Returning to (1) we conclude that $\text{GD}_p(F(B), F(P))^p \leq \text{GD}_p(F(A), F(P))^p$. Finally, property 3 of Definition 2.4 for strictly well-dominated sets ensures that the inequality has to be strict, and the claim follows. \square

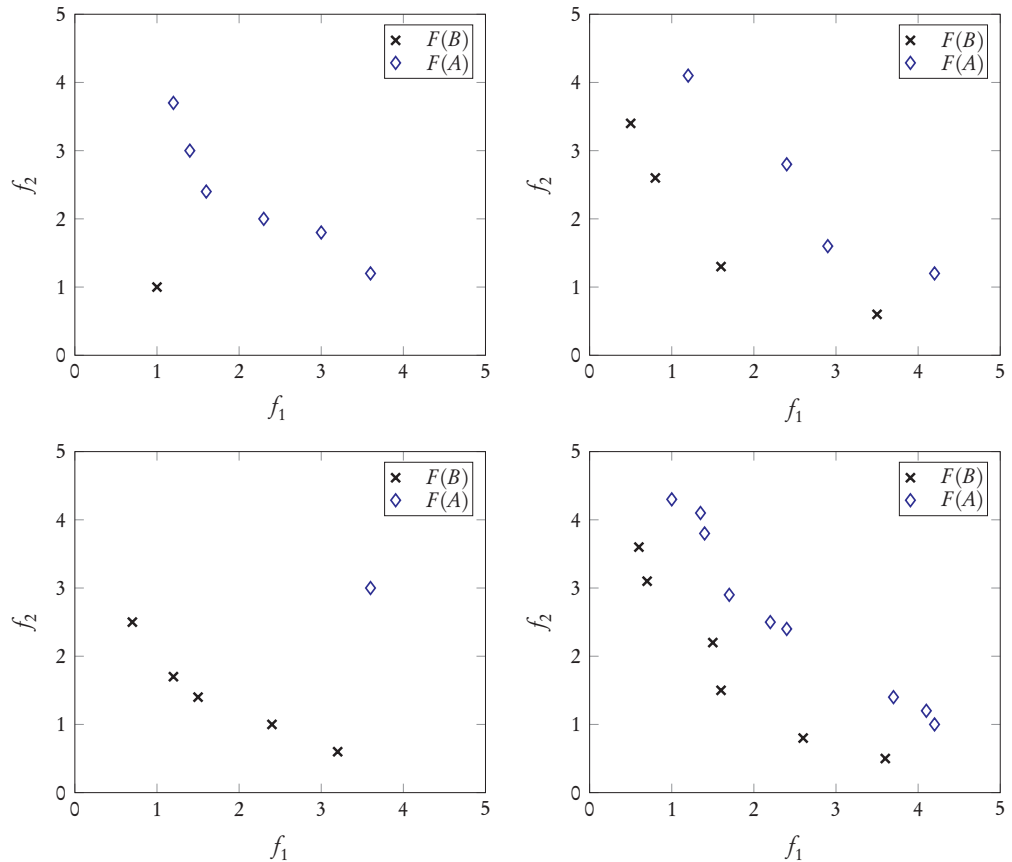


Figure 3. Four different scenarios where the GD_p value of archive B is better (smaller) than the GD_p value of archive A independently of the Pareto front $F(P)$, and where the additional assumptions made in Theorem 3.4 are easily verifiable.

From the step leading to (2) in the proof of Theorem 3.4 it is apparent that the conclusion is still valid if we change the “=” sign between the involved ratios in condition 2 by “ \leq ”, because then $|B_{\leq a}|^{-1} \leq \frac{|A|}{|B|} |A_{\geq b}|^{-1}$, and (2) becomes a “ \leq ”-inequality. Nevertheless, the following result asserts that there is no difference in making that change.

Lemma 3.5. *Let $A, B \subset \mathbb{R}^n$ be finite archives. Then, assuming all the involved sets are non-empty, the following conditions are equivalent:*

1. $\forall a \in A, \forall b \in B: b \preceq a \implies \frac{|B_{\leq a}|}{|B|} = \frac{|A_{\geq b}|}{|A|}$,
2. $\forall a \in A, \forall b \in B: b \preceq a \implies \frac{|B_{\leq a}|}{|B|} \leq \frac{|A_{\geq b}|}{|A|}$.

