Hypervolume in Biobjective Optimization Cannot Converge Faster Than $\Omega(1/p)$

Eugénie Marescaux, Nikolaus Hansen

Hypervolume in Biobjective Optimization Cannot Converge Faster Than $\Omega(1/p)$

Eugénie Marescaux
Ecole Polytechnique, CMAP
91128 Palaiseau, France
forename.lastname@inria.fr

Nikolaus Hansen
Inria
CMAP, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris
91128 Palaiseau, France
forename.lastname@inria.fr

ABSTRACT

The hypervolume indicator is widely used by multi-objective optimization algorithms and for assessing their performance. We investigate a set of p vectors in the biobjective space that maximizes the hypervolume indicator with respect to some reference point, referred to as *p-optimal distribution*. We prove explicit lower and upper bounds on the gap between the hypervolumes of the p-optimal distribution and the ∞-optimal distribution (the Pareto front) as a function of p, of the reference point, and of some Lipschitz constants. On a wide class of functions, this optimality gap can not be smaller than $\Omega(1/p)$, thereby establishing a bound on the optimal convergence speed of any algorithm. For functions with either bilipschitz or convex Pareto fronts, we also establish an upper bound and the gap is hence $\Theta(1/p)$. The presented bounds are not only asymptotic. In particular, functions with a linear Pareto front have the normalized exact gap of $1/(p+1)$ for any reference point dominating the nadir point.

We empirically investigate on a small set of Pareto fronts the exact optimality gap for values of p up to 1000 and find in all cases a dependency resembling $1/(p + \text{CONST})$.

CCS CONCEPTS

• **Theory of computation → Mathematical optimization**; • **Mathematics of computing →** *Continuous optimization*;

KEYWORDS

p-optimal distribution, biobjective optimization, convergence rate analysis

ACM Reference Format:
Eugénie Marescaux and Nikolaus Hansen. 2021. Hypervolume in Biobjective Optimization Cannot Converge Faster Than $\Omega(1/p)$. In *2021 Genetic and Evolutionary Computation Conference (GECCO '21), July 10–14, 2021, Lille, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3449639.3459371

GECCO '21, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8350-9/21/07…$15.00
https://doi.org/10.1145/3449639.3459371

1 INTRODUCTION

Multi-objective optimization aims at minimizing a vector-valued function F. Nowadays, a main goal of multi-objective optimization is to find a good approximation of the Pareto set, the set of all non-dominated feasible vectors of the search space. When measuring the performance in the objective space, there are at least three different ways to define the convergence of a multi-objective optimization algorithm towards the Pareto front. First, the F-values of a subsequence of vectors explored by the algorithm may converge to a vector of the Pareto front. Proof of such convergence already exists for many multi-objective algorithms, such as MultiGLODS [5], Newton's method [8], the projected gradient method [9] and (1+1) evolutionary multiobjective algorithms [3], often with a guarantee on the convergence rate. Second, for evolutionary algorithms, the set of F-values of the population can converge to a good approximation of the Pareto front. In the case of evolutionary algorithms with a hypervolume based selection, ideally, the image of the population would converge to a set of p vectors of the objective space maximizing the hypervolume, with p being the population size. We call such a set *a p-optimal distribution.* Finally, a dynamic subset of the archive can converge to the entire Pareto front, in some sense which is to be defined. For example, it has been known for a long time that under the hypothesis of an uncountably infinite population, the set of all non-dominated vectors explored by an evolutionary algorithm converges almost surely to the Pareto front [12].

Set-quality indicators are widely used for assessing the performance of multi-objective optimization algorithms. They create a total order where there only was a partial one. Many indicators have been invented and are thoroughly used, like the multiplicative and additive epsilon indicators [16] or the hypervolume indicator. The hypervolume and its variants such as the weighted hypervolume [13] are the only known strictly Pareto compliant indicators [15]. The multiplicative epsilon indictor is also called multiplicative approximation ratio when the Pareto front is used as reference set. In the biobjective case, it has been proven that the multiplicative approximation ratio of the set of all vectors explored by the algorithm cannot converge to 1 more rapidly than $\Omega(1/p)$, with p being the number of function evaluations [4]. It is a direct consequence of Corollary 3.2 in [4], which gives a lower bound of the form $1 + \Theta(1/p)$ of the minimum multiplicative approximation ratio of a set of p vectors.

In this paper, we derive lower and upper bounds of the form $\Theta(1/p)$ of the difference in hypervolume between a p-optimal distribution and the Pareto front. We call this difference *optimality gap*.

For bilipschitz Pareto fronts, we have a tight lower bound on the optimality gap of the form $\mathrm{CONST}/(p+1)$. The constant depends on the bilipschitz constants and on the position of the reference point of the hypervolume indicator with respect to the nadir point. The lower bound we found is exact in the case of a linear Pareto front where the reference point dominates the nadir point. In this case, the constant is simply the hypervolume of the Pareto front. We generalize this result to Pareto fronts with a bilipschitz subsection which dominates the reference point. For this wide class of Pareto fronts (see Figure 6), the rate of convergence of multi-objective algorithms in terms of hypervolume cannot be better than $\Omega(1/p)$. For bilipschitz or convex Pareto fronts, we prove an upper bound of the form $\mathrm{CONST} \times (p+1)/p^2$. The constant depends on the extreme values of the Pareto front, on the reference point, and additionally on the bilipschitz constants in the bilipschitz case. Since any convex Pareto front has a bilipschitz subsection, both convex and bilipschitz Pareto fronts abide by the above lower bound. As a consequence, for either bilipschitz or convex Pareto fronts, the optimality gap evolves as $\Theta(1/p)$.

