On the Closest Averaged Hausdorff Archive for a Circularly Convex Pareto Front

Günter Rudolph[1(✉)], Oliver Schütze[2], and Heike Trautmann[3]

[1] Department of Computer Science, TU Dortmund University, Dortmund, Germany
guenter.rudolph@tu-dortmund.de

[2] Department of Computer Science, CINVESTAV, Mexico City, Mexico
schuetze@cs.cinvestav.mx

[3] Department of Information Systems, University of Münster, Münster, Germany
trautmann@uni-muenster.de

Abstract. The averaged Hausdorff distance has been proposed as an indicator for assessing the quality of finitely sized approximations of the Pareto front of a multiobjective problem. Since many set-based, iterative optimization algorithms store their currently best approximation in an internal archive these approximations are also termed archives. In case of two objectives and continuous variables it is known that the best approximations in terms of averaged Hausdorff distance are subsets of the Pareto front if it is concave. If it is linear or circularly concave the points of the best approximation are equally spaced.

Here, it is proven that the optimal averaged Hausdorff approximation and the Pareto front have an empty intersection if the Pareto front is circularly convex. But the points of the best approximation are equally spaced and they rapidly approach the Pareto front for increasing size of the approximation.

Keywords: Multi-objective optimization · Averaged hausdorff distance · Convex front · Optimal archives

1 Introduction

The goal in the *a posteriori* approach of multiobjective optimization is a finitely sized approximation of the Pareto front which itself is innumerable for continuous problems in general. The quality of the approximation is typically measured by scalar-valued quality indicators which are largely based on some notion of distance between approximation set and Pareto front. Popular indicators in the field of multiobjective evolutionary algorithms are the dominated hypervolume [1], generational distance [2], (inverted) generational distance [3], the ε-indicator [4], the R2-indicator [5] and others. Besides measuring the quality of an approximation, these indicators can also be used as a selection operator that drives a set-based optimization algorithm like an evolutionary algorithm towards the Pareto front (see e.g. [6–8]).

When used as quality indicator for assessing the approximation found by some set-based algorithm the indicator implicitly determines the characteristics of optimal approximations (like the distribution of solutions on the Pareto front). Since

G. Squillero and P. Burelli (Eds.): EvoApplications 2016, Part II, LNCS 9598, pp. 42–55, 2016.
DOI: 10.1007/978-3-319-31153-1_4

the Pareto front is typically known analytically in case of constructed benchmark problems, it is often possible to calculate the optimal approximation (or optimal archive of points) for the given indicator. As a consequence, it is then possible to assess the quality of the approximation during the optimization run by the difference between optimal approximation and current solution of the evolutionary algorithm or other set-based metaheuristics. For example, this has been achieved for the dominated hypervolume indicator in case of two objectives [9].

Here, we consider optimal approximations/archives for the recently proposed Δ_p-indicator which is based on the *averaged Hausdorff distance* between two sets [10]. In case of two objectives it is known that the optimal archives are subsets of the Pareto front if it is concave (which includes linear fronts) [11]. In addition, optimal archives and Δ_p-indicator values have been determined for linear and circularly concave Pareto fronts. The appealing feature of optimal Δ_p-archives is the uniform spacing between the archive points. Moreover, the Δ_p-indicator can be efficiently used for expert knowledge based multiobjective optimization using a specific archive technique approximating a predefined aspiration set [12].

But numerical experiments have revealed that the optimal Δ_p-archive for *convex* Pareto fronts is not a subset of the Pareto front which questioned the deployment of the Δ_p-indicator for quality assessments of multiobjective evolutionary algorithms in case of (at least piecewise) continuous Pareto fronts. Actually, we prove in case of a circularly convex Pareto front that the optimal Δ_p-archive and the Pareto front have an empty intersection. This seemingly disappointing result, however, does not discredit the Δ_p-indicator as a measuring device in benchmarks since we prove that the Euclidean distance of each archive point to the Pareto front decreases rapidly for increasing archive size. As a consequence, already for moderately sized archives the deviation from the Pareto front is irrelevant from a measuring point of view.

In Sect. 2 we present some mathematical results that exempt the proofs in subsequent sections from disturbing excursions. Section 3 introduces the averaged Hausdorff inframetric and its properties. The main results regarding optimal Hausdorff archives for circularly convex fronts can be found on Sect. 4. Finally, we draw conclusions in Sect. 5.

2 Mathematical Preliminaries

The results presented in this section will be helpful in Sect. 4. Neither of these results is new; the proofs are supplied only for making this work more self-contained.

A square matrix A of size $n \times n$ is termed positive semidefinite (p.s.d.) if $x'Ax \geq 0$ for all $x \in \mathbb{R}^n$ and positive definite (p.d.) if $x'Ax > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$. Here, x' denotes the transpose of vector x.

Lemma 1. *If A is a p.d. and B a p.s.d. $n \times n$-matrix then $A + B$ is p.d.*

Proof.

$$\forall x \in \mathbb{R}^n \setminus \{0\} : x'(A + B)x = \underbrace{x'Ax}_{>0} + \underbrace{x'Bx}_{\geq 0} > 0.$$

□

A function $f : D \to \mathbb{R}$ with convex domain $D \subseteq \mathbb{R}^n$ is said to be convex, if $f(\gamma\, x + (1-\gamma)\, y) \le \gamma\, f(x) + (1-\gamma)\, f(y)$ for all $x, y \in D$ and $\gamma \in [0,1]$. If the inequality is strict the function is termed strictly convex.