Proof. Clearly 1 implies 2. For the converse, using 2 with any $b \preceq a$ we have the inequality

$$\frac{|A|}{|A_{\succeq b}|} \leq \frac{|B|}{|B_{\preceq a}|}. \quad (3)$$

It is clear that if at both sides of (3) we take the sum over all $a \in A$ and $b \in B$ for which $b \preceq a$ we find that,

$$|A| \sum_{b \preceq a} \frac{1}{|A_{\succeq b}|} \leq |B| \sum_{b \preceq a} \frac{1}{|B_{\preceq a}|}, \quad (4)$$

where the left-hand side of (4) can be rewritten as

$$|A| \sum_{b \preceq a} \frac{1}{|A_{\succeq b}|} = |A| \sum_{a \in A} \sum_{b \in B_{\preceq a}} \frac{1}{|A_{\succeq b}|} = |A| \sum_{b \in B} \frac{|A_{\succeq b}|}{|A_{\succeq b}|} = |A| |B|,$$

and similarly for the right-hand side we have

$$|B| \sum_{b \preceq a} \frac{1}{|B_{\preceq a}|} = |B| \sum_{b \in B} \sum_{a \in A_{\succeq b}} \frac{1}{|B_{\preceq a}|} = |B| \sum_{a \in A} \frac{|B_{\preceq a}|}{|B_{\preceq a}|} = |B| |A|.$$

Hence, both sides of (4) are actually equal. This could not have been the case if for some $b \preceq a$ the inequality (3) were strict, i.e., it has to be always an equality. This implies condition 1. \square

We end our consideration of the GD_p indicator with an alternative form in which the original statement in [9, Prop. 3] can be fixed without asking for unique dominated/dominating elements. The drawback is that it requires previous knowledge on the behavior of the indicator for non-dominating subsets of the dominated archive A .

Proposition 3.6. *Let $A, B \subset \mathbb{R}^n$ be finite archives such that:*

1. $\forall a \in A, \exists b \in B: b \preceq a$;
2. $\forall b \in B, \exists a \in A: b \preceq a$, and $d(F(a), F(P)) \leq \text{GD}_p(F(A_{\not\preceq b}), F(P))$;
3. $\exists b \in B \setminus A, \exists a \in A \setminus B: b \prec a$;
4. $\forall a \in A, \forall b \in B: b \prec a \implies d(F(b), F(P)) < d(F(a), F(P))$.

Then, $\text{GD}_p(F(B), F(P)) < \text{GD}_p(F(A), F(P))$.

Proof. For any $b \in B$, it follows from condition 4 by taking p -averages over all elements $a \in A_{\not\preceq b}$ that the inequality $d(F(b), F(P))^p \leq \text{GD}_p(F(A_{\not\preceq b}), F(P))^p$ holds. Moreover, for any $a \in A_{\succeq b}$ we know using conditions 2 and 4 that

$$d(F(b), F(P))^p \leq d(F(a), F(P))^p \leq \text{GD}_p(F(A_{\not\preceq b}), F(P))^p.$$

Since $A = A_{\neq b} \sqcup A_{\geq b}$, adding the previous inequalities shows that for any $b \in B$,

$$\begin{aligned} (|A_{\neq b}| + |A_{\geq b}|)d(F(b), F(P))^p &\leq \sum_{a \in A_{\neq b}} d(F(a), F(P))^p + \sum_{a \in A_{\geq b}} d(F(a), F(P))^p \\ &= \sum_{a \in A} d(F(a), F(P))^p, \end{aligned}$$

in other words, $d(F(b), F(P))^p \leq \text{GD}_p(F(A), F(P))^p$. Finally, averaging over all $b \in B$ yields the conclusion, taking into account that condition 3 ensures the strict inequality. \square

For the Inverted Generational Distance Now we will concentrate on the Inverted Generational indicator IGD_p . In this context we will always assume that the Pareto front $F(P)$ is either finite or a finite size approximation has been chosen to be able to use the expression given in Definition 2.6. The following is a simple property that will be required among the hypotheses of the remaining results.

Lemma 3.7. *Let $A, B \subset \mathbb{R}^n$ be finite archives and P a Pareto set approximation.*

$$\forall a \in A, \forall b \in B, \forall x \in P_{\leq b} : b \prec a \implies d(F(b), F(x)) < d(F(a), F(x)).$$

Proof. Suppose that for $a \in A$, $b \in B$, and $x \in P$, we have $x \preceq b \prec a$, then $d(F(b), F(x)) = \|F(b) - F(x)\| < \|F(a) - F(x)\| = d(F(a), F(x))$. This property clearly holds for all $x \in P_{\leq b}$. \square

Strengthening the previous property to the whole Pareto set, we obtain our first Pareto compliance criterion.

