

Article

A (p, q) -Averaged Hausdorff Distance for Arbitrary Measurable Sets

Johan M. Bogoya ^{1,*} , Andrés Vargas ¹ , Oliver Cuate ² and Oliver Schütze ²

¹ Department of Mathematics, Pontificia Universidad Javeriana, 110231 Bogotá, Colombia; a.vargasd@javeriana.edu.co

² Computer Science Department, Cinvestav-IPN, 07360 Mexico City, Mexico; ocuate@computacion.cs.cinvestav.mx (O.C.); schuetze@cs.cinvestav.mx (O.S.)

* Correspondence: jbogoya@javeriana.edu.co; Tel.: +57-312-5588102

Received: 22 July 2018; Accepted: 16 September 2018; Published: 18 September 2018

Abstract: The Hausdorff distance is a widely used tool to measure the distance between different sets. For the approximation of certain objects via stochastic search algorithms this distance is, however, of limited use as it punishes single outliers. As a remedy in the context of evolutionary multi-objective optimization (EMO), the averaged Hausdorff distance Δ_p has been proposed that is better suited as an indicator for the performance assessment of EMO algorithms since such methods tend to generate outliers. Later on, the two-parameter indicator $\Delta_{p,q}$ has been proposed for finite sets as an extension to Δ_p which also averages distances, but which yields some desired metric properties. In this paper, we extend $\Delta_{p,q}$ to a continuous function between general bounded subsets of finite measure inside a metric measure space. In particular, this extension applies to bounded subsets of \mathbb{R}^k endowed with the Euclidean metric, which is the natural context for EMO applications. We show that our extension preserves the nice metric properties of the finite case, and finally provide some useful numerical examples that arise in EMO.

Keywords: averaged Hausdorff distance; evolutionary multi-objective optimization; power means; metric measure spaces; performance indicator; Pareto front

1. Introduction

The Hausdorff distance d_H (see [1]) is an established and widely used tool to measure the proximity of different sets. It is, among others, used in several research fields such as image matching (e.g., [2–4]), the approximation of manifolds in dynamical systems ([5–7]), in fractal geometry ([8]), or in the context of convergence analysis in multi-objective optimization ([9–13]). One major reason for the use of d_H is that it defines a metric on the set of all nonempty bounded closed sets in a metric space. However, one characteristic of the Hausdorff distance is that it heavily punishes single outliers which is a severe drawback in many cases. For instance, it is known that stochastic search algorithms are generally quite effective in the (global) approximation of certain objects, however, it is also known that these approximations may come with a few outliers (e.g., [14]). For those cases, the approximation quality is not reflected by the value of the Hausdorff distance.

As a remedy, in the context of evolutionary multi-objective optimization, Schütze et al. [14] have made a first effort to propose the averaged Hausdorff distance Δ_p . As opposed to d_H , this indicator averages the distances involved in the proximity measure of the given sets and is hence much more suitable in the context of stochastic search as single (or few) outliers in a candidate solution set are not punished hard any more. On the other hand, compared to d_H , Δ_p has two shortcomings: (i) it only defines an inframetric instead of a metric; and (ii) it is only defined for *finite* approximations of the solution set. In the particular context of continuous multi-objective optimization, it is known that the solution set, the so-called Pareto set, and its image, the Pareto front, form manifolds of

certain dimensions. Hence, it is natural that the candidate solution set (i.e., the set computed by a given solver) is not restricted to finitely many points, but may also form a continuous set. This is in fact already the case for set-based optimization techniques such as the cell-to-cell mappings ([15–17]) and the subdivision techniques ([10,18,19]). In the context of evolutionary multi-objective optimization, typically a finite set of candidate solutions (a *population*) is generated ([20–23]). However, also here it is a rather natural approach to construct a continuous set out of the final population using, e.g., interpolation techniques (see [24,25]).

In [26], a modification of the Δ_p indicator called the (p, q) -averaged Hausdorff distance $\Delta_{p,q}$ has been introduced by the first two authors. This indicator generalizes the averaged Hausdorff distance Δ_p , is strongly related to the Hausdorff distance d_H , and admits an expression in terms of the matrix $\ell_{p,q}$ -norm $\|\cdot\|_{p,q}$. Moreover, when $1 \leq p, q < \infty$ it is a proper metric, while for the remaining cases where $|p|, |q| \geq 1$ it is still an inframetric. In addition, when finding optimal archives the parameters p and q play crucial geometrical roles. More precisely, in the context of EMO, p handles the closeness to the Pareto front and q handles the dispersion. The indicator, however, is restricted to finite sets.

In this work, we propose a more general version of the $\Delta_{p,q}$ indicator that can be applied to general measurable subsets and that preserves the useful advantages of the finite case. Consideration is also given to the Pareto-compliance of an intermediate indicator $\text{GD}_{p,q}$ that is employed to define $\Delta_{p,q}$. The indicator is hence the first one that can be used in the context of multi-objective optimization using continuous approximations of the Pareto set/front as described above. Numerical results on two well known evolutionary algorithms will show the benefit of such continuous archives compared to discrete ones that have been used so far in lack of a suitable performance indicator.

This paper is organized as follows: In Section 2, we briefly state the background required for the understanding of this work. In Section 3, we introduce the extended version of the $\text{GD}_{p,q}$ and $\Delta_{p,q}$ indicators, discussing their properties and providing some sufficient criteria for the Pareto compliance of the first one. In Section 4, we present some numerical results that show the applicability and the benefit of the novel indicator in particular in the context of multi-objective optimization. Finally, we draw our conclusions and present possible paths for future research in Section 5.

2. Preliminaries

In this section, we briefly present the required background on integral power means and multi-objective optimization that will be needed for our purposes. Throughout the document we employ the notation $\mathbb{R}^\times := \mathbb{R} \setminus \{0\}$ and $\overline{\mathbb{R}} := [-\infty, \infty]$ for simplicity.

2.1. Integral Power Means

The theory can be presented in the general setting of metric measure spaces, briefly outlined below, but for simplicity the reader may assume that the specific context of our interest is that of well-behaved bounded subsets of the n -dimensional Euclidean space \mathbb{R}^n endowed with its standard Lebesgue measure which gives rise to the conventional notion of volume (when it is defined). For a quick review of measure spaces see [27] (Section 1.4), and for a simple explanation of the Lebesgue measure see [28] (Chapter 2). Integral means appear already in [29] (Chapter 6). A comprehensive account on the properties of means can be found in [30].

For greater generality, we recall that (Σ, d, μ) is called a metric measure space if (Σ, d) is a metric space with a measure μ defined on its Borel σ -algebra $\mathfrak{M}(\Sigma)$, i.e., the smallest σ -algebra containing all the open subsets of the metric topology of (Σ, d) . A measure μ is said to be finite if $\mu(\Sigma) < \infty$, and in this case Σ is called a finite-measure space.

Now, given $p \in \mathbb{R}^\times$ and any measurable function $f: X \subset \Sigma \rightarrow [0, \infty)$ over a finite-measure set X , we can define the p -average of f over X (or the p -power mean of f over X), by

$$\mathcal{M}^p(f(x)) := \left(\frac{1}{\mu(X)} \int_X f(x)^p d\mu \right)^{\frac{1}{p}}. \quad (1)$$

Henceforth, the integral at the RHS will be abbreviated as

$$\int_X f^p d\mu := \frac{1}{\mu(X)} \int_X f(x)^p d\mu.$$

If necessary, when the measure μ is clear from the context, the element $d\mu$ will be written as dx to emphasize the variable of integration x . In addition, the notation $\mathcal{M}^p(f(X)) \equiv \mathcal{M}^p(f(x))$ and $|X| \equiv \mu(X)$ will also be employed to simplify expressions whenever appropriate.

Let us note that for $p \geq 1$ we have $\mathcal{M}^p(f(X)) = \mu(X)^{-\frac{1}{p}} \|f\|_p$, where $\|\cdot\|_p$ denotes the p -norm of the Lebesgue space $L^p(X, \mu)$. Furthermore, it is not difficult to show, with the aid of L'Hôpital's rule, that the integral power mean \mathcal{M}^p can be extended to the cases $p = \pm\infty$. Indeed, if $f \not\equiv 0$, denoting the essential supremum and essential infimum of f on X by $\|f\|_\infty := \text{ess sup}_{x \in X} f(x)$ and $\|1/f\|_\infty^{-1} := \text{ess inf}_{x \in X} f(x)$, respectively, it follows that

$$\mathcal{M}^\infty(f(x)) := \lim_{p \rightarrow \infty} \left(\int_X f^p d\mu \right)^{\frac{1}{p}} = \|f\|_\infty \lim_{p \rightarrow \infty} \left(\int_X \left(\frac{f(x)}{\|f\|_\infty} \right)^p d\mu \right)^{\frac{1}{p}} = \|f\|_\infty,$$

because the last integrand is smaller than 1 and the limit is 1. Similarly,

$$\mathcal{M}^{-\infty}(f(x)) := \lim_{p \rightarrow -\infty} \left(\int_X f^p d\mu \right)^{\frac{1}{p}} = \lim_{p \rightarrow -\infty} \left(\int_X \left(\frac{1}{f} \right)^p d\mu \right)^{-\frac{1}{p}} = \left\| \frac{1}{f} \right\|_\infty^{-1}.$$

We recall that $\|\cdot\|_\infty$ is precisely the norm of the Lebesgue space $L^\infty(X, \mu)$. We can also define \mathcal{M}^p when $p = 0$ as follows:

$$\mathcal{M}^0(f(x)) := \exp \left(\int_X \log f d\mu \right).$$

It can be considered the integral generalization of the notion of geometric mean for finitely many elements.

2.2. Multi-objective Optimization

As an application of the (p, q) -distances, we will consider in this work continuous multi-objective optimization problems (MOPs). Problems of this kind can be expressed mathematically as

$$\min \{F(x) : x \in Q \subset \mathbb{R}^n\}, \quad (2)$$

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

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

We will assume here that all objectives $f_i: X \rightarrow \mathbb{R}$, for $i \in \{1, \dots, k\}$, are continuous. The optimality of MOPs is typically defined via the concept of dominance (see [31]).

Definition 1. In the context of MOPs the following are standard notions:

- (i) Let $v = (v_1, \dots, v_k)$ and $w = (w_1, \dots, w_k) \in \mathbb{R}^k$. Then the vector v is less than w (denoted $v <_p w$), if $v_i < w_i$ for all $i \in \{1, \dots, k\}$. The relation \leq_p is defined analogously.

- (ii) A vector $y \in Q$ is dominated by a vector $x \in Q$ (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 $f_j(x) < f_j(y)$.
- (iii) A point $x \in Q$ is called Pareto optimal or a Pareto point if there is no $y \in Q$ which dominates x .
- (iv) The set of all Pareto optimal solutions is called the Pareto set, denoted by P_Q .
- (v) The image of the Pareto set, $F(P_Q)$, is called the Pareto front.

It is known that under certain mild smoothness assumptions the Pareto set and the Pareto front define $(k - 1)$ -dimensional objects [32]. Hence, for set oriented solvers such as cell mapping, subdivision techniques, and evolutionary algorithms, the question naturally arises as to how to measure the approximation quality of the obtained solution set with respect to the Pareto set/front. To accomplish this task, several *performance indicators* have been proposed in the specialized literature. There exist, for instance, the hypervolume indicator [21,33], the R2 indicator [34], the IGD^+ [35], and the DOA [36]. Moreover, in the context of multi-criteria decision-making processes, the properties of some distance measures, as the Hamming, Euclidean, and Hausdorff metrics, is investigated in [37,38]. In this work, we will focus on a new variant of the Hausdorff distance [6]. For convenience of the reader, we recall in the following the most important definitions.

Definition 2. Let $u, v \in \mathbb{R}^n$, arbitrary $A, B \subset \mathbb{R}^n$, and $\|\cdot\|$ be a vector norm. The Hausdorff distance $d_H(\cdot, \cdot)$ is defined as follows:

- (i) $\text{dist}(u, A) := \inf \{\|u - v\| : v \in A\}$,
- (ii) $\text{dist}(B, A) := \sup \{\text{dist}(u, A) : u \in B\}$,
- (iii) $d_H(A, B) := \max \{\text{dist}(A, B), \text{dist}(B, A)\}$.

The Hausdorff distance d_H is widely used in many fields. It is, however, of limited practical use when measuring the distance of the outcome of a stochastic search method such as an evolutionary algorithm to the Pareto set/front. The main reason for this is that evolutionary algorithms may generate outliers that are punished too strongly by d_H . As a remedy, the *averaged* Hausdorff distance has been proposed in [14]. In this study the vector norm is the 2-norm, i.e., the Euclidean norm.