Lemma 2. *Let $f : D \to \mathbb{R}$ be an additively decomposable function with $D = D_1 \times D_2$ and $f(x) = f_1(x_1) + f_2(x_2)$ where the sub-functions $f_1 : D_1 \to \mathbb{R}$ and $f_2 : D_2 \to \mathbb{R}$ are convex. Then:*

(a) $f : D \to \mathbb{R}$ is convex.
(b) $f : D \to \mathbb{R}$ is strictly convex if at least one sub-function is strictly convex.

Proof.

(a) Let $\gamma \in [0,1]$ and $x, y \in D$. According to the definition of convexity we get

$$\begin{aligned} f(\gamma\, x + (1-\gamma)\, y) &= f_1(\gamma\, x_1 + (1-\gamma)\, y_1) + f_2(\gamma\, x_2 + (1-\gamma)\, y_2) \\ &\le \gamma\, f_1(x_1) + (1-\gamma)\, f_1(y_1) + \gamma\, f_2(x_2) + (1-\gamma)\, f_2(y_2) \quad (1) \\ &= \gamma\, (f_1(x_1) + f_2(x_2)) + (1-\gamma)\, (f_1(y_1) + f_2(y_2)) \\ &= \gamma\, f(x) + (1-\gamma)\, f(y). \end{aligned}$$

(b) Inequality (1) is strict if at least one sub-function is strictly convex.

□

Optima of a differentiable function may be determined by inspection of its gradient and Hessian matrix. Sometimes a monotone transformation of the function can make the analysis easier.

Lemma 3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be continuously differentiable. Then:*

(a) $\{x \in \mathbb{R}^n : \nabla f(x) = 0\} \subseteq \{x \in \mathbb{R}^n : \nabla g(f(x)) = 0\}$.
(b) $\{x \in \mathbb{R}^n : \nabla f(x) = 0\} = \{x \in \mathbb{R}^n : \nabla g(f(x)) = 0\}$ if $g(\cdot)$ strictly monotone.
(c) If $g(\cdot)$ is twice differentiable and strictly monotone increasing on the range of $f(\cdot)$ then the location of the local minima of $g(f(x))$ are identical to those of $f(x)$.

Proof.

(a) Since $\nabla g(f(x)) = \nabla f(x) \cdot g'(f(x))$ the left hand side of the equation is zero if $\nabla f(x) = 0$ or the derivative of $g(\cdot)$ has a root in the range of $f(\cdot)$.
(b) Since $g(y)$ is strictly monotone we have either $g'(y) > 0$ or $g'(y) < 0$ but never $g'(y) = 0$ for all $y \in f(\mathbb{R}^n)$. Therefore $\nabla g(f(x))$ is zero if and only if $\nabla f(x) = 0$.
(c) Strictly increasing monotonicity implies $g'(y) > 0$ for all $y \in f(\mathbb{R}^n)$. Owing to part (b) of this Lemma it follows that the set of candidates for local optima are identical. The Hessian matrix of $g(f(x))$ is

$$\begin{aligned} \nabla^2 g(f(x)) = \nabla(\nabla g(f(x))) = \nabla(g'(f(x)) \cdot \nabla f(x)) = \\ g'(f(x)) \cdot \nabla^2 f(x) + g''(f(x)) \cdot \underbrace{\nabla f(x) \cdot \nabla f(x)^T}_{\text{zero matrix if } x=x^*}. \end{aligned} \quad (2)$$

Insertion of a candidate solution x^* in the Hessian (2) leads to

$$\nabla^2 g(f(x^*)) = \underbrace{g'(f(x^*))}_{>0} \cdot \nabla^2 f(x^*)$$

revealing that the Hessian matrix of $f(x)$ is p.d. in x^* if and only if $\nabla^2 g(f(x^*))$ is p.d. □

Finally, we recall the trigonometric identities

$$\left.\begin{array}{lcl} \sin(\alpha+\delta) - \sin(\alpha+\delta) & = & 2\,\cos\alpha\,\sin\delta \\ \sin(\alpha+\delta) + \sin(\alpha+\delta) & = & 2\,\sin\alpha\,\cos\delta \\ \cos(\alpha+\delta) - \cos(\alpha+\delta) & = & -2\,\sin\alpha\,\sin\delta \\ \cos(\alpha+\delta) + \cos(\alpha+\delta) & = & 2\,\cos\alpha\,\cos\delta \end{array}\right\} \tag{3}$$

which follow immediately from entries 4.3.34 to 4.3.37 in [13] whereas

$$\sin(\arctan x) = \frac{x}{\sqrt{1+x^2}} \quad \text{and} \quad \cos(\arctan x) = \frac{1}{\sqrt{1+x^2}} \tag{4}$$

can be extracted from entry 4.3.45 in [13].

3 Averaged Hausdorff Distance

The Hausdorff distance is a well known distance measure between two sets. Actually, it can be shown that it is a metric on sets.

Definition 1. *The value* $d_H(A,B) := \max(d(A,B), d(B,A))$ *is termed the* Hausdorff distance *between two sets* $A, B \subset \mathbb{R}^k$*, where*

$$d(B,A) := \sup\{d(u,A) : u \in B\} \quad \textit{and} \quad d(u,A) := \inf\{\|u-v\| : v \in A\}$$

for a vector norm $\|\cdot\|$*.*

However, the value of the Hausdorff distance is strongly affected by single outliers which may lead to counter-intuitive assessments of closeness between sets [10]. Therefore, the indicators GD_p and IGD_p have been introduced in [10] to construct a new distance measure between finite sets that shares some properties with the Hausdorff distance but which is less prone to outliers.