Proposition 3.8. *Let $A, B \subset \mathbb{R}^n$ be finite and strictly well-dominated archives with $B \preceq A$ and P a finite Pareto set approximation such that*

$$\forall a \in A, \forall b \in B, \forall x \in P : b \prec a \implies d(F(b), F(x)) < d(F(a), F(x)).$$

Then: $\text{IGD}_p(F(B), F(P)) < \text{IGD}_p(F(A), F(P))$.

In the particular case when an approximation P to the Pareto set satisfies $P \preceq B$ (i.e., $P_{\leq B} = P$) it is clear that the conditions of Proposition 3.8 hold immediately by Lemma 3.7, thus in this case the validity of the Pareto compliance for the IGD_p -indicator is guaranteed.

Proof of Proposition 3.8. By the hypothesis, taking successive infima over all $x, x' \in P$ at both sides of the initial inequality, it follows that if $F(P)$ is a compact subset, for all $a \in A$ and $b \in B$ the dominance relation $b \prec a$ implies

$$d(F(b), F(P)) = \inf_{x \in X} d(F(b), F(x)) < \inf_{x' \in X} d(F(a), F(x')) = d(F(a), F(P)), \quad (5)$$

or only that $d(F(b), F(P)) \leq d(F(a), F(P))$ if $F(P)$ were not known to be compact, because in the first case the infimum (actually minimum) at the right-hand side of (5) is reached at some point $F(x_0)$ on the Pareto front $F(P)$, ensuring an strict inequality

$$d(F(b), F(P)) \leq d(F(b), F(x_0)) < d(F(a), F(x_0)) = d(F(a), F(P)).$$

Without the compactness assumption, the infimum at the right-hand side could reach the one to its left, but of course this is not the case here with $F(P)$ finite. Finally, using that $B = \bigcup_{a \in A} B_{\leq a}$ the claim follows from

$$\begin{aligned} \text{IGD}_p(F(B), F(P))^p &= \frac{1}{|P|} \sum_{x \in P} d(F(x), F(B))^p = \frac{1}{|P|} \sum_{x \in P} \min_{a \in A} d(F(x), F(B_{\leq a}))^p \\ &< \frac{1}{|P|} \sum_{x \in P} \min_{a \in A} d(F(x), F(a))^p = \frac{1}{|P|} \sum_{x \in P} d(F(x), F(A))^p \\ &= \text{IGD}_p(F(A), F(P))^p, \end{aligned}$$

as stated. □

Remark. Notice that in Proposition 3.8 the Pareto compliance of the Inverted Generational Distance IGD_p does not depend explicitly on the size of the sets A , B , or their subsets, in contrast to what was found in Theorem 3.4 for the Generational Distance GD_p . This is in part due to the strength of the assumption used for this proposition, which in many cases of interest will unfortunately not hold for all $x \in P$. It is clear from the proof that this assumption can be replaced by any of the following slightly weaker ones:

1. $\forall a \in A, \forall x \in P: d(F(B_{\leq a}), F(x)) < d(F(a), F(x))$,
2. $\forall b \in B, \forall x \in P: d(F(b), F(x)) < d(F(A_{\geq b}), F(x))$.

Indeed, the proof using condition 1 is basically the same, and the one using condition 2 just requires to write $A = \bigcup_{b \in B} A_{\geq b}$ in order to see that

$$\sum_{x \in P} \min_{b \in B} d(F(x), F(b))^p \leq \sum_{x \in P} \min_{b \in B} d(F(x), F(A_{\geq b}))^p = \sum_{x \in P} d(F(x), F(A))^p.$$

Still, these conditions are strong and do not necessarily apply when B does not cover the whole spectrum of admissible optimal solutions, thus alternatives are desirable.

Figure 4 displays two cases where the assumptions of Proposition 3.8 are valid. In fact, for both of them it is easy to see that $P \preceq B \preceq A$, so that their validity is a consequence of Lemma 3.7. Whenever these dominance relations hold, the size or form of the archives under comparison do not matter, and the Pareto compliance of the IGD_p -indicator is satisfied.