Definition 3 (Schütze et al. [14]). For $p \in \mathbb{N}$, and finite sets $A, B \subset \mathbb{R}^n$ the value

$$\Delta_p(A, B) := \max \{GD_p(A, B), IGD_p(A, B)\},$$

where

$$GD_p(A, B) := \left(\frac{1}{|A|} \sum_{a \in A} d(a, B)^p \right)^{\frac{1}{p}} \quad \text{and} \quad IGD_p(A, B) := \left(\frac{1}{|B|} \sum_{b \in B} d(b, A)^p \right)^{\frac{1}{p}},$$

is called the averaged Hausdorff distance between A and B .

The indicator Δ_p can be viewed as a composition of slight variations of the Generational Distance (GD, see [39]) and the Inverted Generational Distance (IGD, see [40]). It is $\Delta_\infty = d_H$, but for finite values of p the indicator Δ_p averages the distances considered in d_H . More precisely, the larger the value of p , the harder single outliers will be punished by Δ_p . Hence, as opposed to d_H , the distance Δ_p does not punish single (or few) outliers in a candidate set. For more discussion about Δ_p and its relation to other indicators we refer to [14,41].

Definition 4 (Vargas–Bogoya [26]). For $p, q \in \mathbb{R}^\times$, and finite sets $A, B \subset \mathbb{R}^n$ the value

$$\Delta_{p,q}(A, B) := \max \{GD_{p,q}(A, B \setminus A), GD_{p,q}(B, A \setminus B)\},$$

where $\text{GD}_{p,q}(A, B) := \left(\frac{1}{|A|} \sum_{a \in A} \left(\frac{1}{|B|} \sum_{b \in B} d(a, b)^q \right)^{\frac{p}{q}} \right)^{\frac{1}{p}}$ is called the (p, q) -averaged Hausdorff distance between A and B .

For finite sets, the indicator $\Delta_{p,q}$, introduced in [26] was also defined for p or $q = 0$, and even for p or $q = \pm\infty$. It is a generalization of Δ_p in the sense that between disjoint subsets we have

$$\lim_{q \rightarrow -\infty} \Delta_{p,q} = \Delta_p.$$

The parameters p and q can be independently modified in order to produce customary spread archives (depending on q) located with customary closeness (depending on p) to the Pareto front.

Finally, let us recall that one characteristic of a performance indicator is *Pareto compliance*: for two subsets A and B we say that $A \preceq B$ if for every $b \in B$ there exists an element $a \in A$ such that $a \preceq b$. If this does not hold, we write $A \not\preceq B$. We say that a performance indicator \mathcal{I} is Pareto compliant if for any two sets A and B with $A \preceq B$ and $B \not\preceq A$ it follows $\mathcal{I}(A) \leq \mathcal{I}(B)$. We refer to [42] for details.

3. The (p, q) -Averaged Hausdorff Distance for Measurable Sets

3.1. Properties of Integral Power Means

We start summarizing some fundamental properties of integral power means that we will need for our subsequent calculations.

Theorem 1. Let X and Y denote finite-measure spaces, $f, g: X \rightarrow [0, \infty)$ non-negative measurable functions, and $d: X \times Y \rightarrow [0, \infty)$ a measurable function with respect to the product measure on $X \times Y$. The integral power mean \mathcal{M} satisfies the following properties:

- (i) If $p \in \overline{\mathbb{R}}$ and $k \in [0, \infty)$, then $\mathcal{M}_{x \in X}^p(k) = k$ and $\mathcal{M}_{x \in X}^p(kf(x)) = k \mathcal{M}_{x \in X}^p(f(x))$.
- (ii) For any $p \in \overline{\mathbb{R}}$, we have $\mathcal{M}_{x \in X}^p \left(\mathcal{M}_{y \in Y}^p(d(x, y)) \right) = \mathcal{M}_{y \in Y}^p \left(\mathcal{M}_{x \in X}^p(d(x, y)) \right)$.
- (iii) If $1 \leq p \leq \infty$, then $\mathcal{M}_{x \in X}^p(f(x) + g(x)) \leq \mathcal{M}_{x \in X}^p(f(x)) + \mathcal{M}_{x \in X}^p(g(x))$.
- (iv) If $p \in \overline{\mathbb{R}}$ and $f(x) \leq g(x)$ for all $x \in X$, then $\mathcal{M}_{x \in X}^p(f(x)) \leq \mathcal{M}_{x \in X}^p(g(x))$.
- (v) For $p, q \in \overline{\mathbb{R}}$ with $0 < p \leq q$, we have that $\mathcal{M}_{x \in X}^p(f(x)) \leq \mathcal{M}_{x \in X}^q(f(x))$.

Proof. The proofs of (i) and (ii) are straightforward. To prove (iii) we only need the Minkowski inequality,

$$\begin{aligned} \mathcal{M}_{x \in X}^p(f(x) + g(x)) &= \mu(X)^{-\frac{1}{p}} \|f + g\|_p \\ &\leq \mu(X)^{-\frac{1}{p}} (\|f\|_p + \|g\|_p) = \mathcal{M}_{x \in X}^p(f(x)) + \mathcal{M}_{x \in X}^p(g(x)). \end{aligned}$$

The proof of (iv) is also straightforward from the definitions and a simple proof of (v) can be given as a particular case of [43] (Theorem 3) which we recall here for completeness. For a positive real v , consider the function

$$\omega_r(v) := \int_1^v t^{r-1} dt = \begin{cases} \frac{v^r - 1}{r}, & r \neq 0; \\ \log v, & r = 0. \end{cases}$$

Since the function t^u , with t a positive constant and $u \geq 0$ is increasing with respect to u , we easily get $\omega_r(v) \leq \omega_s(v)$ for $0 \leq r \leq s$ and every $v \geq 0$. Consider the following linear integral operator

$$\mathcal{J}[f] := \frac{1}{\mu(X)} \int_X f \, d\mu = \int_X f \, d\mu.$$

Assume first, that $p \neq 0$, then for any $x \in X$,

$$\begin{aligned} \mathcal{J}\left[\omega_p\left(\frac{f(x)}{\mathcal{M}^p(f(X))}\right)\right] &= \frac{1}{p} \left(\mathcal{J}\left[\frac{f(x)}{\mathcal{M}^p(f(X))}\right]^p - \mathcal{J}[1] \right) \\ &= \frac{1}{p} \left(\frac{1}{(\mathcal{M}^p(f(X)))^p} \int_X f^p \, d\mu - \int_X 1 \, d\mu \right) = 0. \end{aligned}$$

Similarly, if $p = 0$ we have

$$\begin{aligned} \mathcal{J}\left[\omega_0\left(\frac{f(x)}{\mathcal{M}^0(f(X))}\right)\right] &= \mathcal{J}[\log(f(x)) - \log(\mathcal{M}^0(f(X)))] \\ &= \int_X \log f \, d\mu - \log(\mathcal{M}^0(f(X))) \mathcal{J}[1] = 0. \end{aligned}$$

Suppose that $0 \leq p < q$. Since $\omega_p(\cdot) \leq \omega_q(\cdot)$ implies $\mathcal{J}(\omega_p(\cdot)) \leq \mathcal{J}(\omega_q(\cdot))$, we obtain

$$0 \leq \mathcal{J}\left[\omega_q\left(\frac{f(x)}{\mathcal{M}^p(f(X))}\right)\right] = \frac{1}{q} \left[\left(\frac{\mathcal{M}^q(f(X))}{\mathcal{M}^p(f(X))} \right)^q - 1 \right],$$

from which it follows that $\mathcal{M}^p(f(x)) \leq \mathcal{M}^q(f(x))$. \square

3.2. Definition of $\Delta_{p,q}$ for Measurable Sets

With the aid of Theorem 1 we generalize the results of [26] (Section 3). For easy reference, we provide here slightly abbreviated but complete proofs. Given a metric measure space (Σ, d, μ) , let $\mathfrak{M}(\Sigma)$ denote the σ -algebra of all measurable subsets of Σ and let $\mathfrak{M}_{<\infty}(\Sigma)$ refer to those elements of $\mathfrak{M}(\Sigma)$ having finite measure. As it should be expected from the context, any set relation obtained from calculations involving an underlying measure μ should be understood to hold in a measure-theoretic sense, i.e., almost everywhere (*a.e.*). For example, for $X, Y \in \mathfrak{M}_{<\infty}(\Sigma)$, a result saying $X = Y$, or $X \subset Y$ actually holds almost everywhere, which means that $\mu\{X \neq Y\} = 0$, or $\mu\{X \not\subset Y\} = 0$, respectively. Thus, it is convenient in this setting to identify a set $X \in \mathfrak{M}_{<\infty}(\Sigma)$ with the whole equivalence class $[X] := \{Y \mid X = Y, \text{ a.e.}\}$, and think of these classes as the elements of $\mathfrak{M}_{<\infty}(\Sigma)$ to remove the need for the *a.e.* abbreviation. Also, to avoid an overload of parentheses in the forthcoming expressions, the distance $d(x, y)$ between $x, y \in \Sigma$ will be abbreviated by $d_{x,y}$.

Definition 5. For $p, q \in \mathbb{R}^\times$, the generational (p, q) -distance $\text{GD}_{p,q}(X, Y)$ between two sets $X, Y \in \mathfrak{M}_{<\infty}(\Sigma)$ is given by

$$\text{GD}_{p,q}(X, Y) := \mathcal{M}^p\left(\mathcal{M}^q(d_{x,y})\right) = \left(\int_X \left(\int_Y d_{x,y}^q \, dy \right)^{\frac{p}{q}} \, dx \right)^{\frac{1}{p}},$$

where the sets X and Y are implicitly assumed to be disjoint when $p < 0$ or $q < 0$.

As in the finite case, the definition of $\text{GD}_{p,q}$ can be easily extended for $p, q \in \overline{\mathbb{R}}$, but still has two undesirable drawbacks, first $\text{GD}_{p,q}(X, X)$ can be different from zero, and second, in general the values of $\text{GD}_{p,q}(X, Y)$ and $\text{GD}_{p,q}(Y, X)$ can be different, thus this indicator does not define a metric. To obtain a proper metric we introduce the following modification.

Definition 6. The (p, q) -averaged Hausdorff distance is the map $\Delta_{p,q}: \mathfrak{M}_{<\infty}(\Sigma) \times \mathfrak{M}_{<\infty}(\Sigma) \rightarrow [0, \infty)$ given by

$$\Delta_{p,q}(X, Y) := \max \{ \text{GD}_{p,q}(X, Y \setminus X), \text{GD}_{p,q}(Y, X \setminus Y) \}.$$

Remark 1. For finite subsets $X, Y \subset \mathbb{R}^n$ endowed with the standard counting measure μ , the previous notions of $\text{GD}_{p,q}$ and $\Delta_{p,q}$ coincide with the ones in Definition 4.

Figure 1 illustrates how the shape of $\Delta_{p,q}$ -metric balls $B_\varepsilon := \{x \in \mathbb{R}^2 : \Delta_{p,q}(A, x) \leq \varepsilon\}$ around a discrete set A of ten points (that approximates a segment of negative slope in the plane) varies as p and q take several different values. Notice that for negative values of p and q the balls' shape resemble the shape of A and enclose all of its points.