Definition 2. *The value* $\Delta_p(A,B) = \max(\mathsf{GD}_p(A,B), \mathsf{IGD}_p(A,B))$ *with*

$$\mathsf{GD}_p(A,B) = \left(\frac{1}{|A|}\sum_{a\in A} d(a,B)^p\right)^{1/p} \textit{and} \;\; \mathsf{IGD}_p(A,B) = \left(\frac{1}{|B|}\sum_{b\in B} d(b,A)^p\right)^{1/p}$$

for $p > 0$ *is termed the* averaged Hausdorff distance *between sets* A *and* B *as given in* Definition 1*.*

We note that $\Delta_p(\cdot,\cdot)$ is not a metric but fulfills all axioms of a distance.

Definition 3. (see [14, p. 1])
A function $d : X \times X \to \mathbb{R}$ *is termed a* distance *on set* X *if*

1. $d(x, y) \geq 0$ *(nonnegativity)*
2. $d(x, y) = d(y, x)$ *(symmetry)*
3. $d(x, x) = 0$ *(reflexivity)*

for all $x, y \in X$. □

The averaged Hausdorff distance is not a metric (but a *nearmetric*) only since the triangle property is not fully valid.

Definition 4. (see [14, p. 7])
A function $d : X \times X \to \mathbb{R}$ *is termed a* C-nearmetric *on set* X *if*

1. *it is a distance on* X
2. $d(x, y) > 0$ *for* $x \neq y$
3. $d(x, y) \leq C\,(d(x, z) + d(z, y))$

for all $x, y, z \in X$ *and some* $C \geq 1$. □

Actually, it can be proven [10] that Δ_p is a $N^{1/p}$-nearmetric where N is the maximum cardinality of the sets.

The purpose of the averaged Hausdorff distance is to assess the quality of a finitely sized approximation Y of the Pareto front F^* by observing the value $\Delta_p(Y, F^*)$. In the context of the definitions in [10] the Pareto front F^* is assumed to be discretized appropriately.

Here, we consider only bi-objective problems and assume that the Pareto front is continuous and expressible in parametric form. Thus, let $F^* = \{\varphi(\omega) : \omega \in [0, \pi/2]\} \subset \mathbb{R}^2$ be the Pareto front and $Y \subset \mathbb{R}^2$ with $|Y| = m < \infty$ the approximation.

Two examples are the circularly concave Pareto front defined by $\varphi(\omega) = (\sin\omega, \cos\omega)'$ and the circularly convex Pareto front defined by $\varphi(\omega) = (1 - \cos\omega, 1 - \sin\omega)'$. An illustration of these Pareto fronts is given in Fig. 1.

Since here we deal with continuous Pareto fronts F^*, we have to adapt Δ_p to this context: let $Y = \{y_1, \ldots, y_m\} \subset f(X)$ where $y_1, \ldots, y_m$ are arranged in lexicographic order. We obtain

$$\mathsf{GD}_p(Y, F^*) \;=\; \left[\frac{1}{|Y|} \sum_{y \in Y} d(y, F^*)^p\right]^{\frac{1}{p}} \tag{5}$$

and

$$\mathsf{IGD}_p(Y, F^*) = \left[\frac{1}{L} \int_a^b d(\varphi(s), Y)^p \, ds\right]^{\frac{1}{p}} \tag{6}$$

where L is the length of the curve described by F^*. In case of the two examples we have of course $L = \pi/2$.

In [11] the optimal Δ_p-archive has been determined for the circularly concave front. Here, we consider the determination of the optimal Δ_p-archive in case of the circularly convex front.

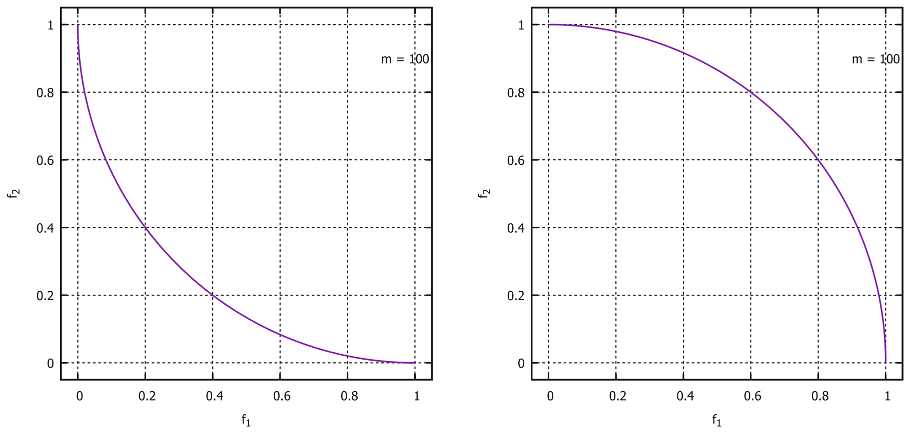

Fig. 1. Left: circularly convex Pareto front. Right: circularly concave Pareto front.

4 Construction of Optimal Archive

Since the averaged Hausdorff distance Δ_p between finite archive and innumerable Pareto front is the maximum of GD_p and IGD_p between the two sets we can find the optimal archive as follows: we know that solutions with minimum GD_p are archives with all members on the Pareto front which results in $\mathsf{GD}_p = 0$ so that the corresponding Δ_p value is just the IGD_p value for the optimal GD_p archive. Next, find solutions with minimum IGD_p. If such a solution has a larger IGD_p value than the corresponding GD_p value then this solution is the minimal Δ_p archive. If this inequality is not valid then this approach fails.