We empirically evaluated the optimality gap for p up to 1000 on six different Pareto fronts, among which three are convex and three are concave. We observe convergence of p times the optimality gap to a constant, even for non-bilipschitz and non-convex Pareto fronts.

The rest of the paper is organized as follows. In Section 2, we define formally the Pareto front, the hypervolume and the p-optimal distribution. Additionally, we introduce the concept of gap region. In Section 3, we derive first lower and upper bounds on the optimality gap, respectively for bilipschitz and for bilipschitz or convex Pareto fronts. In Section 4, we derive sufficient conditions on the objective-functions and the search space for the Pareto front to be convex and bilipschitz. Additionally, we generalize the lower bound for Pareto fronts with only a bilipschitz subsection. In Section 5, we examine the empirical optimality gap on six different Pareto fronts.

Notations. We denote the search space by Ω. We denote elements of Ω by X, which should not be confused with x denoting the first coordinate of a vector of the Pareto front. A vector of the objective space, v, is called feasible if $v \in F(\Omega)$. In order to avoid confusion, we always use the term *area* to refer to the Lebesgue-measure in dimension 2 and never to refer to a part of the objective space. Further notations are defined in the next section.

2 PRELIMINARIES

We focus on biobjective optimization, which aims at minimizing two objective functions, F_1 and F_2 over the search space $\Omega \subset \mathbb{R}^n$. We denote $F : X \in \Omega \mapsto (F_1(X), F_2(X))$.

2.1 Domination and Pareto front

A vector $u \in \mathbb{R}^2$ of the objective space is said to weakly dominate a vector $v \in \mathbb{R}^2$ when $u_1 \leq v_1$ and $u_2 \leq v_2$. We denote it $u \preceq v$. A vector u is said to dominate a vector v when $u \preceq v$ and $u \neq v$. We denote it $u \prec v$. If a vector u does not dominate a vector v, we denote it $u \nprec v$. The *Pareto front* is the set of all non-dominated feasible vectors: $\{u \in F(\Omega) : \forall v \in F(\Omega), v \nprec u\}$. We will assume here that the Pareto front has an explicit representation via f, namely that it can be written as $\{(x, f(x)) : x \in [x_{\min}; x_{\max}]\}$ with $x_{\min} \neq x_{\max}$. By definition of the Pareto front, f must be strictly decreasing. We say that f is $(L_{\min}, L_{\max})$-bilipschitz when $|f(x) - f(y)|$ is between $L_{\min} \times |x-y|$ and $L_{\max} \times |x-y|$ for all $x, y \in \mathbb{R}$. This is one of the main assumptions of interest here. We will also consider Pareto fronts for which f is convex and the (much) wider class of Pareto fronts with a bilipschitz subsection.

Given a reference point r, we denote the extremes of the part of the Pareto front dominating r with $\tilde{u}_{\min,r} := (\tilde{x}_{\min,r}, f(\tilde{x}_{\min,r}))$ and $\tilde{u}_{\max,r} := (\tilde{x}_{\max,r}, f(\tilde{x}_{\max,r}))$ where $\tilde{x}_{\min,r} := x_{\min}$ when $r_2 \geq f(x_{\min})$ and $f^{-1}(r_2)$ otherwise, and $\tilde{x}_{\max,r} := \min(x_{\max}, r_1)$, see Figure 1. In this paper, we will also assume that the reference point r is valid, in the sense that there exists a feasible vector of the objective space dominating r. The *nadir point* is the vector which

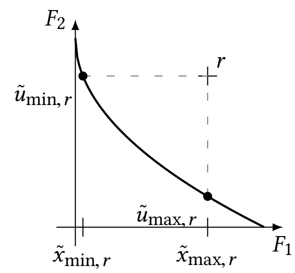


Figure 1: Depiction of $\tilde{u}_{\min,r}$ and $\tilde{u}_{\max,r}$ with $r = (0.7, 0.8)$ for the Pareto front associated with the function $f : x \mapsto 1 - \sqrt{x}$ for $x \in [0, 1]$.

coordinate in each objective is the worst value achieved on the Pareto front for this objective. It equals $(x_{\max}, f(x_{\min}))$.

2.2 Hypervolume and p-optimal distribution

The hypervolume is a set-quality indicator which depends on a reference point r. The hypervolume of a set S is the Lebesgue measure of the region weakly dominated by S and dominating r : $\lambda(\{u \in F(\Omega) : \exists s \in S, s \preceq u \prec r\})$. We denote it $\mathrm{HV}_r(S)$. The hypervolume improvement and hypervolume contribution of a vector u to a set S quantifies how much adding u to or removing u from the set S, respectively, affects its hypervolume: $\mathrm{HVI}_r(u, S) := \mathrm{HV}_r(S \cup \{u\}) - \mathrm{HV}_r(S)$ and $\mathrm{HVC}_r(u, S) := \mathrm{HVI}_r(u, S \setminus \{u\})$.

Here, we will study sets of p feasible vectors of the objective space maximizing the hypervolume. We call them *p-optimal distributions* and denote them S_r^p. We denote $v_{p,1}, \ldots, v_{p,p}$ the vectors of S_r^p ordered by increasing F_1 values and $x_{p,1}, \ldots, x_{p,p}$ their first coordinates. We also denote $x_{p,0} := \tilde{x}_{\min,r}$ and $x_{p,p+1} := \tilde{x}_{\max,r}$.