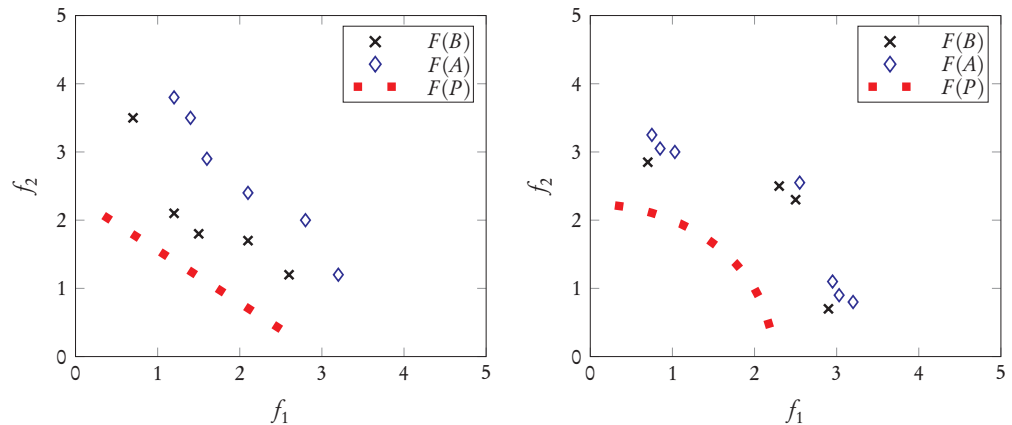


Figure 4. Two different scenarios where the IGD_p value of archive B is better (smaller) than the IGD_p value of archive A , where the hypotheses of Proposition 3.8 hold.

For the following statement, we recall that given a finite archive $A \subset X$ and a fixed point $x \in P$ there always exists at least one element $a_0(x) \in A$ such that $a_0(x) \in \arg \min_{a \in A} d(F(x), F(a))$. The minimal p -average distance between points $F(a)$ and $F(P_{\not\leq B})$ will be abbreviated by

$$\delta_A := \frac{1}{|P_{\not\leq B}|} \min_{a \in A} \left\{ \sum_{x \in P_{\not\leq B}} d(F(x), F(a))^p \right\} = \min_{a \in A} IGD_p(F(a), F(P_{\not\leq B}))^p.$$

Theorem 3.9. Let $A, B \subset \mathbb{R}^n$ be finite and strictly well-dominated archives with $B \preceq A$, such that any of the following conditions hold true:

1. $\forall a \in A, \forall b \in B: b \prec a \implies IGD_p(F(b), F(P_{\not\leq B})) < IGD_p(F(a), F(P_{\not\leq B}))$.
2. $\exists a_0 \in A$ such that $\forall x \in P_{\not\leq B}: a_0 \in \arg \min_{a \in A} d(F(x), F(a))$;
3. $\forall x \in P_{\not\leq B}: d(F(A), F(x))^p = \delta_A$;

then, $IGD_p(F(B), F(P)) < IGD_p(F(A), F(P))$.

Proof. First, notice that Lemma 3.7 and Proposition 3.8 imply that we always have $IGD_p(F(B), F(P_{\preceq B})) \leq IGD_p(F(A), F(P_{\preceq B}))$. From condition 1, and by a similar calculation to the one in the proof of Proposition 3.8 (but restricted to $P_{\not\leq B}$) it follows that $IGD_p(F(B), F(P_{\not\leq B})) \leq IGD_p(F(A), F(P_{\not\leq B}))$. Since $P = P_{\preceq B} \sqcup P_{\not\leq B}$, the conclusion easily follows in this case by adding $|P_{\preceq B}|$ times the first inequality with $|P_{\not\leq B}|$ times the second, and dividing both sides by $|P|$.

Second, notice that for $A \subset X$ the following inequality always holds

$$\sum_{x \in P_{\not\leq B}} \min_{a \in A} d(F(x), F(a))^p \leq \min_{a \in A} \sum_{x \in P_{\not\leq B}} d(F(x), F(a))^p, \quad (6)$$

because at the left-hand side the minimum of each term may be reached for different $a \in A$, while at the right-hand side the same $a \in A$ is required in all terms. For both sides to be equal it is enough (although not necessary) the existence of some fixed $a_0 \in A$ at which the distance from $F(x)$ to $F(A)$ is reached for all $x \in P_{\not\leq B}$. Explicitly, $\forall x \in P_{\not\leq B}: d(F(x), F(A)) = d(F(x), F(a_0))$. Indeed, in this case both sides of (6) attain the same minimum possible value at the point a_0 . Before we proceed with the final calculation, notice that from both conditions 2 and 3 we have

$$\begin{aligned} \min_{a \in A} \sum_{x \in P_{\not\leq B}} d(F(x), F(a))^p &= \\ \text{(by 2)} &= \sum_{x \in P_{\not\leq B}} d(F(x), F(a_0))^p = \sum_{x \in P_{\not\leq B}} d(F(x), F(A))^p, \text{ and} \end{aligned} \tag{7}$$

$$\text{(by 3)} = |P_{\not\leq B}| \delta_A = \sum_{x \in P_{\not\leq B}} \delta_A = \sum_{x \in P_{\not\leq B}} d(F(x), F(A))^p. \tag{8}$$