Presume that this approach does work. First, we need a method to calculate the GD_p value for an arbitrary given archive. This method is developed in Sect. 4.1. Second, we need a method to determine an archive with minimal IGD_p value. We demonstrate the method in case of a circularly convex Pareto front in Sect. 4.2. Finally, a comparison with the corresponding GD_p value in Sect. 4.3 reveals that this solution is the optimal Δ_p archive.

4.1 Averaged Generational Distance (GD_p)

In the remainder of this work we assume that the vector norm used in Definition 1 is the Euclidean norm and that $\varphi(\cdot)$ is continuously differentiable.

Theorem 1. *Let $F^* = \{\varphi(\omega) : \omega \in [0, \pi]\} \subset \mathbb{R}^2$ and $Y \subset \mathbb{R}^2$ with $|Y| = m < \infty$. Then*

$$\mathsf{GD}_p(Y, F^*) = \mathsf{GD}_p(\{y^{(1)}, \ldots, y^{(m)}\}, F^*) = \left[\frac{1}{m} \sum_{i=1}^{m} d(\varphi(\omega_i^*), y^{(i)})^p \right]^{1/p}$$

where ω_i^* is an appropriate solution of

$$\frac{\partial}{\partial \omega_i} d(\varphi(\omega_i), y^{(i)}) \stackrel{!}{=} 0$$

for $i = 1, \ldots, m$ regardless of $p \geq 1$.

Proof. Suppose that $Y \cap F^* = \emptyset$ since elements of Y that are on F^* do not contribute to the value of GD_p. Let

$$g(\omega) = g(\omega_1, \ldots, \omega_m) = \frac{1}{m} \sum_{i=1}^{m} d(\varphi(\omega_i), y^{(i)})^p.$$

Partial derivation leads to

$$\frac{\partial}{\partial \omega_i} g(\omega)^{1/p} = \left[\frac{\partial}{\partial \omega_i} g(\omega) \right] \cdot \underbrace{\frac{1}{p} \cdot g(\omega)}_{>0} \stackrel{!}{=} 0 \tag{7}$$

for $i = 1, \ldots, m$. Therefore, it suffices to look at

$$\begin{aligned} \frac{\partial}{\partial \omega_i} g(\omega) \;&=\; \frac{1}{m} \cdot \frac{\partial}{\partial \omega_i} d(\varphi(\omega_i), y^{(i)})^p \;= \\ \frac{1}{m} \left[\frac{\partial}{\partial \omega_i} d(\varphi(\omega_i), y^{(i)}) \right] &\cdot \underbrace{p \cdot d(\varphi(\omega_i), y^{(i)})^{p-1}}_{>0} \stackrel{!}{=} 0 \end{aligned} \tag{8}$$

for $i = 1, \ldots, m$. Thus, it is sufficient to solve

$$\frac{\partial}{\partial \omega_i} d(\varphi(\omega_i), y^{(i)}) \stackrel{!}{=} 0$$

independently for each $i = 1, \ldots, m$ and regardless of $p \geq 1$ to find the candidates for optimal angles ω_i^*. It remains to show that these candidates lead to a minimum, i.e., that the Hessian matrix of GD_p is positive definite (p.d.). The second partial derivatives of GD_p are obtained by partial derivation of (7) yielding

$$\frac{\partial}{\partial \omega_j} \left[\frac{\partial}{\partial \omega_i} g(\omega) \right] \cdot \frac{1}{p} \cdot g(\omega) = \frac{1}{p} \left(\frac{\partial^2 g(\omega)}{\partial \omega_i \, \partial \omega_j} \, g(\omega) + \frac{\partial g(w)}{\partial \omega_i} \cdot \frac{\partial g(w)}{\partial \omega_j} \right)$$

which can be expressed in matrix form as

$$p \cdot \nabla^2 \mathsf{GD}_p(\omega) \;=\; \underbrace{g(\omega)}_{>0} \cdot \nabla^2 g(\omega) + \underbrace{\nabla g(\omega) \, \nabla g(\omega)^T}_{p.s.d.}.$$

The Hessian matrix of GD_p is p.d. if $\nabla^2 g(\omega)$ is p.d. (Lemma 1). Partial derivation of the first partial derivatives of $g(\omega)$ given in (8) w.r.t. ω_j yields

$$\begin{aligned} \frac{\partial^2 g(\omega)}{\partial \omega_i \partial \omega_j} = \frac{p}{m} \cdot & \underbrace{\frac{\partial^2 d(\varphi(\omega_i), y^{(i)})}{\partial \omega_i \partial \omega_j}}_{=0 \text{ if } i \neq j} \cdot d(\varphi(\omega_i), y^{(i)})^{p-1} \\ & + \quad \frac{p}{m} \cdot \frac{\partial d(\varphi(\omega_i), y^{(i)})}{\partial \omega_i} \cdot \underbrace{\frac{\partial d(\varphi(\omega_i), y^{(i)})^{p-1}}{\partial \omega_j}}_{=0 \text{ if } i \neq j}. \end{aligned}$$