2.3 Gap regions

In this paper, we will examine the dependency of the *optimality gap* $\mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S_r^p)$ in p. The optimality gap is the area of the region of the objective space dominated by the Pareto front PF but not by the p-optimal distribution S_r^p. We call this region of the objective space *total gap region* and denote it $\mathcal{G}_r^p$. We can now write the optimality gap as $\lambda(\mathcal{G}_r^p)$. Since the p-optimal distribution S_r^p is a subset of the Pareto front, the total gap region can be decomposed into $p+1$ disjoints regions, that we call *gap regions*, see Figure 2. The

i-th gap region of S_r^p is the region of the objective space dominated by the Pareto front and dominating the reference point r_i^p with $r_1^p{:=}(x_{p,1}, r_2)$, $r_{p+1}^p{:=}(r_1, f(x_{p,p}))$ and $r_i^p{:=}(x_{p,i}, f(x_{p,i-1}))$ for all $i \in [\![2, p]\!]$. We denote it $\mathcal{G}_{r,i}^p$.

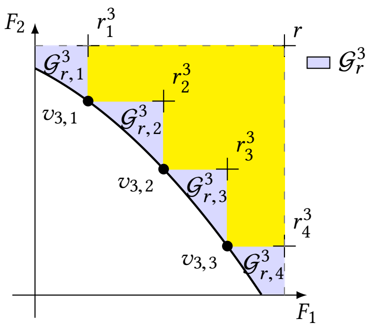


Figure 2: The four gap regions of the 3-optimal distribution S_r^3 for the concave and bilipschitz function $f : x \mapsto 1 - 0.5x - 0.5x^2$ for $x \in [0, 1]$.

3 THEORETICAL BOUNDS ON THE OPTIMALITY GAP $\mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S_r^p)$

In this section, we will prove both upper and lower theoretical bounds on the optimality gap $\mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S_r^p)$ when f is bilipschitz, and a theoretical upper bound when f is convex. The lower bound is generalized to a wider class of functions in Section 4.2. All bounds are equivalent to a constant times $1/p$.

3.1 Lower bound on the optimality gap

We will prove that if f is ($L_{\min}$,$L_{\max}$)-bilipschitz, then the optimality gap $\lambda(\mathcal{G}_r^p) = \mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S_r^p)$ is greater than $\frac{1}{p+1}$ times a constant depending only on the hypervolume of the Pareto front, on the reference point and on the bilipschitz constants $L_{\min}$ and $L_{\max}$.

Theorem 1. *If f restricted to $[\tilde{x}_{\min,r}, \tilde{x}_{\max,r}]$ is $(L_{\min}, L_{\max})$-bilipschitz, then the normalized optimality gap is bounded from below as*

$$\frac{\lambda(\mathcal{G}_r^p)}{HV_r(PF)} \geq \frac{1}{p+1} \times \frac{L_{\min}}{L_{\max}} \times \frac{1}{1 + 2 \times (q_1 + q_2 + q_1 \times q_2)} \tag{1}$$

with $q_1 := \frac{r_1 - \tilde{x}_{\max,r}}{\tilde{x}_{\max,r} - \tilde{x}_{\min,r}}$ and $q_2 := \frac{r_2 - f(\tilde{x}_{\min,r})}{f(\tilde{x}_{\min,r}) - f(\tilde{x}_{\max,r})}$.

In particular, when r dominates the nadir point $(x_{\max}, f(x_{\min}))$, both q_1 and q_2 equal 0, and thus the lower bound is simply $\frac{1}{p+1} \times \frac{L_{\min}}{L_{\max}}$.

Proof. We note $c_{p,i} := x_{p,i} - x_{p,i-1}$ for $i \in [\![1, p+1]\!]$, $\Delta_1 := r_1 - \tilde{x}_{\max,r}$ and $\Delta_2 := r_2 - f(\tilde{x}_{\min,r})$, see Figure 3.

The optimality gap $\mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S_r^p)$ is greater than $\sum_{i=1}^{p+1} \int_{x=x_{p,i-1}}^{x_{p,i}} (f(x) - f(x_{p,i}))dx$. Since the function f is decreasing and its restriction to $[\tilde{x}_{\min,r}, \tilde{x}_{\max,r}]$ is ($L_{\min}$,$L_{\max}$)-bilipschitz, $f(x) - f(x_{p,i})$ is greater than $L_{\min}$ times $x_{p,i} - x$. Therefore, the optimality gap $\lambda(S_r^p)$ is greater than $\frac{1}{2} \times L_{\min} \times \sum_{i=1}^{p+1} c_{p,i}^2$.

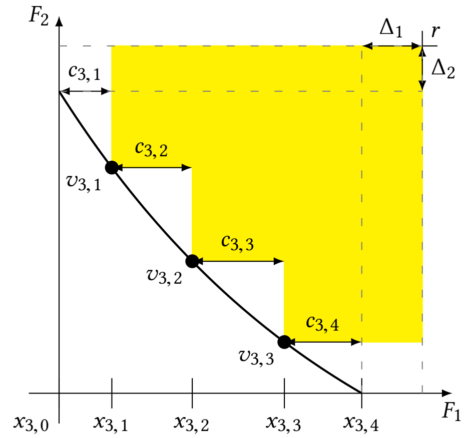


Figure 3: Illustration of the notations of the proof of Theorem 1 for the convex and bilipschitz function $f : x \mapsto \frac{e}{e-1} \times e^{-x} + 1 - \frac{e}{e-1}$ for $x \in [0, 1]$ and $r = (1.2, 1.15)$.