From these observations, and using that $P_{\leq B} \preceq B$ with Proposition 3.8, it follows that

$$\begin{aligned} \text{IGD}_p(F(B), F(P))^p &= \frac{1}{|P|} \sum_{x \in P} d(F(x), F(B))^p \\ &\leq \frac{1}{|P|} \left[\sum_{x \in P_{\leq B}} \min_{a \in A} d(F(x), F(B_{\leq a}))^p + \sum_{x \in P_{\not\leq B}} \min_{a \in A} \{d(F(x), F(B_{\leq a}))^p\} \right] \\ &\leq \frac{1}{|P|} \left[\sum_{x \in P_{\leq B}} \min_{a \in A} d(F(x), F(a))^p + \min_{a \in A} \left\{ \min_{b \leq a} \sum_{x \in P_{\not\leq B}} d(F(x), F(b))^p \right\} \right] \\ &\leq \frac{1}{|P|} \left[\sum_{x \in P_{\leq B}} d(F(x), F(A))^p + \min_{a \in A} \sum_{x \in P_{\not\leq B}} d(F(x), F(a))^p \right], \end{aligned}$$

where we used condition 1 to recast the second term. Using (7) or (8) we conclude that

$$\begin{aligned} &= \frac{1}{|P|} \left[\sum_{x \in P_{\leq B}} d(F(x), F(A))^p + \sum_{x \in P_{\not\leq B}} d(F(x), F(A))^p \right] \\ &= \frac{1}{|P|} \sum_{x \in P} d(F(x), F(A))^p = \text{IGD}_p(F(A), F(P))^p. \quad \square \end{aligned}$$

Remark. Condition 2 of Theorem 3.9 requires that for some $a_0 \in A$ we have

$$a_0 \in \bigcap_{x \in P_{\not\leq B}} \arg \min_{a \in A} d(F(a), F(x)) \neq \emptyset.$$

In dimension $n = 2$, from the fact that A and P consist only of mutually non-dominating points, it is possible to show that the intersection admits more than one element only when $|P_{\not\leq B}| = 1$ and $|A| > 1$, otherwise, the

condition is that the intersection should be precisely a_0 . For $n > 2$ the situation is, in general, more complicated. Also, condition 3 of Theorem 3.9 can be rephrased by saying that each $x \in P_{\not\leq B}$ has an associated element $a_x \in A$ at which its distance to A is reached, and for any of those x this distance has the same value δ_A (given by the average of all distances $d(x, a)$ to points $a \in A$), i.e.,

$$\min_{a \in A} d(F(a), F(x))^p = d(F(a_x), F(x))^p = \delta_A.$$

Note that from the definition of δ_A as a minimal average, the seemingly weaker condition $\forall x \in P_{\not\leq B}: d(F(A), F(x))^p \geq \delta_A$, implies that all the “ \geq ”-inequalities need to be actual equalities, because a collection of values cannot all be greater than, or equal to their average, unless all are equal to it.

Figure 5 illustrates four situations where Theorem 3.9 applies. At the upper left appears a variant of the second case in Figure 4 where condition 1 holds, while at the upper right the diamond in the corner of $F(A)$ corresponds to the image $F(a_0)$ of the point a_0 required in condition 2. Finally, both of the lower diagrams correspond to cases where the points of $F(P)$ are equidistant of $F(A)$ so that condition 3 holds, and δ_A is precisely this distance (to the power p).

Remark. We conclude our analysis of the IGD_p -indicator highlighting useful cases where each one of the assumptions needed for Theorem 3.9 hold, and which are illustrated in Figure 5.

1. The first condition holds, in particular, whenever $P \preceq B \preceq A$.
2. The second condition is valid when the dominated archive A is much more convex than the Pareto front. For example, this can occur if $F(P)$ is concave, or flat, and A is sufficiently convex (i.e., curved in the opposite direction of $F(P)$).
3. The third condition holds whenever the Pareto front $F(P)$ and the dominated archive A have corresponding equidistant points.