Thus, $\nabla^2 g(\omega)$ is a diagonal matrix $\mathrm{diag}(d_1, \ldots, d_m)$ that is p.d. if every diagonal entry d_i is positive. Since

$$d_i = \frac{p}{m} \cdot \frac{\partial^2 d(\varphi(\omega_i), y^{(i)})}{\partial \omega_i^2} \cdot \underbrace{d(\varphi(\omega_i), y^{(i)})^{p-1}}_{>0} + \frac{p}{m} \cdot \underbrace{\left(\frac{\partial d(\varphi(\omega_i), y^{(i)})}{\partial \omega_i}\right)^2}_{\geq 0} \cdot \underbrace{(p-1)}_{\geq 0} \cdot \underbrace{d(\varphi(\omega_i), y^{(i)})^{p-2}}_{>0}$$

the sufficient condition reduces to

$$\frac{\partial^2 d(\varphi(\omega_i), y^{(i)})}{\partial \omega_i^2} \overset{!}{>} 0,$$

which is always fulfilled since the Euclidean norm is strictly convex. $\square$

Suppose the Pareto front is given by $F^* = \{\varphi(\omega) : \omega \in [0, \pi/2]\}$ with $\varphi(\omega) = (1 - \cos\omega, 1 - \sin\omega)^T$ and let $Y \subset [0, 1)^2$ with $|Y| = m < \infty$. Owing to Theorem 1 it suffices to solve

$$\frac{\partial d(\varphi(\omega), y)}{\partial \omega} \overset{!}{=} 0$$

for a single pair (ω, y). Lemma 3(c) asserts that the solution of the squared problem delivers the desired ω^*. Thus,

$$\begin{aligned} \frac{\partial d(\varphi(\omega), y)^2}{\partial \omega} &= \frac{\partial}{\partial \omega}[(1 - \cos\omega - y_1)^2 + (1 - \sin\omega - y_2)^2] \\ &= 2\,(1 - \cos\omega - y_1)(-\sin\omega) + 2\,(1 - \sin\omega - y_2)\cos\omega \overset{!}{=} 0 \end{aligned}$$

which is equivalent to

$$\frac{1 - y_2}{1 - y_1} \overset{!}{=} \frac{\sin\omega}{\cos\omega} = \tan\omega \quad \Leftrightarrow \quad \arctan\left(\frac{1 - y_2}{1 - y_1}\right) \overset{!}{=} \omega^*.$$

Insertion of ω^* and usage of the trigonometric identities (4) leads to

$$d(\varphi(\omega^*(y)), y) = \left[(1 - y_1)^2 + (1 - y_2)^2 - 2\sqrt{(1 - y_1)^2 + (1 - y_2)^2} + 1\right]^{\frac{1}{2}} \tag{9}$$

which can be used as building block to calculate GD_p for an arbitrary given set $\{y^{(1)}, \ldots, y^{(m)}\} \subset [0, 1)^2$.

4.2 Averaged Inverted Generational Distance (IGD_p)

We solve the squared problem. At first the integrals are eliminated before we apply partial differentiation and proceed in the standard way to identify the optimal points of set Y analytically. To this end let $0 = \alpha_0 \leq \alpha_1 \leq \ldots \leq \alpha_m = \pi/2$.

$$\frac{\pi}{2}\,\mathsf{IGD}_2(\{y^{(1)}, \ldots, y^{(m)}\}, F^*)^2 \;=\; \sum_{i=1}^{m} \int_{\alpha_{i-1}}^{\alpha_i} d(\varphi(\omega), y^{(i)})^2 \, d\omega$$

$$= \sum_{i=1}^{m} \int_{\alpha_{i-1}}^{\alpha_i} \left[(1 - \cos\omega - y_1^{(i)})^2 + (1 - \sin\omega - y_2^{(i)})^2 \right] d\omega$$

$$= \sum_{i=1}^{m} \int\limits_{\alpha_{i-1}}^{\alpha_i} \left[(1 - y_1^{(i)})^2 + (1 - y_2^{(i)})^2 + 1 - 2\,(1 - y_1^{(i)}) \cos\omega - (1 - y_2^{(i)}) \sin\omega \right] d\omega$$

$$\begin{aligned} &= \sum_{i=1}^{m} \left[(z_1^{(i)})^2 + (z_2^{(i)})^2 + 1 \right] (\alpha_i - \alpha_{i-1}) \\ &\quad - 2 \sum_{i=1}^{m} z_1^{(i)} (\sin\alpha_i - \sin\alpha_{i-1}) \\ &\quad + 2 \sum_{i=1}^{m} z_2^{(i)} (\cos\alpha_i - \cos\alpha_{i-1}) \end{aligned} \tag{10}$$

where we temporarily made the variable replacement $z_k^{(i)} = 1 - y_k^{(i)}$ for $k = 1, 2$. Partial differentiation of (10) w.r.t. $z_k^{(i)}$ leads to the first set of necessary conditions with $i = 1, \ldots, m$:

$$\frac{\partial V}{\partial z_1^{(i)}} = 2\, z_1^{(i)} (\alpha_i - \alpha_{i-1}) - 2\,(\sin\alpha_i - \sin\alpha_{i-1}) \overset{!}{=} 0$$

$$\frac{\partial V}{\partial z_2^{(i)}} = 2\, z_2^{(i)} (\alpha_i - \alpha_{i-1}) + 2\,(\cos\alpha_i - \cos\alpha_{i-1}) \overset{!}{=} 0$$

which are equivalent to

$$z_1^{(i)} \overset{!}{=} \frac{\sin\alpha_i - \sin\alpha_{i-1}}{\alpha_i - \alpha_{i-1}} \quad \text{and} \quad z_2^{(i)} \overset{!}{=} -\frac{\cos\alpha_i - \cos\alpha_{i-1}}{\alpha_i - \alpha_{i-1}}. \tag{11}$$