The hypervolume of the Pareto front is equal to

$$\Delta_1 \times (f(\tilde{x}_{\min,r}) - f(\tilde{x}_{\max,r})) + \Delta_2 \times (\tilde{x}_{\max,r} - \tilde{x}_{\min,r}) + \Delta_1 \times \Delta_2 + \int_{\tilde{x}_{\min,r}}^{\tilde{x}_{\max,r}} (f(x) - f(\tilde{x}_{\max,r}))dx.$$

The integral and $\Delta_1 \times (f(\tilde{x}_{\min,r}) - f(\tilde{x}_{\max,r}))$ are respectively smaller than $\frac{1}{2} \times L_{\max} \times (\tilde{x}_{\max,r} - \tilde{x}_{\min,r})^2$ and $\Delta_1 \times L_{\max} \times (\tilde{x}_{\max,r} - \tilde{x}_{\min,r})$. By rewriting $\tilde{x}_{\max,r} - \tilde{x}_{\min,r}$ as the sum over $i \in [\![1, p+1]\!]$ of $c_{p,i}$, we obtain that the optimality gap divided by the hypervolume of the Pareto front is greater than

$$\frac{\frac{1}{2} L_{\min} \times \sum_{i=1}^{p+1} c_{p,i}^2}{\frac{1}{2} L_{\max} \times (\sum_{i=1}^{p+1} c_{p,i})^2 + (\Delta_1 L_{\max} + \Delta_2 \times (\sum_{i=1}^{p+1} c_{p,i})) + \Delta_1 \Delta_2}.$$

The terms $\sum_{i=1}^{p+1} c_{p,i}^2$ and $\left(\sum_{i=0}^{p} c_{p,i}\right)^2$ being respectively $\|c_p\|_2^2$ and $\|c_p\|_1^2$ with $c_p := (c_{p,i})_{i \in [\![1,p+1]\!]} \in \mathbb{R}^{p+1}$, it is well-known that their ratio is superior to $\frac{1}{p+1}$. Therefore, the normalized optimality gap is greater than

$$\frac{1}{p+1} \times \frac{L_{\min}}{L_{\max} + 2 \times \frac{\Delta_1 \times L_{\max}}{\sum_{i=1}^{p+1} c_{p,i}} + 2 \times \frac{\Delta_2}{\sum_{i=1}^{p+1} c_{p,i}} + 2 \times \frac{\Delta_1 \times \Delta_2}{(\sum_{i=1}^{p+1} c_{p,i})^2}}.$$

We can rewrite back $\sum_{i=1}^{p+1} c_{p,i}$ as $\tilde{x}_{\max,r} - \tilde{x}_{\min,r}$. Additionally, since f is ($L_{\min}$,$L_{\max}$)-bilipschitz, $\frac{\Delta_2 \times L_{\max}}{f(\tilde{x}_{\max,r}) - f(\tilde{x}_{\min,r})}$ is greater than $\frac{\Delta_2}{\tilde{x}_{\max,r} - \tilde{x}_{\min,r}}$. $\square$

The larger $L_{\max}/L_{\min}$ is, the less information we have on the shape of ($L_{\min}$,$L_{\max}$)-bilipschitz Pareto fronts, and the looser is the bound. The quantities q_1 and q_2 reflect the distance between r and the nadir point normalized by the scale of the Pareto front, respectively in the first and in the second objective. When the nadir point dominates the reference point r, the bound gets looser as r moves away from the nadir point. In case of a linear front, we get the following tight result.

Corollary 1. *When r dominates the nadir point and $L_{\min} = L_{\max}$, the normalized optimality gap $\frac{\lambda(\mathcal{G}_r^p)}{HV_r(PF)}$ equals $\frac{1}{p+1}$, which corresponds to the lower bound* (1) *in Theorem 1.*

Proof. - In this case, the Pareto front is linear, and thus S_r^p is evenly distributed [1, Theorem 6]. As a consequence, the normalized optimality gap of the p-optimal distribution is exactly $1/(p+1)$, which is the lower bound given by Theorem 1. □

3.2 Upper bound on the optimality gap

We will prove that if f is either ($L_{\min}$,$L_{\max}$)-bilipschitz or convex, then the optimality gap is smaller than $\frac{p+1}{p^2}$ times a constant. The proof idea is largely inspired by the proof of Theorem 4.4 in [4]. This theorem gives an upper bound on the multiplicative approximation ratio of the p-optimal distribution of the form $1 + \frac{C}{p-4}$ with C being a constant depending on the Pareto front.

The following lemma is a natural extension of [4, Lemma 4.3], where we consider not only the $p-2$ inner vectors of the p-optimal distribution S_r^p but also its two extreme vectors. It states that the smallest hypervolume contribution of any element in S_r^p is below $1/p^2$ times a constant depending only on the extreme values of the Pareto front and on the reference point r.