Consequences for the Averaged Hausdorff Distance Criteria for the Pareto compliance of Δ_p can now be stated as a consequence of the compliance of their intermediate generational indicators GD_p and IGD_p . Using the same notation for δ_A as in the previous part, we can add up the results of the previous sections in the following

Theorem 3.10. *Let $A, B \subset \mathbb{R}^n$ be finite and well-dominated archives with $B \preceq A$ such that*

$$\forall a \in A, \forall b \in B: b \preceq a \implies \frac{|B_{\preceq a}|}{|B|} = \frac{|A_{\succeq b}|}{|A|},$$

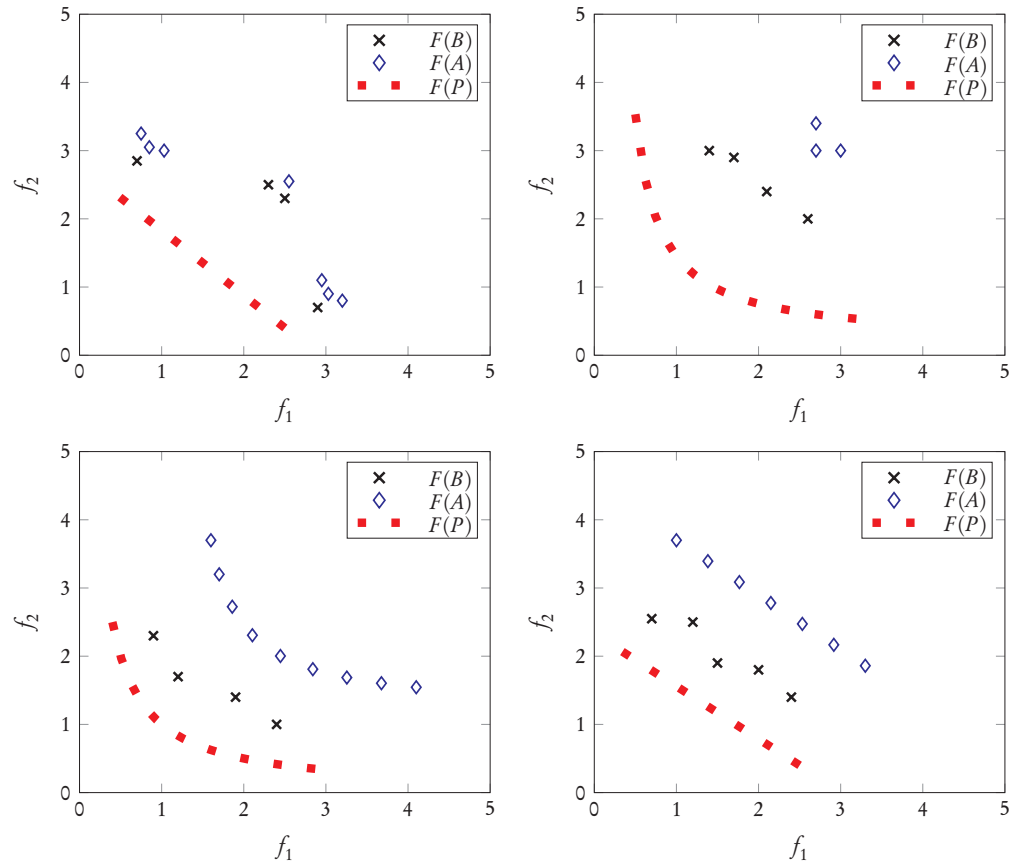


Figure 5. Four situations where the IGD_p value of archive B is smaller (better) than the IGD_p value of archive A , and where the conditions of Theorem 3.9 are satisfied.

and any of the following conditions hold

1. $\forall a \in A, \forall b \in B, \forall x \in P: b \prec a \implies d(F(b), F(x)) < d(F(a), F(x));$
2. $\exists a_0 \in A$ such that $\forall x \in P_{\neq B}: a_0 \in \arg \min_{a \in A} d(F(x), F(a));$
3. $\forall x \in P_{\neq B}: d(F(A), F(x))^p = \delta_A;$

Then, $\Delta_p(F(B), F(P)) < \Delta_p(F(A), F(P)).$

Proof. From condition 1 it follows easily that

$$b \prec a \implies d(F(b), F(P)) < d(F(a), F(P)),$$

which is precisely condition 1 of Theorem 3.4, and furthermore, it also implies condition 1 of Theorem 3.9. Hence, this list of conditions satisfy all the requirements of Theorems 3.4 and Theorem 3.9. From the definition of Δ_p we immediately arrive at the desired conclusion. \square

It is possible to study in detail which of the GD_p or IGD_p -distances to the Pareto front is larger for two candidate archives, in order to find the indicator that will correspond to the value of Δ_p in each case. We will not explore this question further, but clearly, if the same indicator realizes the value of Δ_p for both archives the compliance can be evaluated through the previously presented results for that indicator alone.