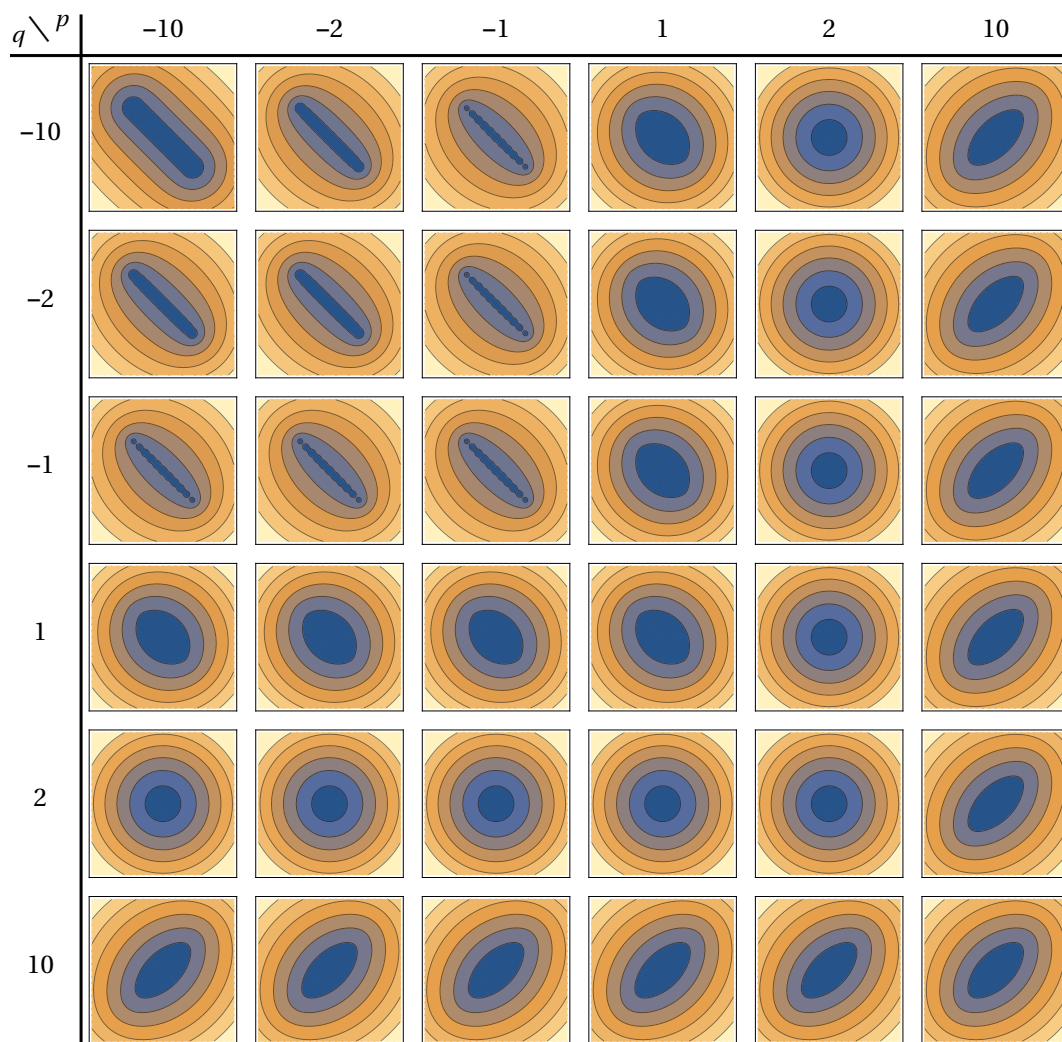


Figure 1. Table of $\Delta_{p,q}$ -neighborhoods of increasing radius around a discrete set of ten equidistant points along the line $y = -x$ in \mathbb{R}^2 , showing how their shape change for different values of p and q .

3.3. Metric Properties

The extension of $\Delta_{p,q}$ to measurable sets given in Definition 6 preserves the nice metric properties of the finite version considered in [26] (Section 3). In particular, Theorem 1 enables us to show a result analogous to [26] (Theorem 3.3).

Theorem 2. Suppose that $1 \leq p, q < \infty$. Then the generational (p, q) -distance $\text{GD}_{p,q}$ satisfies the triangle inequality, namely

$$\text{GD}_{p,q}(X, Z) \leq \text{GD}_{p,q}(X, Y) + \text{GD}_{p,q}(Y, Z)$$

for any sets $X, Y, Z \in \mathfrak{M}_{<\infty}(\Sigma)$.

Proof. From the triangle inequality for the metric $d(\cdot, \cdot)$ we have

$$d_{x,z} \leq d_{x,y} + d_{y,z} \quad (x \in X, y \in Y, z \in Z).$$

Taking the q -average over Z at both sides and using Theorem 1 (i)–(iii), yields

$$\mathcal{M}_{z \in Z}^q(d_{x,z}) \leq \mathcal{M}_{z \in Z}^q(d_{x,y} + d_{y,z}) \leq \mathcal{M}_{z \in Z}^q(d_{x,y}) + \mathcal{M}_{z \in Z}^q(d_{y,z}) = d_{x,y} + \mathcal{M}_{z \in Z}^q(d_{y,z}). \quad (3)$$

Now, we consider two cases for the parameters $1 \leq p, q < \infty$, independently.

Case $p \leq q$: Taking at both sides of (3) the p -average over X and using Theorem 1 (i), (iii), and (iv), we get

$$\mathcal{M}_{x \in X}^p\left(\mathcal{M}_{z \in Z}^q(d_{x,z})\right) \leq \mathcal{M}_{x \in X}^p\left(d_{x,y} + \mathcal{M}_{z \in Z}^q(d_{y,z})\right) = \mathcal{M}_{x \in X}^p(d_{x,y}) + \mathcal{M}_{z \in Z}^q(d_{y,z}). \quad (4)$$

In this expression, the LHS is precisely $\text{GD}_{p,q}(X, Z)$ which does not depend on Y . We now take the p -average over Y at both sides of (4) and use Theorem 1 (i), (iii), and (iv), to obtain

$$\text{GD}_{p,q}(X, Z) \leq \mathcal{M}_{y \in Y}^p\left(\mathcal{M}_{x \in X}^p(d_{x,y}) + \mathcal{M}_{z \in Z}^q(d_{y,z})\right) = \mathcal{M}_{y \in Y}^p\left(\mathcal{M}_{x \in X}^p(d_{x,y})\right) + \text{GD}_{p,q}(Y, Z).$$

To finish this case note that from Theorem 1 (ii), (iv), and (v), we have that

$$\mathcal{M}_{y \in Y}^p\left(\mathcal{M}_{x \in X}^p(d_{x,y})\right) = \mathcal{M}_{x \in X}^p\left(\mathcal{M}_{y \in Y}^p(d_{x,y})\right) \leq \mathcal{M}_{x \in X}^p\left(\mathcal{M}_{y \in Y}^q(d_{x,y})\right) = \text{GD}_{p,q}(X, Y)$$

which proves the claim.

Case $q \leq p$: Here, we note that the LHS of (3) does not depend on Y , and take at both sides of (3) the q -average over Y . Hence, Theorem 1 (i), (iii)–(v) yield

$$\mathcal{M}_{z \in Z}^q(d_{x,z}) \leq \mathcal{M}_{y \in Y}^q\left(d_{x,y} + \mathcal{M}_{z \in Z}^q(d_{y,z})\right) \leq \mathcal{M}_{y \in Y}^q(d_{x,y}) + \text{GD}_{p,q}(Y, Z).$$

Lastly, we take the p -average over X and use Theorem 1 (ii)–(iv), to obtain

$$\text{GD}_{p,q}(X, Z) \leq \mathcal{M}_{x \in X}^p\left(\mathcal{M}_{y \in Y}^q(d_{x,y}) + \text{GD}_{p,q}(Y, Z)\right) = \text{GD}_{p,q}(X, Z) + \text{GD}_{p,q}(Y, Z),$$

which is the required result. \square

Corollary 1. For $p, q \in \mathbb{R}^{\times}$ the (p, q) -averaged Hausdorff distance $\Delta_{p,q}$ is a semimetric on the collection $\mathfrak{M}_{<\infty}(\Sigma)$ of all measurable subsets of Σ with finite measure. Moreover, between disjoint sets, $\Delta_{p,q}$ is a proper metric on $\mathfrak{M}_{<\infty}(\Sigma)$ for $1 \leq p, q < \infty$.

Proof. Definition 6 easily implies that $\Delta_{p,q}(\cdot, \cdot) \geq 0$ as well as $\Delta_{p,q}(X, Y) = \Delta_{p,q}(Y, X)$, for every pair $X, Y \in \mathfrak{M}_{<\infty}(\Sigma)$ and all $p, q \in \mathbb{R}^{\times}$. From Definition 5 we can see that $\text{GD}_{p,q}(X, Y \setminus X) = 0$ if and only if $X = \emptyset$ or $Y \subseteq X$ (and hence $Y \setminus X = \emptyset$). We thus find, for $X, Y \neq \emptyset$, that

$$\Delta_{p,q}(X, Y) = 0 \text{ if and only if } X = Y.$$

We have shown that $\Delta_{p,q}$ is a semimetric on $\mathfrak{M}_{<\infty}(\Sigma)$, and since the maximum of two functions satisfying the triangle inequality also satisfies it, Theorem 2 shows that $\Delta_{p,q}$ satisfies the triangle inequality when $1 \leq p, q < \infty$. \square

Theorem 3. Suppose that for any sets $X, Y, Z \in \mathfrak{M}_{<\infty}(\Sigma)$ there exist some constants $0 < r < R$ such that $r \leq d_{u,v} \leq R$ holds for all pairs (u, v) in $X \times Y$, $X \times Z$, or $Y \times Z$. Then, for all non-simultaneously positive $p, q \in \mathbb{R}^\times$ with $|p|, |q| \geq 1$ the generational (p, q) -distance $\text{GD}_{p,q}$ satisfies the following relaxed triangle inequality

$$\text{GD}_{p,q}(X, Z) \leq \frac{R^2}{r^2} \left(\text{GD}_{p,q}(X, Y) + \text{GD}_{p,q}(Y, Z) \right).$$

Proof. We prove the theorem in three steps.

Step 1: Take $p \in \mathbb{R}^\times$ and assume that $q < 0$, we will show that

$$\text{GD}_{p,|q|}(X, Y) \leq \frac{R}{r} \text{GD}_{p,q}(X, Y). \quad (5)$$

For any $x \in X$ and all $y_1, y_2 \in Y$ we have $\frac{r}{R} \leq \frac{d_{x,y_1}}{d_{x,y_2}} \leq \frac{R}{r}$, thus

$$\frac{R}{r} \geq \left(\int_Y \int_Y \left[\frac{d_{x,y_1}}{d_{x,y_2}} \right]^{|q|} dy_1 dy_2 \right)^{\frac{1}{|q|}} = \left(\int_Y d_{x,y_1}^{|q|} dy_1 \right)^{\frac{1}{|q|}} \left(\int_Y d_{x,y_2}^{-|q|} dy_2 \right)^{\frac{1}{|q|}}.$$

Using the fact that $q = -|q|$, we get

$$\left(\int_Y d_{x,y_1}^{|q|} dy_1 \right)^{\frac{1}{|q|}} \leq \frac{R}{r} \left(\int_Y d_{x,y_2}^q dy_2 \right)^{\frac{1}{q}},$$

which by (1), proves that $\mathcal{M}^{|q|}(d_{x,y}) \leq \frac{R}{r} \mathcal{M}^q(d_{x,y})$. Calculating the p -average \mathcal{M}^p of both sides, and from Theorem 1 (i) and (iv), we finally get $\mathcal{M}^p \left(\mathcal{M}^{|q|}(d_{x,y}) \right) \leq \frac{R}{r} \mathcal{M}^p \left(\mathcal{M}^q(d_{x,y}) \right)$, which, by Definition 5, is precisely (5).

Step 2: Now, take $q \in \mathbb{R}^\times$ and assume that $p < 0$, we will show that

$$\text{GD}_{|p|,q}(X, Y) \leq \frac{R}{r} \text{GD}_{p,q}(X, Y). \quad (6)$$

By hypothesis, for any $y \in Y$ and all $x_1, x_2 \in X$ we have $\frac{r}{R} \leq \frac{d_{x_1,y}}{d_{x_2,y}} \leq \frac{R}{r}$. Therefore, proceeding as before and applying again Theorem 1 (i) and (iv) we conclude that $\mathcal{M}^q(d_{x_1,y}) \leq \frac{R}{r} \mathcal{M}^q(d_{x_2,y})$. Hence,

$$\left(\int_X \left(\mathcal{M}^q(d_{x_1,y}) \right)^{|p|} dx_1 \right)^{\frac{1}{|p|}} \left(\int_X \left(\mathcal{M}^q(d_{x_2,y}) \right)^p dx_2 \right)^{\frac{1}{|p|}} = \left(\int_X \int_X \left[\frac{\mathcal{M}^q(d_{x_1,y})}{\mathcal{M}^q(d_{x_2,y})} \right]^{|p|} dx_1 dx_2 \right)^{\frac{1}{|p|}} \leq \frac{R}{r},$$

from which we deduce

$$\left(\int_X \left(\mathcal{M}^q(d_{x_1,y}) \right)^{|p|} dx_1 \right)^{\frac{1}{|p|}} \leq \frac{R}{r} \left(\int_X \left(\mathcal{M}^q(d_{x_2,y}) \right)^p dx_2 \right)^{\frac{1}{p}}.$$

Using (1), the previous inequality can be written as

$$\mathcal{M}^{|p|} \left(\mathcal{M}^q(d_{x,y}) \right) \leq \frac{R}{r} \mathcal{M}^p \left(\mathcal{M}^q(d_{x,y}) \right),$$

which, by Definition 5, is precisely (6).