Lemma 1. *For any f such that there exists a p-optimal distribution S_r^p, we have*

$$\min_{v \in S_r^p} HVC_r(v, S_r^p) \leq \frac{(r_1 - \tilde{x}_{\min,r})(r_2 - f(\tilde{x}_{\max,r}))}{p^2}. \tag{2}$$

Proof. We denote $a_{p,i} := x_{p,i} - x_{p,i-1}$ and $b_{p,i} := f(x_{p,i}) - f(x_{p,i-1})$ for $i \in [\![1, p+1]\!]$ except for $a_{p,p+1}$ and $b_{p,0}$ which are respectively $r_1 - x_{p,p}$ and $r_2 - f(x_{p,1})$, see Figure 4. For all $i \in [\![1, p]\!]$, the hypervolume contribution of $v_{p,i}$ to S_r^p is $a_{p,i+1} \times b_{p,i}$. In particular, this implies that for any i the quantity $a_{p,i+1}$ is greater than $\frac{\min_{v \in S_r^p} \mathrm{HVC}_r(v, S_r^p)}{b_{p,i}}$. Summing over i, we obtain that $(\sum_{i=1}^{p} \frac{1}{b_{p,i}}) \times \min_{v \in S_r^p} \mathrm{HVC}_r(v, S_r^p)$ is lower than $\sum_{i=1}^{p} a_{p,i+1}$, and thus lower than $r_1 - \tilde{x}_{\min,r}$. Additionally, the harmonic mean of the vector $(b_{p,i})_{i \in [\![1,p]\!]}$ is lower than its arithmetic mean:

$$\frac{p}{\sum_{i=1}^{p} \frac{1}{b_{p,i}}} \leq \frac{\sum_{i=1}^{p} b_{p,i}}{p}.$$

As a consequence, $\min_{v \in S_r^p} \mathrm{HVC}_r(v, S_r^p)$ is lower than $(r_1 - \tilde{x}_{\min,r})$ times $\frac{\sum_{i=1}^{p+1} b_{p,i}}{p^2}$, that is $\frac{(r_1 - \tilde{x}_{\min,r}) \times (r_2 - f(\tilde{x}_{\max,r}))}{p^2}$. □

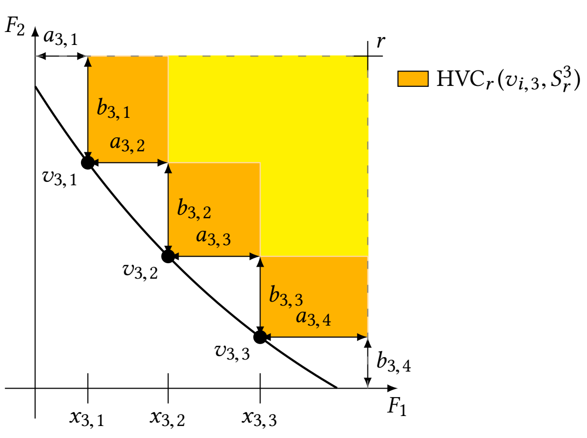


Figure 4: Illustration of the notations of the proof of Lemma 1 for $f : x \mapsto \frac{e}{e-1} \times e^{-x} + 1 - \frac{e}{e-1}$ for $x \in [0, 1]$ with $r = (1.1, 1.1)$ and $p = 3$.

We will also use the lower bounds on the maximum hypervolume of a single vector of the Pareto front for f either convex or ($L_{\min}$,$L_{\max}$)-bilipschitz stated in [10]. We recall these results below, with a slight reformulation, for sake of completeness.

Proposition 1. *If f restricted to $[\tilde{x}_{\min,r}, \tilde{x}_{\max,r}]$ is convex, we have*

$$\max_{u \in PF} HV_r(u) \geq \frac{1}{2} \times HV_r(PF). \tag{3}$$

If f restricted to $[\tilde{x}_{\min,r}, \tilde{x}_{\max,r}]$ is $(L_{\min}, L_{\max})$-bilipschitz, we have

$$\max_{u \in PF} HV_r(u) \geq \frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times HV_r(PF). \tag{4}$$

For all p, for all $i \in [\![1, p+1]\!]$, the hypervolume associated with the reference point r_i^p of a vector u is equal to $\mathrm{HVI}_r(u, S_r^p)$ when $u \in \mathcal{G}_{r,i}^p$ and 0 otherwise. Thus, $\max_{u \in \mathrm{PF}} \mathrm{HV}_{r_i^p}(u)$ equals $\max_{u \in \mathcal{G}_{r,i}^p} \mathrm{HVI}_r(u, S_r^p)$. Additionally, the hypervolume associated with r_i^p of the Pareto front is simply the area of the gap region $\mathcal{G}_{r,i}^p$. Using the optimality of the p-optimal distribution, we can deduce from Lemma 1 and Proposition 1 an upper bound on the area of any gap region $\mathcal{G}_{r,i}^p$, and thus an upper bound on the optimality gap at iteration p.

By applying Proposition 1 in the convex case, we obtain the following theorem.

Theorem 2. *If f restricted to $[\tilde{x}_{\min,r}, \tilde{x}_{\max,r}]$ is convex, then the optimality gap is bounded from above as*

$$\lambda(\mathcal{G}_r^p) \leq 2 \times (r_1 - \tilde{x}_{\min,r}) \times (r_2 - f(\tilde{x}_{\max,r})) \times \frac{p+1}{p^2} \tag{5}$$

Proof. The p-optimal distribution S_r^p is a set of p feasible vectors of the objective space maximizing the hypervolume. As a consequence, for any $u \in F(\Omega)$, for any $v \in S_r^p$, the hypervolume of $S_r^p \setminus \{v\} \cup \{u\}$ is lower than the hypervolume of S_r^p. In other words, the hypervolume improvement with respect to the set $S_r^p \setminus \{v\}$ of any feasible vector u is lower than the one of v itself. Additionally, for any feasible u, the hypervolume improvement of u to S_r^p is lower than the hypervolume improvement of u to $S_r^p \setminus \{v\}$. Indeed, they are equal to the area of the region dominated by u but not by respectively S_r^p and $S_r^p \setminus \{v\}$. As a consequence, for any feasible vector u, $\mathrm{HVI}_r(u, S_r^p)$ is lower than $\mathrm{HVI}_r(v, S_r^p \setminus \{v\})$, that is $\mathrm{HVC}_r(v, S_r^p)$.