Conclusions and Perspectives

In this study, different criteria have been obtained to establish the Pareto compliance for the averaged Hausdorff distance, and its associated generational p -indicators, when dealing with finite solutions/approximations. Inaccuracies found in the current literature have also been corrected, and as far as the constraints and scope of the problem permits, general results have been provided. As an important observation, from Theorem 3.4 and the examples presented, it can be stated that the Pareto compliance of GD_p (and therefore of Δ_p) is not, in general, completely independent on the size of the solution sets to be compared (except under very particular circumstances). Additionally, two alternative conditions were given to guarantee the compliance of the IGD_p -indicator and that help to elucidate its behavior. Although the conditions depend on the characteristics of the Pareto set under consideration, it is important to note that they are useful in cases where the convexity/concavity of the solutions and/or the expected ones of the Pareto set are known.

As it was already mentioned in the introduction, it is natural to consider generalizations of GD_p , IGD_p , and Δ_p that can enhance their usefulness. Some proposals have already been made for GD_p , e.g., in [6], and furthermore, a (p, q) -averaged Hausdorff distance $\Delta_{p,q}$ has been introduced in [12] for the case of finite archives, and in [1] for continuous ones, in terms of generalized generational distances $GD_{p,q}$ and $IGD_{p,q}$ suited to each context.

From a theoretical perspective, further consideration of the convenience of those generalizations require not only a deeper understanding of the properties of GD_p , IGD_p and Δ_p as indicators themselves, but also a set of well-established and testable assumptions allowing for appropriate comparisons. This work is an intermediate contribution to that aim from the purely theoretical side, and its extension to the (p, q) -generalizations is a matter of further investigation.

Apart from EMO algorithms, performance indicators play an important role on all set oriented methods, i.e., methods that generate an entire set of solutions in one run of an algorithm. This is the case, e.g., of subdivision and cell mapping techniques (see [11, 5]). It is a matter of further study how the results presented here can be extended to those contexts.

Acknowledgements

The author is indebted to the referees for their valuable comments that helped to improve this manuscript. This work was partially financed by Vicerrectoría de Investigación, Pontificia Universidad Javeriana, Bogotá, Colombia.

Conflict of Interest

None.

References

- [1] Bogoya JM, Vargas A, Cuate O, Schütze O. A (p, q) -averaged Hausdorff distance for arbitrary measurable sets. *Mathematical and Computational Applications*, 23(3) 51, 2018.
doi: 10.3390/mca23030051
- [2] Coello CAC, Cortés NC. Solving multiobjective optimization problems using an artificial immune system. *Genetic Programming and Evolvable Machines*, 6(2):163–190, 2005.
doi: 10.1007/s10710-005-6164-x
- [3] Hansen M, Jaszkiwicz A. *Evaluating the quality of approximations to the non-dominated set*. IMM, Department of Mathematical Modelling, Technical University of Denmark, 1998.
- [4] Heinonen J. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, New York, 2001.
doi: 10.1007/978-1-4613-0131-8
- [5] Hillermeier C. *Nonlinear Multiobjective Optimization. A Generalized Homotopy Approach*. Birkhäuser, 2001.
doi: 10.1007/978-3-0348-8280-4
- [6] Ishibuchi H, Masuda H, Tanigaki Y, Nojima Y. *Modified Distance Calculation in Generational Distance and Inverted Generational Distance*. In: Gaspar-Cunha A, Antunes CH, Coello Coello CA (eds.), *Evolutionary Multi-Criterion Optimization 8th International Conference EMO 2015*, Guimarães, Portugal. *Lecture Notes in Computer Science* vol. 9019, pp. 110–125, Springer, 2015.
doi: 10.1007/978-3-319-15892-1_8