Thus, conditions (11) tell us how to choose set Y for any given partitioning of F^* with switch points $\varphi(\alpha_1), \ldots, \varphi(\alpha_{m-1})$. If $m = 1$ there is no switch point since $\alpha_0 = 0$ and $\alpha_1 = \pi/2$. Therefore

$$y_1^{(1)} = 1 - z_1^{(1)} = 1 - \frac{2}{\pi} \quad \text{and} \quad y_2^{(1)} = 1 - z_2^{(1)} = 1 - \frac{2}{\pi}$$

with optimal

$$\mathsf{IGD}_2(\{y^{(1)}\}, F^*) = \sqrt{1 - \frac{8}{\pi^2}} \approx 0.43524$$

and associated

$$\mathsf{GD}_2(\{y^{(1)}\}, F^*) = 1 - \frac{\sqrt{8}}{\pi} \approx 0.09968$$

after insertion in (10) and (9), respectively. For $m > 1$ we also need to consider the optimality conditions for the angles. Partial differentiation of (10) w.r.t. α_i leads to the second set of necessary conditions

$$\begin{aligned}\frac{\partial V}{\partial \alpha_i} &= (z_1^{(i)})^2 + (z_2^{(i)})^2 - z_1^{(i)} \cos \alpha_i - z_2^{(i)} \sin \alpha_i \\ &\quad - \left[(z_1^{(i+1)})^2 + (z_2^{(i+1)})^2 - z_1^{(i+1)} \cos \alpha_i - z_2^{(i+1)} \sin \alpha_i \right] \overset{!}{=} 0 \qquad (12)\end{aligned}$$

where $i = 1, \dots, m-1$. After insertion of (11) in (12) we obtain

$$\begin{aligned}&2 \cos \alpha_i \left[\frac{\cos \alpha_{i+1}}{(\alpha_{i+1} - \alpha_i)^2} - \frac{\cos \alpha_{i-1}}{(\alpha_i - \alpha_{i-1})^2} \right] + 2 \sin \alpha_i \left[\frac{\sin \alpha_{i+1}}{(\alpha_{i+1} - \alpha_i)^2} - \frac{\sin \alpha_{i-1}}{(\alpha_i - \alpha_{i-1})^2} \right] \\ &+ \cos \alpha_i \left[\frac{\sin \alpha_{i+1}}{\alpha_{i+1} - \alpha_i} + \frac{\sin \alpha_{i-1}}{\alpha_i - \alpha_{i-1}} \right] - \sin \alpha_i \left[\frac{\cos \alpha_{i+1}}{\alpha_{i+1} - \alpha_i} + \frac{\cos \alpha_{i-1}}{\alpha_i - \alpha_{i-1}} \right] \overset{!}{=} 0. \qquad (13)\end{aligned}$$

If $m = 2$ conditions (13) reduce to a single equation with a single variable:

$$\frac{2\,(1 - \cos \alpha_1)}{\alpha_1^2} - \frac{2\,(1 - \sin \alpha_1)}{(\pi/2 - \alpha_1)^2} - \frac{2\,\sin \alpha_1}{\alpha_1} + \frac{2\,\cos \alpha_1}{\pi/2 - \alpha_1} \overset{!}{=} 0. \qquad (14)$$

Evidently, the choice of $\alpha_1^* = \pi/4$ solves the equation. Insertion of $\alpha_1^* = \pi/4$ in the second derivative (i.e., the derivative of (14) w.r.t. α_1) yields a positive value (> 0.7) revealing that this choice is at least a *local* minimum. The positions of the corresponding archive points

$$y^{(1)} = \begin{pmatrix} 1 - \dfrac{\sqrt{8}}{\pi} \\ \\ 1 - \dfrac{4 - \sqrt{8}}{\pi} \end{pmatrix} \quad \text{and} \quad y^{(2)} = \begin{pmatrix} 1 - \dfrac{4 - \sqrt{8}}{\pi} \\ \\ 1 - \dfrac{\sqrt{8}}{\pi} \end{pmatrix} \qquad (15)$$

are obtained after insertion of α_1^* in (11). These values lead to minimal

$$\mathsf{IGD}_2(\{y^{(1)}, y^{(2)}\}, F^*) = \sqrt{1 - \frac{16\,(2 - \sqrt{2})}{\pi^2}} \approx 0.22441$$

with corresponding

$$\mathsf{GD}_2(\{y^{(1)}, y^{(2)}\}, F^*) = 1 - \frac{4\sqrt{2 - \sqrt{2}}}{\pi} \approx 0.0255046.$$

A closer look at the coordinates in (15) discloses a symmetry in the solution. This observation gives rise to the **conjecture** that also solutions in the general case with $m > 2$ should exhibit this kind of symmetry, namely

$$y_1^{(i)} = y_2^{(m-i+1)}$$

for $i = 1, \ldots, m$. If the conjecture is true then the difference between consecutive angles must be equal, i.e., $\alpha_i^* - \alpha_{i-1}^* = \delta > 0$ for $i = 1, \ldots, m$ or, equivalently,

$$\alpha_i^* = i \cdot \delta \quad \text{with} \quad \delta = \frac{\pi}{2\,m}$$

for $i = 0, 1, \ldots, m$. As a consequence, $\alpha_{i-1} = \alpha_i - \delta$ and $\alpha_{i+1} = \alpha_i + \delta$. Using this setting of angles and the trigonometric identities in (3) the necessary conditions for the angles (13) are fulfilled leading to optimal archive points

$$y^{(i)} = \left(1 - \frac{\sin(i \cdot \delta) - \sin((i-1) \cdot \delta)}{\delta},\; 1 + \frac{\cos(i \cdot \delta) - \cos((i-1) \cdot \delta)}{\delta} \right)^{\mathsf{T}}$$

where $\delta = \pi/(2\,m)$. Figure 2 shows optimal archives of size $m = 1, 2, 3, 4, 5$ and 10. As can be seen, the archive points move closer to the Pareto front for increasing cardinality.