Let $\mathcal{G}_{r,i}^p$ be the i-th gap region of S_r^p. By Proposition 1, the area of $\mathcal{G}_{r,i}^p$ is lower than $2 \times \max_{u \in \mathcal{G}_{r,i}^p} \mathrm{HVI}_r(u, S_r^p)$. We just proved that $\max_{u \in \mathcal{G}_{r,i}^p} \mathrm{HVI}_r(u, S_r^p)$ is lower than $\min_{v \in S_r^p} \mathrm{HVC}_r(v, S_r^p)$, and thus it is lower than $\frac{(r_1-\tilde{x}_{\min,r})(r_2-f(\tilde{x}_{\max,r}))}{p^2}$ by Lemma 1. Therefore, the area of any gap region $\mathcal{G}^p_{r,i}$ of S^p_r is lower than two times $\frac{(r_1-\tilde{x}_{\min,r})(r_2-f(\tilde{x}_{\max,r}))}{p^2}$. Since the optimality gap $\mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S^p_r)$ is the sum of the areas of the $p+1$ gap regions of S^p_r, we can conclude. □

By applying Proposition 1 in the bilipschitz case instead of the convex case, we obtain the following looser upper bound on the optimality gap.

Theorem 3. *If f restricted to $[\tilde{x}_{\min,r}, \tilde{x}_{\max,r}]$ is $(L_{\min}, L_{\max})$-bilipschitz, then the optimality gap is bounded from above by*

$$\lambda(\mathcal{G}^p_r) \leq 2 \times \frac{L_{\max}}{L_{\min}} \times (r_1 - \tilde{x}_{\min,r}) \times (r_2 - f(\tilde{x}_{\max,r})) \times \frac{p+1}{p^2} \tag{6}$$

4 SUFFICIENT ASSUMPTIONS AND GENERALIZATION OF THE LOWER BOUND

In this section, we give sufficient conditions for deriving bounds on the optimality gap from Theorems 1 and 3. We also generalize Theorem 1 to Pareto fronts that only need to have some bilipschitz subsection.

4.1 Sufficient assumptions on the objective functions

First, we will examine the bilipschitz assumption, under which we have both a lower and an upper bound on the optimality gap. It is simple to prove that, as soon as both objective functions F_1 and F_2 are bilipschitz, the function f characterizing the Pareto front is bilipschitz too. We will say that F_i is ($L_{\min}$,$L_{\max}$)-bilipschitz when for all $X, Y \in \Omega$, the quantity $|F_i(Y) - F_i(X)|$ is between $L_{\min} \times \|Y - X\|$ and $L_{\max} \times \|Y - X\|$.

Proposition 2. *If F_1 and F_2 are respectively ($L_{\min,1}$, $L_{\max,1}$)-bilipschitz and ($L_{\min,2}$, $L_{\max,2}$)-bilipschitz.*
Then, f is $(\frac{L_{\min,2}}{L_{\max,1}}, \frac{L_{\max,2}}{L_{\min,1}})$-bilipschitz.

Proof. Let x, y be in $[x_{\min}, x_{\max}]$. Since the vectors $(x, f(x))$ and $(y, f(y))$ belong to the Pareto front, they are feasible, and thus there exist $X, Y \in \Omega$ such that $(x, f(x)) = (F_1(X), F_2(X))$ and $(y, f(y)) = (F_1(Y), F_2(Y))$. Thus, $|f(y) - f(x)|$ equals $|F_2(Y) - F_2(X)|$, which is superior to $L_{\min,2}$ times $\|Y - X\|$, which is superior to $\frac{L_{\min,2}}{L_{\max,1}}$ times $|F_1(Y) - F_1(X)|$, that is $|x - y|$. Conversely, $|x-y|$ is superior to $\frac{L_{\min,1}}{L_{\max,2}}$ times $|f(y) - f(x)|$. □

However, as soon as an objective function has a critical point in the interior of the search space Ω, it is not bilipschitz. In particular, this is the case when the objective function is derivable and has a local minimum. In that case, there is no guarantee that f will be bilipschitz. Setting the reference point such that the optimum in question is excluded can account for this problem. From a practical perspective, this is not a strong restriction.

Likewise, the convexity assumption on f is met as soon as both objective functions F_1 and F_2 are convex. This is a known result (see [7, p68]), but since we did not find a proof in the literature, we include the proof below.

Proposition 3. *If the search space Ω and the objective functions F_1 and F_2 are convex, then f is convex.*

Proof. Let $u := (u_1, u_2)$ and $v := (v_1, v_2)$ be two vectors of the epigraph of f. Since the vectors $(u_1, f(u_1))$ and $(v_1, f(v_1))$ belong to the Pareto front, they are feasible, and thus there exist $X, Y \in \Omega$ such that $(u_1, f(u_1)) = (F_1(X), F_2(X))$ and $(v_1, f(v_1)) = (F_1(Y), F_2(Y))$. Let note $Z := \frac{X+Y}{2}$. By convexity of Ω, Z also belongs to Ω. By convexity of F_1, $F_1(Z)$ is smaller than $\frac{F_1(X)+F_1(Y)}{2}$, that is $\frac{u_1+v_1}{2}$. Therefore, f being decreasing, $f\left(\frac{u_1+v_1}{2}\right)$ is smaller than $f(F_1(Z))$, that is $F_2(Z)$. By convexity of F_2, $F_2(Z)$ is smaller than $\frac{F_2(X)+F_2(Y)}{2}$, that is $\frac{f(u_1)+f(v_1)}{2}$ and thus than $\frac{u_2+v_2}{2}$ since u and v belong to the epigraph of f. Therefore, $\frac{u+v}{2}$ also belongs to epi$f$. We can conclude that the epigraph of f, and thus the function f itself, are convex. □

Convexity of each objective is a sufficient but not a necessary condition. For example, the Pareto front of the test problem ZDT1 [14] is convex, while the second objective function is not.