Step 3: From the previous two steps we easily obtain

$$\text{GD}_{|p|,|q|}(X, Y) \leq \frac{R}{r} \text{GD}_{p,q}(X, Y) \leq \frac{R^2}{r^2} \text{GD}_{p,q}(X, Y). \quad (7)$$

Theorem 1 (iv) and Definition 5 imply that $\text{GD}_{p,q}(X, Z) \leq \text{GD}_{|p|,|q|}(X, Z)$. Finally, the triangle inequality for $\text{GD}_{|p|,|q|}$ (Theorem 2) and (7), produces the desired relation

$$\text{GD}_{p,q}(X, Z) \leq \text{GD}_{|p|,|q|}(X, Y) + \text{GD}_{|p|,|q|}(Y, Z) \leq \frac{R^2}{r^2} (\text{GD}_{p,q}(X, Y) + \text{GD}_{p,q}(Y, Z)). \quad \square$$

Remark 2. When the pair (p, q) lies in the light-gray or violet regions of Figure 2, the distance $\text{GD}_{p,q}$ satisfies a relaxed triangle inequality, with the drawback that the constant R^2/r^2 depends on the condition that $r \leq d_{u,v} \leq R$, for all pairs $(u, v) \in X \times Y, X \times Z$, or $Y \times Z$. For bounded and separated sets this condition always holds, and on those sets the associated (p, q) -averaged Hausdorff distance $\Delta_{p,q}$ becomes an inframetric as the following corollary implies.

Corollary 2. Under the same hypothesis of Theorem 3 we have

$$\Delta_{p,q}(X, Z) \leq \frac{R^2}{r^2} (\Delta_{p,q}(X, Y) + \Delta_{p,q}(Y, Z)).$$

Proof. It follows immediately from Theorem 3 and Definition 6. \square

Theorem 4. Let $X, Y \in \mathfrak{M}_{<\infty}(\Sigma)$ and suppose that $p, p', q, q' \in \overline{\mathbb{R}}$ satisfy $p \leq p'$ and $q \leq q'$. Then

$$\Delta_{p,q}(X, Y) \leq \Delta_{p',q}(X, Y) \quad \text{and} \quad \Delta_{p,q}(X, Y) \leq \Delta_{p,q'}(X, Y).$$

Proof. It follows easily from Theorem 1 (v) and Definition 6. \square

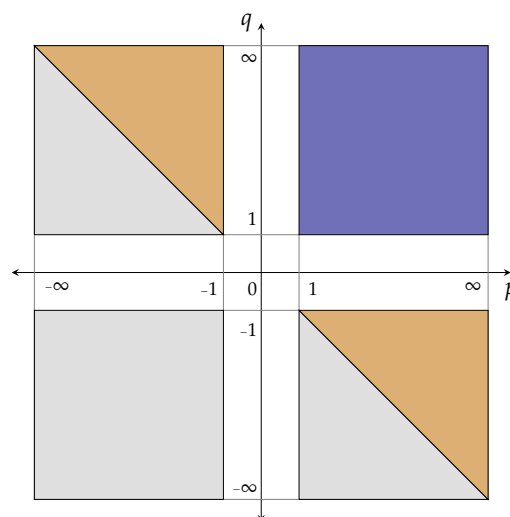


Figure 2. Representation of key regions on the (p, q) -plane. Corollary 1 shows that $\Delta_{p,q}$ is a proper metric in the violet region and Corollary 2 shows that it is an inframetric in the orange and light-gray ones. Numerical evidence suggests that $\Delta_{p,q}$ is still a proper metric in the orange regions.

3.4. Pareto-Compliance

We return now to the setting of MOPs to consider the behavior of the generalized $GD_{p,q}$ and $\Delta_{p,q}$ distances as performance indicators by studying their Pareto-compliance. A discussion of the Pareto-compliance for the indicators GD_p and Δ_p appeared in [14] (Section 3). Similar observations are valid for these new (p, q) -indicators, but a detailed and complete account of the details is part of ongoing research and will appear elsewhere. Here, as a first approach to the compliance question we present a basic result that describes the behavior of the indicator $GD_{p,q}$ under stronger assumptions than the compliance notion mentioned at the end of Section 2.2.

Let us assume that given a decision space $Q \subset \mathbb{R}^n$, a MOP has an associated objective function $F: Q \rightarrow \mathbb{R}^k$, with objective space $F(Q) \subset \mathbb{R}^k$ endowed with the Euclidean distance $d(\cdot, \cdot)$ and the inherited Lebesgue measure μ . Furthermore, let P_Q denote the Pareto set and $F(P_Q) \subset \mathbb{R}^k$ the corresponding Pareto front. If $X \subset Q$ denotes an approximating subset (or archive), the explicit $GD_{p,q}$ -performance indicators assigned to X is given by

$$\mathcal{I}_{p,q}^{\text{GD}}(X) := GD_{p,q}(F(X), F(P_Q)).$$

For the following statement, let us recall here that a partition of a set X is a collection of disjoint and non-empty subsets of X whose union is the whole of X . Furthermore, for any $q \in \overline{\mathbb{R}}$ we abbreviate the q -averaged distance of $F(u) \in F(Q)$ to the Pareto front $F(P_Q)$ by $\delta_q(u) := \mathcal{M}_q^q(d(F(u), F(v)))_{v \in P_Q}$.

Theorem 5. Suppose that for fixed $p, q \in \overline{\mathbb{R}}$ a pair of measurable archives $X, Y \subset Q$, satisfy that:

- (i) X and Y admit finite partitions $X = \bigsqcup_{i=1}^m X_i$ and $Y = \bigsqcup_{i=1}^m Y_i$ such that for each $i \in \{1, \dots, m\}$:
 - (a) $X_i \subset X$ and $Y_i \subset Y$ are subsets of non-null finite measure.
 - (b) $\forall x \in X_i, \forall y \in Y_i: x \preceq y$,
- (ii) $\forall x \in X, \forall y \in Y: \text{if } x \preceq y \implies \delta_q(x) \leq \delta_q(y)$.

Then $\mathcal{I}_{p,q}^{\text{GD}}(X) \leq \mathcal{I}_{p,q}^{\text{GD}}(Y)$.

Proof. From condition (i) the archives X and Y admit partitions into the same number m of subsets and from (ii) it is clear that for any $i \in \{1, \dots, m\}$ if $x \in X_i$ and $y \in Y_i$ then $\delta_q(x) \leq \delta_q(y)$. Hence, taking integral p -averages over X_i , and then over Y_i of the quantities at both sides of this inequality we obtain for each i that

$$a_i^p := \frac{1}{|X_i|} \int_{X_i} \delta_q(x)^p dx \leq \frac{1}{|Y_i|} \int_{Y_i} \delta_q(y)^p dy =: b_i^p. \quad (8)$$

Now, for each $i \in \{1, \dots, m\}$ for which the inequality $\frac{|X_i|}{|X|} \leq \frac{|Y_i|}{|Y|}$ does not hold, we can further subdivide X_i into a sufficiently large partition of m_i non-null finite measure subsets $X_{i,1}, X_{i,2}, \dots, X_{i,m_i}$, so that for all $j \in \{1, \dots, m_i\}$ we can guarantee that

$$w_{i,j} := \frac{|X_{i,j}|}{|X|} \leq \frac{|Y_i|}{|Y|} =: \tilde{w}_i. \quad (9)$$

Please note that this should be possible due to the assumption that X_i has non-null finite measure. Since part (b) of condition (i) still holds for these subsets, (i.e., $\forall x \in X_{i,j}, \forall y \in Y_i: x \preceq y$), an analogous relation to Inequality (8) is valid for them. Explicitly, for each $i \in \{1, \dots, n\}$ and all $j \in \{1, \dots, m_i\}$ we have

$$a_{i,j}^p := \frac{1}{|X_{i,j}|} \int_{X_{i,j}} \delta_q(x)^p dx \leq \frac{1}{|Y_i|} \int_{Y_i} \delta_q(y)^p dy =: b_i^p.$$

Due to the chosen partitions of X and Y , it is clear that $|X| = \sum_{i=1}^m |X_i|$, where $|X_i| = \sum_{j=1}^{m_i} |X_{i,j}|$, and $|Y| = \sum_{i=1}^m |Y_i|$. Therefore, with the notation of (9) it follows $\sum_{i=1}^m \sum_{j=1}^{m_i} w_{i,j} = \sum_{i=1}^m \tilde{w}_i = 1$,

which implies that the quantities $w_{i,j}$ and \tilde{w}_i can be regarded as normalized weights appropriate for taking weighted averages. Using that $0 \leq a_{i,j} \leq b_i$ and $0 \leq w_{i,j} \leq \tilde{w}_i \leq 1$, simple properties of (discrete) weighted power means ensure that $\sum_{i=1}^m \sum_{j=1}^{m_i} w_{i,j} a_{i,j}^p \leq \sum_{i=1}^m \tilde{w}_i b_i^p$. Thus, we can finally write

$$\begin{aligned} \mathcal{I}_{p,q}^{\text{GD}}(Y)^p &= \frac{1}{|X|} \sum_{i=1}^m \sum_{j=1}^{m_i} \int_{X_{i,j}} \delta_q(x)^p dx = \sum_{i=1}^m \sum_{j=1}^{m_i} \frac{|X_{i,j}|}{|X|} a_{i,j}^p = \sum_{i=1}^m \sum_{j=1}^{m_i} w_{i,j} a_{i,j}^p \\ &\leq \sum_{i=1}^m \tilde{w}_i b_i^p = \sum_{i=1}^m \frac{|Y_i|}{|Y|} b_i^p = \frac{1}{|Y|} \sum_{i=1}^m \int_{Y_i} \delta_q(y)^p dy = \mathcal{I}_{p,q}^{\text{GD}}(Y)^p, \end{aligned}$$

proving the statement. \square

Remark 3. Condition (i) of Theorem 5 implies the simpler (and weaker) dominance conditions:

- (a') $X \preceq Y$ (i.e., $\forall y \in Y, \exists x \in X$ such that $x \preceq y$), and
- (b') $\forall x \in X, \exists y \in Y$ such that $x \preceq y$.

In many simple examples for which (a') and (b') hold, it is not difficult to find the partitions needed for Theorem 5 (i), however this is not always possible, and the question of when such partitions exist will not be considered here. Figure 3, show examples where (a') and (b') hold and the inequality $\mathcal{I}_{p,q}^{\text{GD}}(X) \leq \mathcal{I}_{p,q}^{\text{GD}}(Y)$ is both, true (left) and false (right). In these cases it can be shown that X and Y satisfy (left), and do not satisfy (right) the requirements of Theorem 5 (i), respectively.

Remark 4. Another important observation is that condition (ii) of Theorem 5 allows for some freedom in the choice of an appropriate $q \in \overline{\mathbb{R}}$ such that the inequality $\delta_q(x) \leq \delta_q(y)$ holds for $x \preceq y$, ensuring the compliance to Pareto optimality. This freedom is not available for the indicator GD_p because in that case $\delta_q(x)$ should be replaced by the corresponding quantity when $q \rightarrow \infty$ which is the standard distance from a set to a point $d(F(x), F(P_Q))$. The possibility to choose a value of q according to the problem is clearly an advantage, and provides an argument in favor of the generalized version $\text{GD}_{p,q}$.

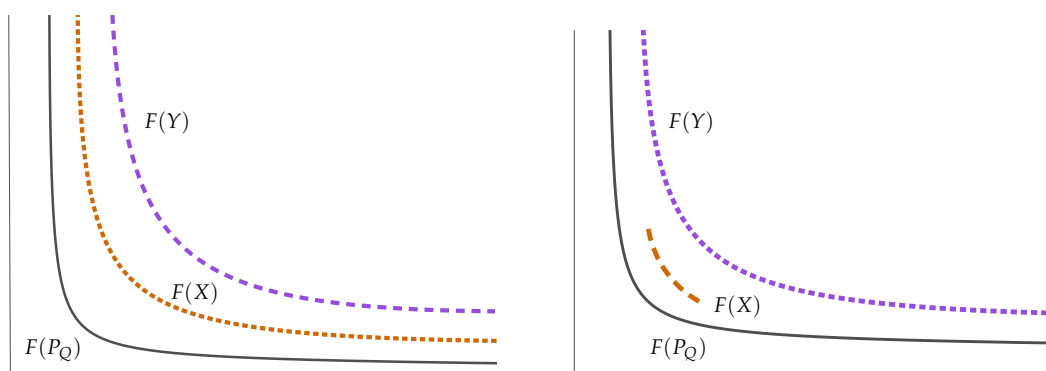


Figure 3. (Left) Example of a Pareto front $F(P_Q)$ with two archives satisfying condition (i) of Theorem 5 for which $\mathcal{I}_{p,q}^{\text{GD}}(X) \leq \mathcal{I}_{p,q}^{\text{GD}}(Y)$. (Right) Example of a Pareto front $F(P_Q)$ with two archives satisfying conditions (a') and (b') of Remark 3 but for which $\mathcal{I}_{p,q}^{\text{GD}}(X) \not\leq \mathcal{I}_{p,q}^{\text{GD}}(Y)$. In this case partitions of the archives satisfying Theorem 5 (i) do not exist.