Table 1 provides an impression about the minimal IGD_2 and corresponding GD_2 values for increasing archive size $m \in \mathbb{N}$. The results give rise to the educated guess that the IGD_2 value decreases with order $1/m$ whereas the GD_2 value decreases even with order $1/m^2$. A closer inspection provides at least numerical evidence of the relation

$$\mathsf{GD}_2 = 1 - \sqrt{1 - \mathsf{IGD}_2^2}$$

if IGD_2 is the optimal value.

It remains to show that the conjecture of equidistant angles is true. Notice that equidistant angles fulfill the necessary conditions (13) and (11). If there

Table 1. Minimal IGD_2 values and corresponding GD_2 values for increasing archive size $m \in \mathbb{N}$.

m	IGD_2	GD_2
1	4.3524×10^{-1}	9.96837×10^{-2}
2	2.2441×10^{-1}	2.55046×10^{-2}
3	1.5046×10^{-1}	1.13841×10^{-2}
4	1.1307×10^{-2}	6.41315×10^{-3}
5	9.0541×10^{-2}	4.10726×10^{-3}
6	7.5489×10^{-2}	2.85334×10^{-3}
7	6.4724×10^{-2}	2.09681×10^{-3}
8	5.6649×10^{-2}	1.60561×10^{-3}
9	5.0358×10^{-2}	1.26876×10^{-3}
10	4.5326×10^{-2}	1.02777×10^{-3}
100	4.5345×10^{-3}	1.02808×10^{-5}
1,000	4.5345×10^{-4}	1.02808×10^{-7}
10,000	4.5345×10^{-5}	1.02808×10^{-9}

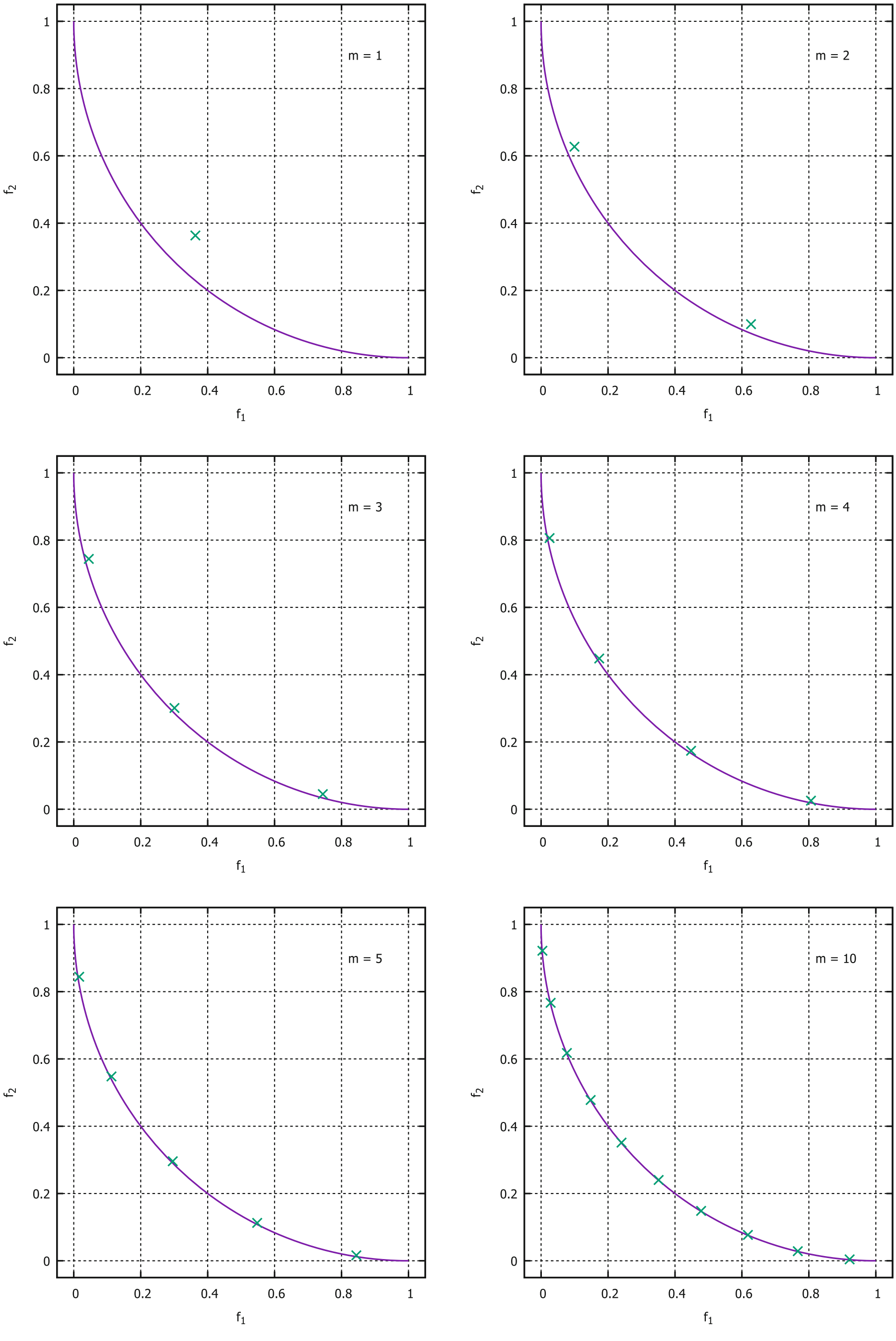

Fig. 2. Optimal archives of size $m = 1, 2, 3, 4, 5, 10$ for the circularly convex Pareto front.

is no other solution that fulfills the necessary conditions then there is no other candidate solution in the interior of $[0, \pi/2]^m$ with possibly better function value. Since (10) can be written as a sum of strictly convex sub-functions

$$f_i(z^{(i)}) = \delta_i \left((z_1^{(i)})^2 + (z_2^{(i)})^2 \right) + a_i\, z_1^{(i)} + b_i\, z_2^{(i)} + \delta_i$$

where $\delta_i := \alpha_i - \alpha_{i-1} > 0$ and $a_i, b_i \in \mathbb{R}$ for $i = 1, \ldots, m$, Lemma 2(b) asserts that (10) is strictly convex. As a consequence the local minimizer is unique and also the global solution.