- [7] Pareto V. *Manual of Political Economy: A Critical and Variorum Edition*. Edited by Montesano A, Zanni A, Bruni L, Chipman JS, McLureet M. Oxford University Press, 2014.
- [8] Rudolph G, Schütze O, Grimme C, Domínguez-Medina C, Trautmann H. Optimal averaged Hausdorff archives for bi-objective problems: theoretical and numerical results. *Computational Optimization and Applications*, 64:589–618, 2016.
doi: 10.1007/s10589-015-9815-8
- [9] Schütze O, Esquivel X, Lara A, Coello CA. Using the averaged Hausdorff distance as a performance measure in evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 16(4):504–522, 2012.
doi: 10.1109/TEVC.2011.2161872
- [10] Siwel J, Yew-Soon O, Jie Z, Liang F. Consistencies and contradictions of performance metrics in multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 44(12):2329–2404, 2014.
doi: 10.1109/TCYB.2014.2307319
- [11] Sun JQ, Xiong FR, Schütze O, Hernández C. *Cell Mapping Methods - Algorithmic Approaches and Applications*. Nonlinear Systems and Complexity 99, Springer, 2019.
doi: 10.1007/978-981-13-0457-6
- [12] Vargas A, Bogoya J. A generalization of the averaged Hausdorff distance. *Computación y Sistemas*, 22(2):331–345, 2018.
doi: 10.13053/CyS-22-2-2950
- [13] Veldhuizen D. Multiobjective evolutionary algorithms: classifications, analyses, and new innovations. 1999. *Air Force Institute of Technology*, page 249, 1999.
- [14] Zitzler E, Thiele L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, 1999.
doi: 10.1109/4235.797969
- [15] Zitzler E, Thiele L, Laumanns M, Fonseca CM, Da Fonseca VG. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.
doi: 10.1109/TEVC.2003.810758

Sobre la Sujeción de Pareto para la Distancia Promedio de Hausdorff como Indicador de Desempeño

Resumen. La distancia promedio de Hausdorff Δ_p es una inframétrica recientemente introducida en optimización multiobjetivo evolutiva (EMO) como una herramienta para medir la optimalidad de aproximaciones finitas al frente de Pareto asociado con un problema de optimización multiobjetivo (MOP). Presentamos aquí un estudio puramente teórico sobre la sujeción del indicador Δ_p a la noción de optimalidad de Pareto. Puesto que Δ_p está definida en términos de una versión modificada de otros indicadores bien conocidos como lo son la distancia generacional GD_p y la distancia generacional invertida IGD_p , discutimos en detalle criterios específicos para la sujeción de tipo Pareto de cada uno de ellos. Adicionalmente, presentamos una revisión del comportamiento previamente conocido de estos indicadores, corrigiendo imprecisiones que se encuentran en la literatura y establecemos resultados nuevos y más generales, incluyendo pruebas detalladas y ejemplos ilustrativos.

Palabras clave: distancia promedio de Hausdorff; distancia generacional; distancia generacional invertida; optimización multiobjetivo; optimalidad de Pareto; indicador de desempeño.

Sobre a Sujeição de Pareto para a Distância Média de Hausdorff como Indicador de Desempenho

Resumo A distância média de Hausdorff Δ_p é uma inframétrica introduzida recentemente em otimização multiobjetivo evolucionária (EMO) como uma ferramenta para medir a otimalidade de aproximações finitas para o frente de Pareto associado com um problema de otimização multiobjetivo (MOP). Apresentamos aqui um estudo puramente teórico sobre a sujeição do indicador Δ_p à noção de otimalidade de Pareto. Desde Δ_p é definido em termos de uma versão modificada de outros indicadores bem conhecidos, tais como a distância geracional GD_p e a distância geracional invertida IGD_p , discutimos em detalhes critérios específicos para a sujeição de Pareto de cada um deles. Além disso, apresentamos uma revisão do comportamento previamente conhecido desses indicadores, corrigindo imprecisões encontradas na literatura e estabelecemos novos e mais gerais resultados, incluindo testes detalhados e exemplos ilustrativos.

Palavras-chave: distância média de Hausdorff; distância geracional; distância geracional invertida; otimização multiobjetivo; otimalidade de Pareto; indicador de desempenho.

Andrés Vargas

Currently, Associate Professor at the Department of Mathematics, Pontificia Universidad Javeriana, Bogotá, Colombia. He obtained his PhD in mathematics at Universität Bonn, Germany, working in Differential Geometry. His research interests include also Mathematical Physics, PDEs, Metric Geometry and applications.