4. Numerical Examples

In this section, we demonstrate the applicability and usefulness of the new distance measure on two examples.

4.1. General Example

As a first example we consider the following sets within the Euclidean plane \mathbb{R}^2 : the first set, A , is a line segment connecting two points $a = (-1, 0)$ and $b = (1, 0)$, i.e.,

$$A = \overline{ab}. \quad (10)$$

Next, for some given $\delta > 0$ and any fixed value of $\varepsilon > 0$ we consider sets B_δ defined as the union of line segments

$$B_\delta = \overline{cd_\delta} \cup \overline{e_\delta f_\delta} \cup \overline{g_\delta h}, \quad (11)$$

where $c = (-1, \varepsilon)$, $d_\delta = (-\delta, \varepsilon)$, $e_\delta = (-\delta, 1)$, $f_\delta = (\delta, 1)$, $g_\delta = (\delta, \varepsilon)$, and $h = (1, \varepsilon)$ are the segment end-points in \mathbb{R}^2 . Hereby, a set B_δ can be seen as a certain approximation of A , where the segment $\overline{e_\delta f_\delta}$ can be considered to be the outlier in the approximation.

Figure 4 shows the sets A and B_δ for the values $\delta \in \{0.05, 0.10, 0.20, 0.40\}$ and $\varepsilon = 0.10$. Apparently, for smaller δ , the outlier region gets smaller, and hence, the approximation B_δ of A gets “better”. This is reflected by the values of the (p, q) -distance in Table 1.

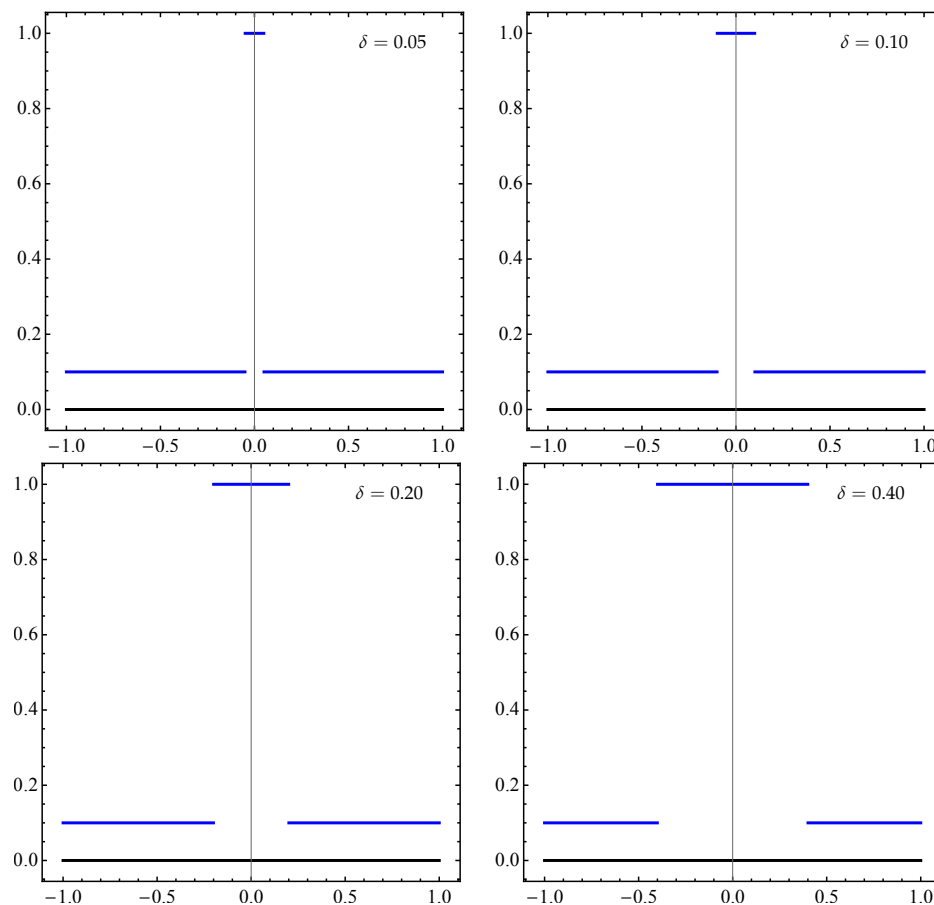


Figure 4. Example of four approximations of A (black horizontal segment) with B_δ (blue piecewise function) for four different values of δ and fixed $\varepsilon = 0.10$.

Table 1. $\Delta_{p,q}$ values for A and B_δ in (10) and (11), for different values of p , q , and δ , with fixed $\varepsilon = 0.10$.

p	q	$\Delta_{pq}(A, B_{0.05})$	$\Delta_{pq}(A, B_{0.10})$	$\Delta_{pq}(A, B_{0.20})$	$\Delta_{pq}(A, B_{0.40})$
1	1	0.7149	0.7464	0.8091	0.9324
1	−1	0.4105	0.4506	0.5311	0.6945
1	−100	0.1503	0.1961	0.2878	0.4711
1	−200	0.1479	0.1934	0.2844	0.4663
1	−10,000	0.1451	0.1901	0.2802	0.4602

On the other hand, if choosing the classical Hausdorff distance, all values of $d_H(A, B_\delta)$ are equal to 1, regardless of the choice of $\delta > 0$. Hence, the (p, q) -distance is more appropriate in this example to identify “better” approximations.

4.2. Approximation of Pareto Sets/Fronts

As a second example we consider the approximation of the Pareto set and front of a given MOP. For this, we define the following bi-objective problem that is known as the *Lamé super-sphere function* [32]:

$$\min_x F: \mathbb{R}^n \rightarrow \mathbb{R}^2, \quad (12)$$

where $F(x) = (f_1(x), f_2(x))$ is given by

$$f_1(x) = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right)^{\frac{\gamma}{2}} \quad \text{and} \quad f_2(x) = \left(\frac{1}{n} \sum_{i=1}^n (x_i - 1)^2 \right)^{\frac{\gamma}{2}}$$

for $x \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}$. Figures 5 and 6 show the Pareto sets and fronts for the special cases $n = 2$ with $\gamma = 2$ and $\gamma = 1/2$, respectively.

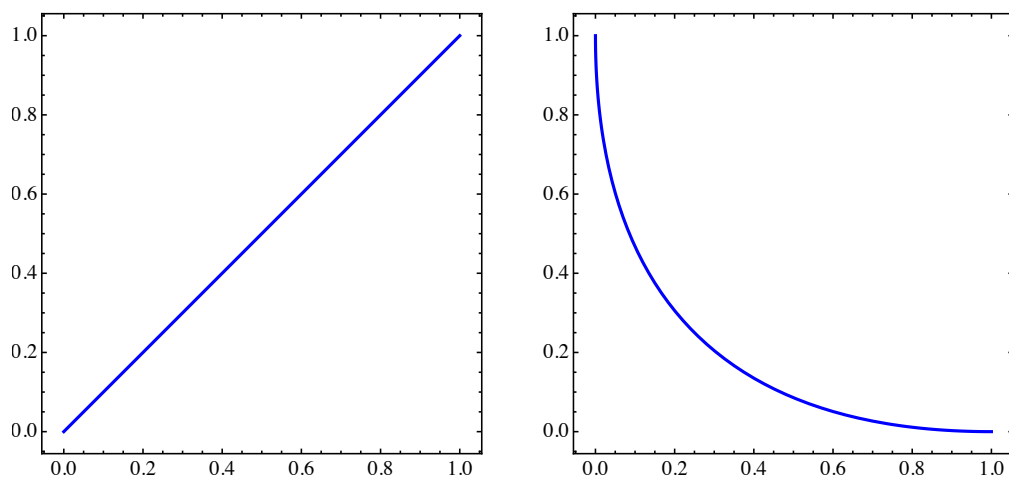


Figure 5. Pareto set (left) and front (right) of MOP (12) for $n = 2$ with $\gamma = 2$.

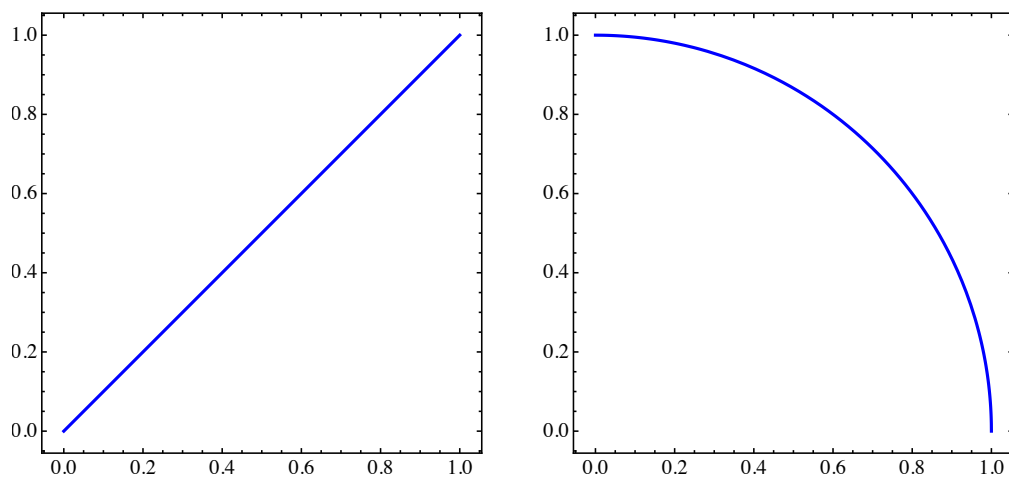


Figure 6. Pareto set (left) and front (right) of MOP (12) for $n = 2$ with $\gamma = 1/2$.

For the first step, we consider a simple hypothetical example to illustrate the concept of continuous archives in the context of evolutionary multi-objective optimization. For these, assume we are given the discrete archive $A = \{x_1, \dots, x_5\} \subset \mathbb{R}^2$, where

$$x_1 = (-0.0129, -0.0421), \quad x_2 = (0.2525, 0.2912), \quad x_3 = (0.4903, 0.4035), \\ x_4 = (0.6258, 0.6912), \quad x_5 = (1.0212, 0.9930).$$

The set A is hence consisting of only five candidate solutions. Analogously, the image $F(A)$ of A can be considered as an approximation of the Pareto front that consists as well of five candidate solutions. Now, in order to improve the quality of the approximation, instead of A one can consider the polygon that is defined by the elements of A , namely

$$B := \overline{x_1 x_2} \cup \dots \cup \overline{x_4 x_5}. \quad (13)$$

In what follows, we will call this polygon *the continuous archive*. The approximations A , B , $F(A)$, and $F(B)$ can be seen in the Figures 7 and 8. By visual inspection, the approximation qualities increase significantly when using the linear interpolation, in particular in objective space. This is reflected by the (p, q) -distances which are shown in Table 2 where we can find the following general behavior: first, the distances within decision and objective spaces, decreases from finite to continuous archives, and this phenomena is stronger in the objective space; and second, following the result of Theorem 4, the distances decreases as q decreases.

In a next step, we consider discrete and continuous archives resulting from two of the most famous EMO algorithms: NSGA-II [44] and MOEA/D [45], see Table 3 for the parameter setting of these algorithms. To this end, we first consider the result of NSGA-II with a population size of 12 after 500 generations, see Figures 9 and 10 and Table 4 for the numerical results. Finally, we consider the MOEA/D generational algorithm to get 500 finite archives of 12 elements each, see Figures 11 and 12 and Table 5 for the numerical results.

For both EMO algorithms, it can be observed that the indicator values for the continuous archives are significantly better than for the respective discrete archives. Next, note that the $\Delta_{p,q}$ values oscillate for NSGA-II which is a typical behavior for this dominance-based algorithm. These oscillations, however, are less distinct for the continuous archives.