4.3 Optimal Averaged Hausdorff Archive ($\boldsymbol{\Delta_p}$)

In the introduction to this Sect. 4 it was elaborated that the optimal IGD_p-archive is also the optimal Δ_p-archive if the associated GD_p value for this solution is smaller than the minimal IGD_p value. As seen in the preceding subsection, this relationship is true for archive sizes $m = 1, 2$ and it can be assured numerically exact for larger values of m by the formulas we have proven.

The expression (11) for the coordinates of the optimal archive points reveals that the archive elements will not be elements of the Pareto front regardless of the finite archive size m. Only if $m \to \infty$, which in turn implies $\delta \to 0$, the optimal archive points converge to $(1 - \cos\omega, 1 - \sin\omega)'$ for every $\omega \in (0, \pi/2)$.

5 Conclusions

The observation that the optimal Δ_p-approximation does not lie on the convex part(s) of the Pareto front has questioned the deployment of the Δ_p-indicator for the quality assessment of multiobjective evolutionary algorithms. The theoretical analysis presented here proves for circular convex fronts (1) that this observation is actually not a speculation but a fact and (2) that the deviation from the Pareto front decreases rapidly for increasing size of the approximation / population leading to insignificant inaccuracies already for population size ≥ 100. In summary, our results are not in opposition to the deployment of the Δ_p-indicator for benchmarking multiobjective evolutionary algorithms.

The results presented in this work can be extended in various ways. The 'numerically exact' proof that the IGD_p value for the optimal archive is larger than the associated GD_p value should be replaced by an analytic version. The method may be applied and extended, if necessary, to nonsymmetric convex test functions. Finally, we must admit that the extension of this approach from the 2-dimensional to 3-dimensional case is an open question.

Acknowledgment. Support from CONACYT project no. 207403 and DAAD project no. 57065955 is gratefully acknowledged. Additionally, Heike Trautmann acknowledges support by the European Center of Information Systems (ERCIS).

References

1. Zitzler, E., Thiele, L.: Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study. In: Eiben, A.E., Bäck, T., Schoenauer, M., Schwefel, H.-P. (eds.) PPSN 1998. LNCS, vol. 1498, pp. 292–301. Springer, Heidelberg (1998)
2. van Veldhuizen, D.A.: Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations. PhD thesis, Department of Electrical and Computer Engineering. Graduate School of Engineering. Air Force Institute of Technology, Wright-Patterson AFB, Ohio, May 1999
3. Bosman, P., Thierens, D.: The balance between proximity and diversity in multiobjective evolutionary algorithms. IEEE Trans. Evol. Comput. **7**(2), 174–188 (2003)
4. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C., Grunert da Fonseca, V.: Performance assessment of multiobjective optimizers: an analysis and review. IEEE Trans. Evol. Comput. **7**(2), 117–132 (2003)
5. Hansen, M.P., Jaszkiewicz, A.: Evaluating the quality of approximations of the non-dominated set. IMM Technical Report IMM-REP-1998-7, Institute of Mathematical Modeling, Technical University of Denmark, Lyngby, March 1998
6. Beume, N., Naujoks, B., Emmerich, M.: SMS-EMOA: Multiobjective selection based on dominated hypervolume. Eur. J. Oper. Res. **181**(3), 1653–1669 (2007)
7. Gerstl, K., Rudolph, G., Schütze, O., Trautmann, H.: Finding evenly spaced fronts for multiobjective control via averaging Hausdorff-measure. In: Proceedings of 8th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), pp. 1–6. IEEE Press (2011)
8. Brockhoff, D., Wagner, T., Trautmann, H.: R2 indicator based multiobjective search. Evol. Comput. **23**(3), 369–395 (2015)
9. Auger, A., Bader, J., Brockhoff, D., Zitzler, E.: Theory of the hypervolume indicator: Optimal μ-distributions and the choice of the reference point. In: Proceedings of the 10th ACM SIGEVO Workshop on Foundations of Genetic Algorithms FOGA 2009, pp. 87–102. ACM, New York (2009)
10. 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. **16**(4), 504–522 (2012)
11. 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. (2015). doi:10.1007/s10589-015-9815-8, online 23 December 2015
12. Rudolph, G., Schütze, O., Grimme, C., Trautmann, H.: An aspiration set EMOA based on averaged Hausdorff distances. In: Pardalos, P.M., Resende, M.G.C., Vogiatzis, C., Walteros, J.L. (eds.) LION 8. LNCS, vol. 8246, pp. 153–156. Springer, Heidelberg (2014)
13. Abramowitz, M., Stegun, I. (eds.): Handbook of Mathematical Formulas. National Bureau of Standards, Washington (DC) (1972). 10th printing
14. Deza, M., Deza, E.: Encyclopedia of Distances, 3rd edn. Springer, Heidelberg (2014)