4.2 Generalization of the lower bound to functions with a bilipschitz subsection

We prove that the optimality gap associated with any reference point r' dominating the reference point r provides a lower bound on the optimality gap associated with r.

Lemma 2. *Given a reference point r' that dominates the reference point r. The optimality gap $HV_r(PF) - HV_r(S^p_r)$ associated with r is bounded from below by the optimality gap $HV_{r'}(PF) - HV_{r'}(S^p_{r'})$ associated with r'.*

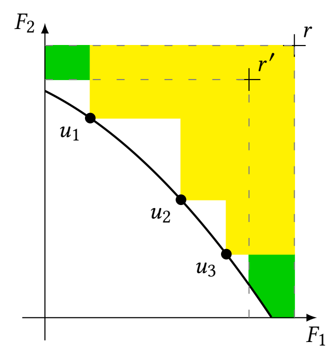


Figure 5: Illustration of the proof of Lemma 2.

Proof. For any set S, the difference $\mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S)$ is greater than $\mathrm{HV}_{r'}(\mathrm{PF}) - \mathrm{HV}_{r'}(S)$, see Figure 5. Indeed, the difference between the hypervolumes of S and of the Pareto front is the area of the intersection of the region dominated by the Pareto front but not by the set S and the region dominating the reference point. Since r' dominates r, the region of the objective space dominating r' is included in the region dominating r. Therefore, $\mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S^p_r)$

is greater than $\mathrm{HV}_{r'}(\mathrm{PF}) - \mathrm{HV}_{r'}(S_r^p)$, which is itself greater than $\mathrm{HV}_{r'}(\mathrm{PF}) - \mathrm{HV}_{r'}(S_{r'}^p)$ by definition of S_r^p. □

Hence, to get a lower bound on the optimality gap for any reference point r, it suffices to find a reference point r' such that Theorem 1 applies and r' dominates r. As a consequence, as soon as any part of the Pareto front dominating r is bilipschitz, see Figure 6, Theorem 1 provides a, generally non-tight, lower bound on the optimality gap for any reference point that covers at least some part of a bilipschitz subsection.

THEOREM 4. *Assume there exists a reference point r' dominating both r and the nadir point such that f restricted to $[\tilde{x}_{\min,r}, \tilde{x}_{\max,r}]$ is ($L_{\min}$,$L_{\max}$)-bilipschitz. Then, the optimality gap $HV_r(PF) - HV_r(S_r^p)$ is greater than $HV_{r'}(PF) \times \frac{L_{\min}}{L_{\max}} \times \frac{1}{p+1}$.*

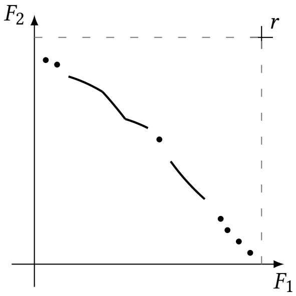


Figure 6: An example of Pareto front with a bilipschitz section dominating the reference point r. Both continuous sections are bilipschitz.

In particular, this is the case when the function f is either convex or concave. Indeed, since f is strictly decreasing, it has finite and nonzero left and right derivatives everywhere outside the extremes $x_{\min}$ and $x_{\max}$. This generalization of the lower bound extends the scope of the study to non-continuous Pareto fronts such as piecewise continuous Pareto fronts. Lemma 2 and Theorem 1 also provide a way to find a lower bound for Pareto fronts which do not even have an explicit formula, contrary to the assumptions detailed in the preliminaries. It suffices that a part of the Pareto front has such a characterization.

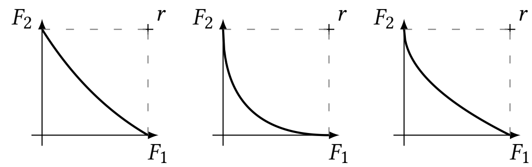
5 EXPERIMENTAL RESULTS

In this section, we will compare the dependency in p of the optimality gap $\mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S_r^p)$ and the theoretical bounds on six different Pareto fronts.

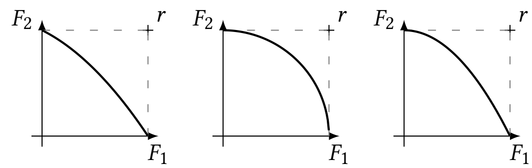
5.1 Benchmark Pareto fronts

We will look at six different Pareto fronts, among which three are convex and three are concave, see Figure 7 (a)-(f).