Table 2. $\Delta_{p,q}$ values for the Pareto set/front approximations for MOP (12).

	p	q	Decision Space		Objective Space	
			Finite Archive	Continuous Archive	Finite Archive	Continuous Archive
$\gamma = 2$	1	1	0.5314	0.4841	0.4369	0.3943
	1	−1	0.2732	0.1750	0.2095	0.0945
	1	−100	0.1140	0.0213	0.0893	0.0018
	1	−200	0.1131	0.0208	0.0886	0.0017
	1	−10,000	0.1122	0.0202	0.0879	0.0017
$\gamma = \frac{1}{2}$	1	1	0.5314	0.4841	0.5629	0.5024
	1	−1	0.2732	0.1750	0.2807	0.1072
	1	−100	0.1140	0.0213	0.1202	0.0015
	1	−200	0.1131	0.0208	0.1192	0.0015
	1	−10,000	0.1122	0.0202	0.1183	0.0014

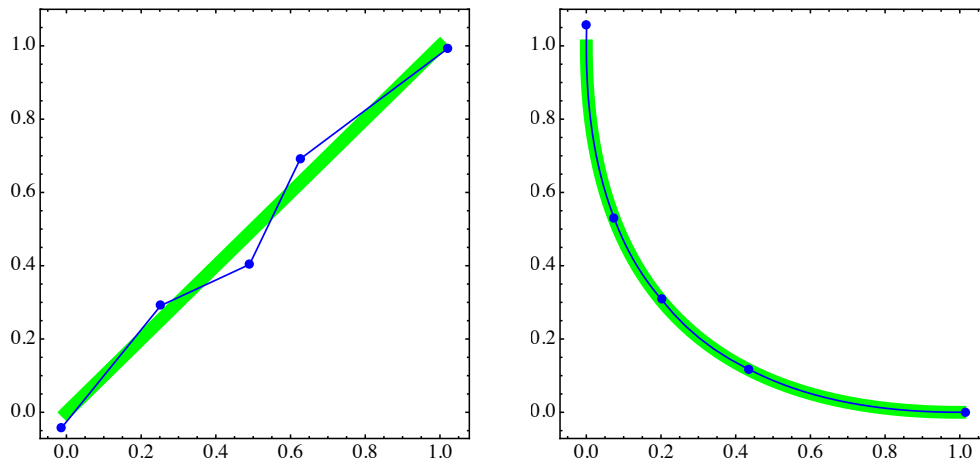


Figure 7. Left: Approximations A (blue dots) and B (blue polygon line) of the Pareto set (green thick line) of MOP (12) for $n = 2$. Right: corresponding approximations $F(A)$ and $F(B)$ of the Pareto front, for $\gamma = 2$.

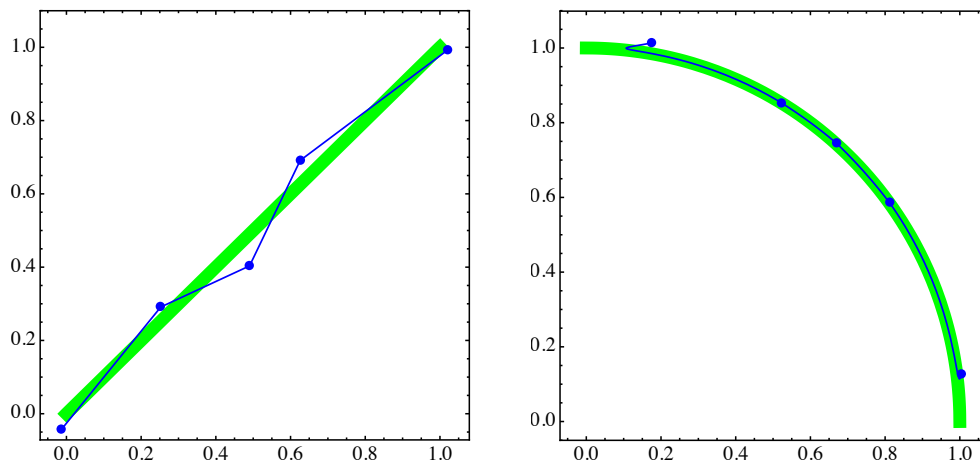


Figure 8. Left: Approximations A (blue dots) and B (blue polygon line) of the Pareto set (green thick line) of MOP (12) for $n = 2$. Right: corresponding approximations $F(A)$ and $F(B)$ of the Pareto front, for $\gamma = 1/2$.

Table 3. Parameter setting for NSGA-II and MOEA/D.

Algorithm	Parameter	Value
NSGA-II	Population size	12
	Number of generations	500
	Crossover probability	0.8
	Mutation probability	$1/n$
	Distribution index for crossover	20
	Distribution index for mutation	20
MOEA/D	Population size	12
	# weight vectors	12
	Number of generations	500
	Crossover probability	1
	Mutation probability	$1/n$
	Distribution index for crossover	30
	Distribution index for mutation	20
	Aggregation function	Tchebycheff
	Neighborhood size	3

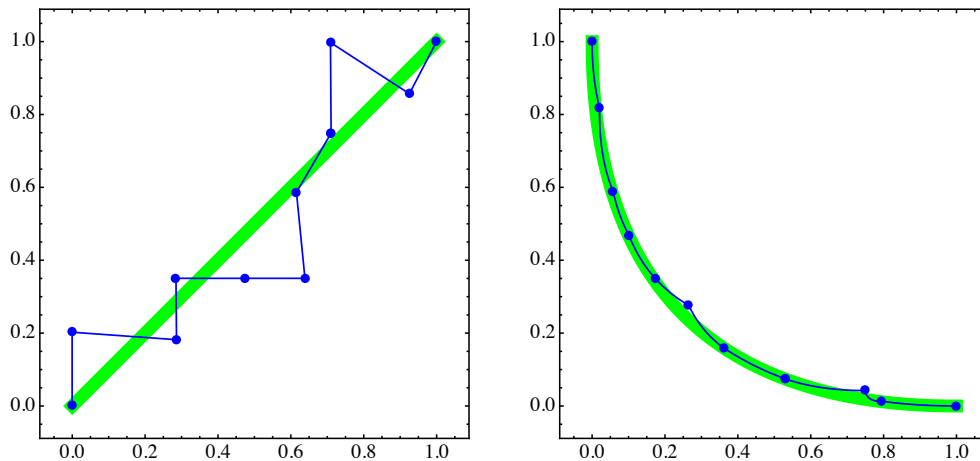


Figure 9. Left: approximations A (blue dots) corresponding to the 500th generation of the NSGA-II algorithm, and B (blue polygon line) of the Pareto set (green thick line) of MOP (12) for $n = 2$. Right: respective approximations $F(A)$ and $F(B)$ of the Pareto front for $\gamma = 2$.

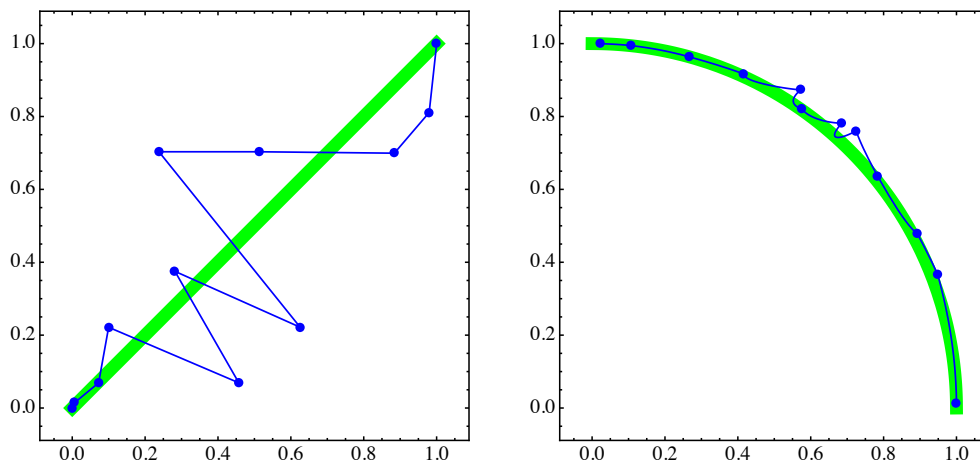


Figure 10. Left: approximations A (blue dots) corresponding to the 500th generation of the NSGA-II algorithm, and B (blue polygon line) of the Pareto set (green thick line) of MOP (12) for $n = 2$. Right: respective approximations $F(A)$ and $F(B)$ of the Pareto front for $\gamma = 1/2$.

Table 4. $\Delta_{p,q}$ values for the Pareto front approximations for MOP (12) using the NSGA-II archives and with $p = 1, q = -10$.

Generation	$\gamma = 1/2$		$\gamma = 2$	
	Finite Archive	Continuous Archive	Finite Archive	Continuous Archive
50	0.0439	0.0147	0.0696	0.0160
100	0.0498	0.0109	0.0540	0.0102
200	0.0613	0.0118	0.0716	0.0207
250	0.0651	0.0265	0.0572	0.0061
400	0.0602	0.0102	0.0723	0.0276
450	0.0630	0.0154	0.0584	0.0088
460	0.0612	0.0154	0.0658	0.0098
470	0.0523	0.0102	0.0566	0.0083
480	0.0754	0.0269	0.0684	0.0241
490	0.0510	0.0091	0.0584	0.0118
500	0.0722	0.0097	0.0560	0.0103

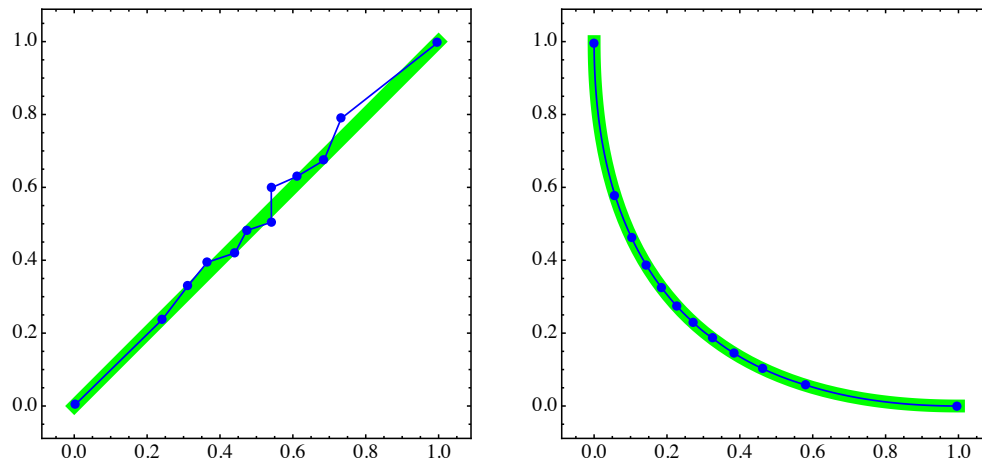


Figure 11. Left: approximations A (blue dots) corresponding to the 500th generation of the MOEA/D algorithm), and B (blue polygon line) of the Pareto set (green thick line) of MOP (12) for $n = 2$. Right: respective approximations $F(A)$ and $F(B)$ of the Pareto front for $\gamma = 2$.

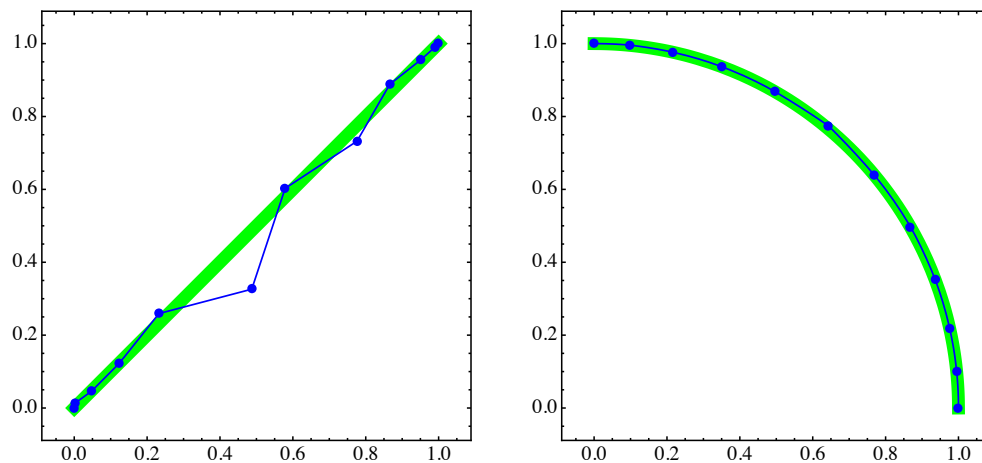


Figure 12. Left: approximations A (blue dots) corresponding to the 500th generation of the MOEA/D algorithm), and B (blue polygon line) of the Pareto set (green thick line) of MOP (12) for $n = 2$. Right: respective approximations $F(A)$ and $F(B)$ of the Pareto front for $\gamma = 1/2$.

Table 5. $\Delta_{p,q}$ values for the Pareto front approximations for MOP (12) using the MOEA/D archives and with $p = 1, q = -10$.

Generation	$\gamma = 1/2$		$\gamma = 2$	
	Finite Archive	Continuous Archive	Finite Archive	Continuous Archive
50	0.0610	0.0171	0.0648	0.0119
100	0.0519	0.0051	0.1093	0.0016
200	0.0536	0.0037	0.0781	0.0009
250	0.0522	0.0037	0.0790	0.0008
400	0.0511	0.0017	0.0784	0.0009
450	0.0511	0.0017	0.0784	0.0009
460	0.0509	0.0012	0.0784	0.0009
470	0.0509	0.0012	0.0784	0.0009
480	0.0509	0.0010	0.0783	0.0009
490	0.0509	0.0010	0.0783	0.0009
500	0.0509	0.0010	0.0783	0.0009

To further investigate the last statement, we consider finally the convex bi-objective problem $f_1, f_2: \mathbb{R}^3 \rightarrow \mathbb{R}$, where $x = (x_1, x_2, x_3)$ and

$$\begin{aligned} f_1(x) &= (x_1 + 1)^2 + x_2^2 + x_3^2 \\ f_2(x) &= (x_1 - 1)^2 + x_2^2 + x_3^2. \end{aligned} \quad (14)$$

The Pareto set of MOP (14) is the line segment connecting the points $(0, 0, 0)$ and $(1, 0, 0)$, and the Pareto front is as shown in Figure 13.

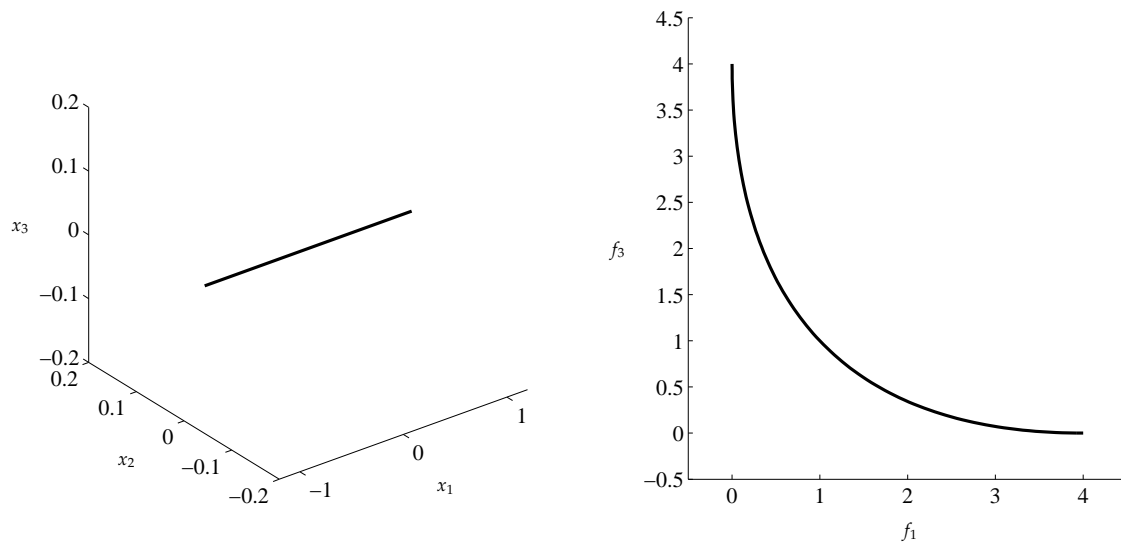


Figure 13. Pareto set (left) and front (right) of MOP (14).

Figure 14 and Table 6 show the $\Delta_{p,q}$ values for both the discrete and continuous archives obtained by NSGA-II using a population size of 20. As it can be seen, again the continuous archives achieve much better indicator values, and the amplitudes of the oscillations are significantly smaller compared to the discrete archives. This is confirmed by Figures 15–17 that show the results of the discrete and continuous archives after 300, 400, and 500 generations, respectively. As it can be seen, NSGA-II is able to compute solutions along the Pareto front, however, with varying distribution along this set (In fact, it is known that there is no “limit archive” for NSGA-II since this algorithm is not indicator-based). In turn, for each of the results of NSGA-II, all of the continuous archives represent—at least by visual inspection—perfect approximations of the Pareto front, which is reflected by the good $\Delta_{p,q}$ values.

Concluding, the results presented in this section strongly indicate the convenience of the new indicator that is able to assess the performance of continuous archives. Though in principle also other indicators can be extended to continuous sets, this has not been done so far, and this is not a straightforward task. Hence, no comparisons to other indicators can be considered here. The presented results further indicate the benefit of the use of continuous archives instead of discrete ones that are being used classically. This would, among others, allow for the usage of smaller population sizes which would in turn allow to reduce the computational burden of the evolutionary algorithms (note that the time complexity for all MOEAs in each generation is quadratic in the population size). The verification of this statement, however, is left for future work as this goes beyond the scope of this study.

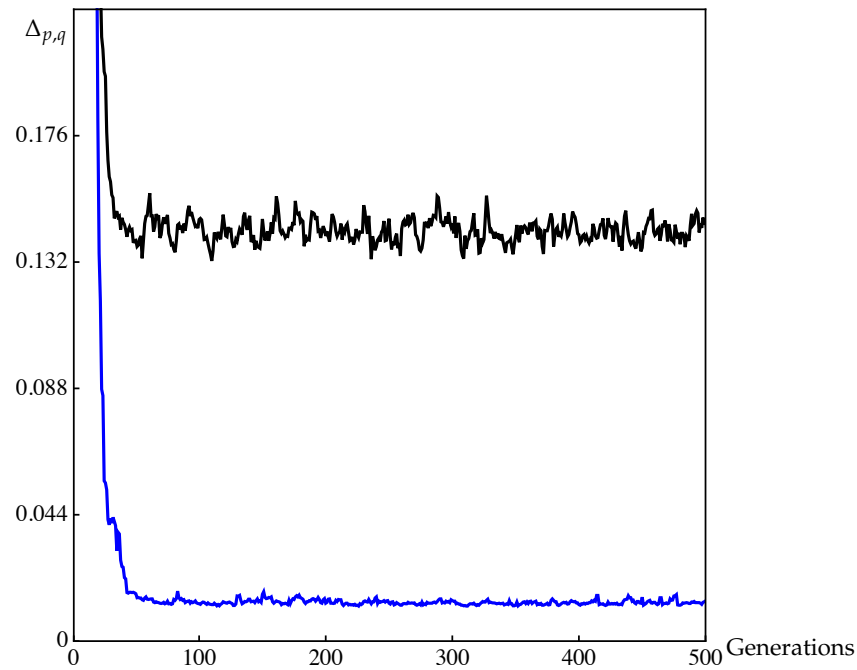


Figure 14. $\Delta_{p,q}$ values for the discrete (black curve) and the continuous archives (blue curve) of NSGA-II for MOP (14).

Table 6. $\Delta_{p,q}$ values for the discrete and continuous archives of NSGA-II for MOP (14). The results are averaged over 20 independent runs.

Generation	Continuous Archive	Finite Archive
20	0.1333	0.2401
40	0.0176	0.1451
60	0.0090	0.1561
80	0.0088	0.1355
100	0.0065	0.1472
120	0.0074	0.1412
140	0.0081	0.1395
160	0.0075	0.1549
180	0.0092	0.1468
200	0.0074	0.1429
220	0.0066	0.1408
240	0.0075	0.1397
260	0.0066	0.1460
280	0.0074	0.1439
300	0.0084	0.1421
320	0.0070	0.1352
340	0.0070	0.1373
360	0.0081	0.1454
380	0.0079	0.1413
400	0.0066	0.1388
420	0.0063	0.1400
440	0.0097	0.1384
460	0.0067	0.1418
480	0.0067	0.1421
500	0.0076	0.1426

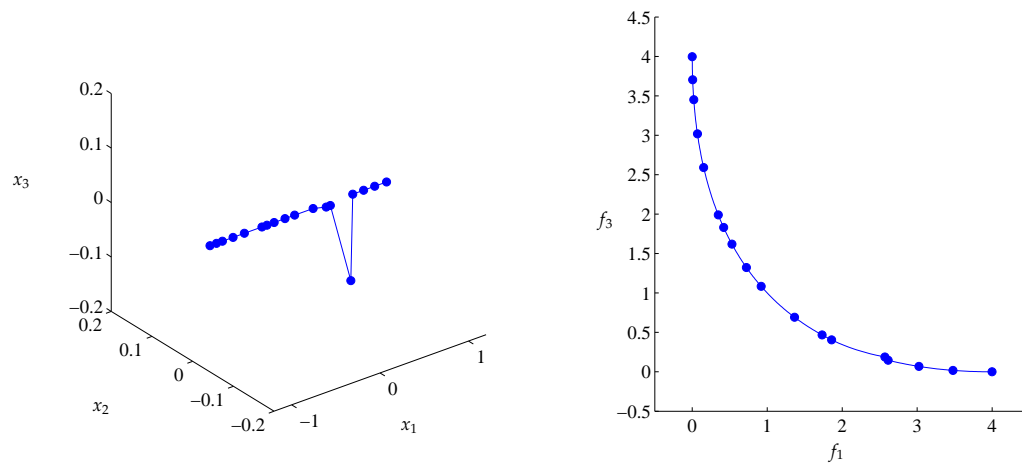


Figure 15. Left: Approximations A (blue dots) and B (blue continuous polygon line) of the Pareto set of MOP (14) in the 300th generation. Right: corresponding approximations $F(A)$ and $F(B)$ of the Pareto front.

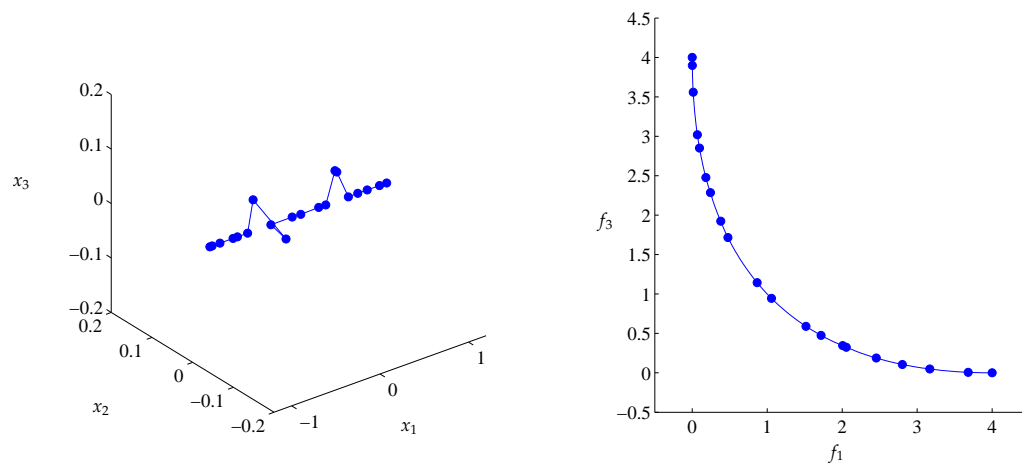


Figure 16. Left: Approximations A (blue dots) and B (blue continuous polygon line) of the Pareto set of MOP (14) in the 400th generation. Right: corresponding approximations $F(A)$ and $F(B)$ of the Pareto front.

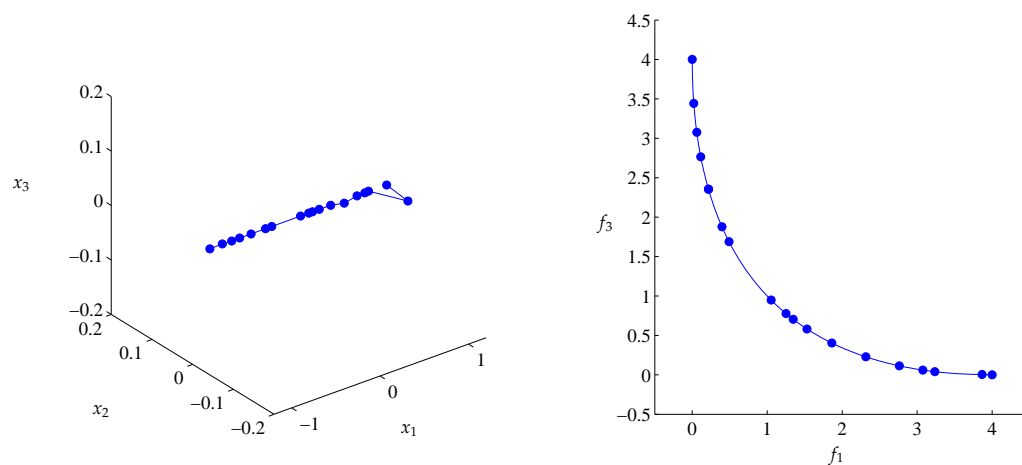


Figure 17. Left: Approximations A (blue dots) and B (blue continuous polygon line) of the Pareto set of MOP (14) in the 500th generation. Right: corresponding approximations $F(A)$ and $F(B)$ of the Pareto front.

5. Conclusions and Future Work

In this work, we have proposed extensions of the existing $GD_{p,q}$ and $\Delta_{p,q}$ performance indicators that allow to compute the distance between two general measurable sets. In particular, this is a natural setting in multi-objective optimization because the solution of such a problem typically forms a set of certain dimension (and is thus not given by finitely many points). We have shown that the extended indicators keep the nice metric properties from its finite-version predecessors (see [14,26]). Moreover, for $GD_{p,q}$, sufficient conditions have been provided ensuring that certain compliance to Pareto optimality of this indicator can be guaranteed. Further study is needed to determine the precise relation between these conditions and other ones appearing in the literature.

We have demonstrated the applicability and usefulness of the novel indicator on examples related to evolutionary multi-objective optimization.

As part of future work, we intend to further investigate the use of $\Delta_{p,q}$ within evolutionary multi-objective optimization. For instance, it might be interesting to integrate this performance indicator within an evolutionary multi-objective optimization algorithm as it was done, e.g., in [46] for its predecessor Δ_p . Although it is clear that the individual roles of p and q are related with the convexity of the metric neighborhoods of point and sets, further research is needed to elucidate more precisely useful ways to take advantage of their joint behavior in concrete situations. Additionally, to understand the behavior of $\Delta_{p,q}$ in relation to Pareto compliance and to complete the partial results that have been established in Section 3.4 for $GD_{p,q}$, consideration should be given to the inverted generational indicator $IGD_{p,q}$. Finally, one interesting aspect is to see if the indicator can be used as a proximity measure in other research fields.

Author Contributions: J.B. and A.V. obtained the theoretical results concerning the (p, q) -averaged Hausdorff distance. O.S. and O.C. conceived and designed the experiments; J.B. and O.C. performed the experiments and provided the related figures and tables; O.S. analyzed the data and revised the text. J.B. and A.V. wrote the paper.

Acknowledgments: We would like to thank the referees for the for their useful comments and suggestions that helped to improve our manuscript. The first two authors were partially supported by the research project ID-PRY: 7919 of the Faculty of Sciences, Pontificia Universidad Javeriana, Bogotá, Colombia. The last two authors acknowledge support from the Conacyt Basic Science Group project no. 285599.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Heinonen, J. *Lectures on Analysis on Metric Spaces*; Springer: New York, NY, USA, 2001.
2. Huttenlocher, D.P.; Klanderman, G.A.; Rucklidge, W.A. Comparing Images Using the Hausdorff Distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **1993**, *15*, 850–863. [[CrossRef](#)]
3. Yi, X.; Camps, O.I. Line-Based Recognition Using A Multidimensional Hausdorff Distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 901–916.
4. De Carvalho, F.; de Souza, R.; Chavent, M.; Lechevallier, Y. Adaptive Hausdorff distances and dynamic clustering of symbolic interval data. *Pattern Recognit. Lett.* **2006**, *27*, 167–179. [[CrossRef](#)]
5. Dellnitz, M.; Hohmann, A. A subdivision algorithm for the computation of unstable manifolds and global attractors. *Numerische Mathematik* **1997**, *75*, 293–317. [[CrossRef](#)]
6. Aulbach, B.; Rasmussen, M.; Siegmund, S. Approximation of attractors of nonautonomous dynamical systems. *Discret. Contin. Dyn. Syst. Ser. B* **2005**, *5*, 215–238.
7. Emmerich, M.; Deutz, A.H. Test Problems Based on Lamé Superspheres. In Proceedings of the 4th International Conference on Evolutionary Multi-criterion Optimization, Matsushima, Japan, 5–8 March 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 922–936.
8. Falconer, K. *Fractal Geometry: Mathematical Foundations and Applications*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2003.
9. Schütze, O. Set Oriented Methods for Global Optimization. Ph.D. Thesis, University of Paderborn, Paderborn, Germany, 2004.

10. Dellnitz, M.; Schütze, O.; Hestermeyer, T. Covering Pareto Sets by Multilevel Subdivision Techniques. *J. Optim. Theory Appl.* **2005**, *124*, 113–155. [\[CrossRef\]](#)
11. Padberg, K. Numerical Analysis of Transport in Dynamical Systems. Ph.D. Thesis, University of Paderborn, Paderborn, Germany, 2005.
12. Schütze, O.; Coello Coello, C.A.; Mostaghim, S.; Talbi, E.G.; Dellnitz, M. Hybridizing Evolutionary Strategies with Continuation Methods for Solving Multi-Objective Problems. *Eng. Optim.* **2008**, *40*, 383–402. [\[CrossRef\]](#)
13. Schütze, O.; Laumanns, M.; Coello Coello, C.A.; Dellnitz, M.; Talbi, E.G. Convergence of Stochastic Search Algorithms to Finite Size Pareto Set Approximations. *J. Glob. Optim.* **2008**, *41*, 559–577. [\[CrossRef\]](#)
14. Schütze, O.; Esquivel, X.; Lara, A.; Coello Coello, C.A. Using the averaged Hausdorff distance as a performance measure in evolutionary multiobjective optimization. *IEEE Trans. Evol. Comput.* **2012**, *16*, 504–522. [\[CrossRef\]](#)
15. Hernández, C.; Naranjani, Y.; Sardahi, Y.; Liang, W.; Schütze, O.; Sun, J.Q. Simple Cell Mapping Method for Multi-objective Optimal Feedback Control Design. *Int. J. Dyn. Control* **2013**, *1*, 231–238. [\[CrossRef\]](#)
16. Siwel, J.; Yew-Soon, O.; Jie, Z.; Liang, F. Consistencies and contradictions of performance metrics in multiobjective optimization. *IEEE Trans. Evol. Comput.* **2014**, *44*, 2329–2404.
17. Sun, J.Q.; Xiong, F.R.; Schütze, O.; Hernández, C. *Cell Mapping Methods—Algorithmic Approaches and Applications*; Springer: Singapore, 2019.
18. Jahn, J. Multiobjective search algorithm with subdivision technique. *Comput. Optim. Appl.* **2006**, *35*, 161–175. [\[CrossRef\]](#)
19. Schütze, O.; Vasile, M.; Junge, O.; Dellnitz, M.; Izzo, D. Designing optimal low thrust gravity assist trajectories using space pruning and a multi-objective approach. *Eng. Optim.* **2009**, *41*, 155–181. [\[CrossRef\]](#)
20. Deb, K. *Multi-Objective Optimization Using Evolutionary Algorithms*; John Wiley & Sons: Chichester, UK, 2001.
21. Beume, N.; Naujoks, B.; Emmerich, M. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *Eur. J. Oper. Res.* **2007**, *181*, 1653–1669. [\[CrossRef\]](#)
22. Garg, H. Solving structural engineering design optimization problems using an artificial bee colony algorithm. *J. Ind. Manag. Optim.* **2014**, *10*, 777–794. [\[CrossRef\]](#)
23. Garg, H. A hybrid PSO-GA algorithm for constrained optimization problems. *Appl. Math. Comput.* **2016**, *274*, 292–305. [\[CrossRef\]](#)
24. Zapotecas-Martínez, S.; López-Jaimes, A.; García-Nájera, A. LIBEA: A Lebesgue Indicator-Based Evolutionary Algorithm for Multi-objective Optimization. *Swarm Evolut. Comput.* **2018**. [\[CrossRef\]](#)
25. Hartikainen, M.; Miettinen, K.; Wiecek, M. PAINT: Pareto front interpolation for nonlinear multiobjective optimization. *Comput. Optim. Appl.* **2012**, *52*, 845–867. [\[CrossRef\]](#)
26. Vargas, A.; Bogoya, J.M. A generalization of the averaged Hausdorff distance. *Computación y Sistemas* **2018**, *22*, 331–345. [\[CrossRef\]](#)
27. Tao, T. *An Introduction to Measure Theory (Graduate Studies in Mathematics)*; American Mathematical Society: Providence, RI, USA, 2011.
28. Jones, F. *Lebesgue Integration on Euclidean Space*; Jones and Bartlett Publishers: Boston, MA, USA, 2001.
29. Hardy, G.H.; Littlewood, J.E.; Pólya, G. *Inequalities*, 2nd ed.; Cambridge University Press: Cambridge, UK, 1952.
30. Bullen, P.S. *Handbook of Means and Their Inequalities*; Kluwer Academic Publishers Group: Dordrecht, The Netherlands, 2003.
31. Pareto, V. *Manual of Political Economy*; The Macmillan Press: London, UK, 1971.
32. Hillermeier, C. *Nonlinear Multiobjective Optimization: A Generalized Homotopy Approach*; Springer Science & Business Media: Berlin, Germany, 2001.
33. Zitzler, E.; Thiele, L. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Trans. Evol. Comput.* **1999**, *3*, 257–271. [\[CrossRef\]](#)
34. Brockhoff, D.; Wagner, T.; Trautmann, H. On the Properties of the R2 Indicator. In Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation, Philadelphia, PA, USA, 7–11 July 2012; ACM: New York, NY, USA, 2012; pp. 465–472.
35. Ishibuchi, H.; Masuda, H.; Nojima, Y. A Study on Performance Evaluation Ability of a Modified Inverted Generational Distance Indicator. In Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation, Madrid, Spain, 11–15 July 2015; ACM: New York, NY, USA, 2015; pp. 695–702.

36. Dilettoso, E.; Rizzo, S.A.; Salerno, N. A Weakly Pareto Compliant Quality Indicator. *Math. Comput. Appl.* **2017**, *22*, 25.
37. Garg, H.; Kumar, K. Distance measures for connection number sets based on set pair analysis and its applications to decision-making process. *Appl. Intell.* **2018**, *48*, 1–14. [[CrossRef](#)]
38. Singh, S.; Garg, H. Distance measures between type-2 intuitionistic fuzzy sets and their application to multicriteria decision-making process. *Appl. Intell.* **2017**, *46*, 788–799. [[CrossRef](#)]
39. Van Veldhuizen, D.A.; Lamont, G.B. Multiobjective evolutionary algorithm test suites. In Proceedings of the 1999 ACM Symposium on Applied Computing, San Antonio, TX, USA, 28 February–2 March 1999; ACM: New York, NY, USA, 1999; pp. 351–357.
40. Coello Coello, C.A.; Cruz Cortés, N. Solving Multiobjective Optimization Problems using an Artificial Immune System. *Genet. Program. Evolvable Mach.* **2005**, *6*, 163–190. [[CrossRef](#)]
41. Rudolph, G.; Schütze, O.; Grimme, C.; Domínguez-Medina, C.; Trautmann, H. Optimal averaged Hausdorff archives for bi-objective problems: Theoretical and numerical results. *Comput. Optim. Appl.* **2016**, *64*, 589–618. [[CrossRef](#)]
42. Zitzler, E.; Thiele, L.; Laumanns, M.; Fonseca, C.M.; da Fonseca, V.G. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Trans. Evol. Comput.* **2003**, *7*, 117–132. [[CrossRef](#)]
43. Witkowski, A. A new proof of the monotonicity property of power means. *JIPAM. J. Inequal. Pure Appl. Math.* **2004**, *5*, 73.
44. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [[CrossRef](#)]
45. Zhang, Q.; Li, H. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Trans. Evol. Comput.* **2007**, *11*, 712–731. [[CrossRef](#)]
46. Schütze, O.; Domínguez-Medina, C.; Cruz-Cortés, N.; de la Fraga, L.G.; Sun, J.Q.; Toscano, G.; Landa, R. A scalar optimization approach for averaged Hausdorff approximations of the Pareto front. *Eng. Optim.* **2016**, *48*, 1593–1617. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).