The **doublesphere** Pareto front (b) corresponds to the objective functions $F_1 : X \mapsto \|X - X_1^*\|_2^2$ and $F_2 : X \mapsto \|X - X_2^*\|_2^2$ with $\|X_1^* - X_2^*\|_2 = 1$, see [11]. **zdt1** (c) and **zdt2** (f) belong to the ZDT test suite [14] while **dtlz2** (e) belongs to the DTLZ test suite [6]. None of these Pareto fronts are bilipschitz. For this reason, we construct ourselves **convex-bil** (a) and **concave-bil** (d) to



(a) convex-bil **(b) doublesphere** **(c) zdt1**



(d) concave-bil **(e) dtlz2** **(f) zdt2**

Figure 7: The benchmark Pareto fronts, corresponding to the functions (a) : $f : x \mapsto \frac{e}{e-1} \times e^{-x} + 1 - \frac{e}{e-1}$ (b) : $f : x \mapsto 1 + x - 2 \times \sqrt{x}$ (c) : $f : x \mapsto 1 - \sqrt{x}$ (d) : $f : x \mapsto 1 - 0.5x - 0.5x^2$ (e) : $f : x \mapsto \sqrt{1 - x^2}$ (f) : $f : x \mapsto 1 - x^2$.

be bilipschitz convex and concave Pareto fronts, respectively. We can easily build constrained multi-objective optimization problems whose Pareto fronts are respectively **convex-bil** and **concave-bil**. For example, these Pareto fronts correspond to the problem of minimizing $F_1 : X \mapsto X_1$ and $F_2 : X \mapsto f(X_1) + \sum_{i=2}^{n} X_i$ for $X \in [0,1]^n$ with $f : x \mapsto \frac{e}{e-1} \times e^{-x} + 1 - \frac{e}{e-1}$ and $f : x \mapsto 1 - 0.5x - 0.5x^2$, respectively. The letter n represents the dimension of the search space.

We chose these Pareto fronts because they have a known analytic formula. It allows us to estimate the optimality gap for a p-optimal distribution for high p in reasonable time with high confidence.

5.2 Computation of a p-optimal distribution and the optimality gap

Since S_r^p is a subset of the Pareto front, if we know the explicit representation of the Pareto front $\{(x, f(x)) : x \in [x_{\min}, x_{\max}]\}$, a p-optimal distribution S_r^p can be obtained from the first coordinates of its vectors, that is, $x_{p,i}$ for $i \in [\![1, p]\!]$. They are a solution of

$$\max_{x_1,\ldots,x_p \in [x_{\min}, x_{\max}]} \mathrm{HV}_r(\{(x_1, f(x_1)), \ldots, (x_p, f(x_p))\}) \ .$$

We do not solve this problem directly. We exploit the following parametrization to solve it faster: $\delta_i := x_i - x_{i-1}$ for all $i \in [\![2, p-1]\!]$, $\delta_1 := x_1 - x_{\min}$ and $\delta_p := x_{\max} - x_p$, see Figure 8. For **dtlz2**, we use a slightly different parametrization to cancel the bad conditioning: $\delta_p' := \mathrm{sgn}(\delta_p) \times \sqrt{|\delta_p|}$.

We use the algorithm CMA-ES with bounds between 0 and 1. These bounds do not guarantee that the x_i corresponding to the δ_i are in $[0,1]$. We ensure that a x_i outside $[0,1]$ does not contribute to the hypervolume by setting $f(x) = r_2$ outside $[0,1]$. Additionally, to prevent having a flat fitness for δ_i such that the corresponding x_i are outside $[0,1]$, we add the penalization $\sum_{i=1}^{p}(x-1)^2 \mathbf{1}_{x_i>1} + x^2 \mathbf{1}_{x_i<0}$. The source code is available at https://github.com/eugeniemarescaux/gecco2021.

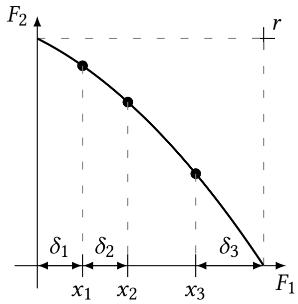


Figure 8: Illustration of the reparametrization of a set of p points of the Pareto front by δ.

5.3 Results

Figure 9 shows p times the optimality gap of a p-optimal distribution for p between 1 and 1000. In the same figure, we also plotted for comparison p times the theoretical lower and upper bounds described in Section 3. For non-bilipschitz Pareto fronts, we exploit the generalization of the theoretical lower bound done in Section 4.2. For Pareto fronts which are neither bilipschitz nor convex, since there is no known theoretical upper bound, only two curves are represented. For a reference point r equal to the nadir point $(1, 1)$, i.e. in the first and third column of Figure 9, we see a close to constant shift between the curves of the empirical optimality gap and of the theoretical lower bound. Hence, the optimality gap evolves roughly as $1/(p + 1)$ even for the smallest values of p. For a reference point r equal to $(11, 11)$, i.e. in the second and fourth column of Figure 9, the shift between the empirical curve and the theoretical bound curve takes very different values for p small. However, the empirical curve seems to converge to a horizontal line, which is a marker that p times the optimality gap of S_r^p converges to a constant. For every Pareto front, the curve of the empirical optimality gap is almost horizontal at least for p greater than 10.

In the proof of Theorem 1, more precisely in the first phrase after the introduction of the notations, we neglected the area of the part of the total gap region which does not dominate the nadir point. It could explain why the optimality gap decreases with p in the same way as the theoretical lower bound for r equal to the nadir point $(1, 1)$ but not for r equal to $(11, 11)$.

6 CONCLUSION

We have proven that for a wide class of biobjective Pareto fronts, the hypervolume of all the solutions visited by a biobjective optimization algorithm converges to the hypervolume of the Pareto front in $\Omega(1/p)$. This is true for any algorithm, but also for objective functions as easy as convex quadratic functions, and for a search space of any dimension. The maximum rate of convergence to the entire Pareto front is slow compared to the convergence rate observed when converging to a single point in the Pareto set or, likewise, in single-objective optimization.

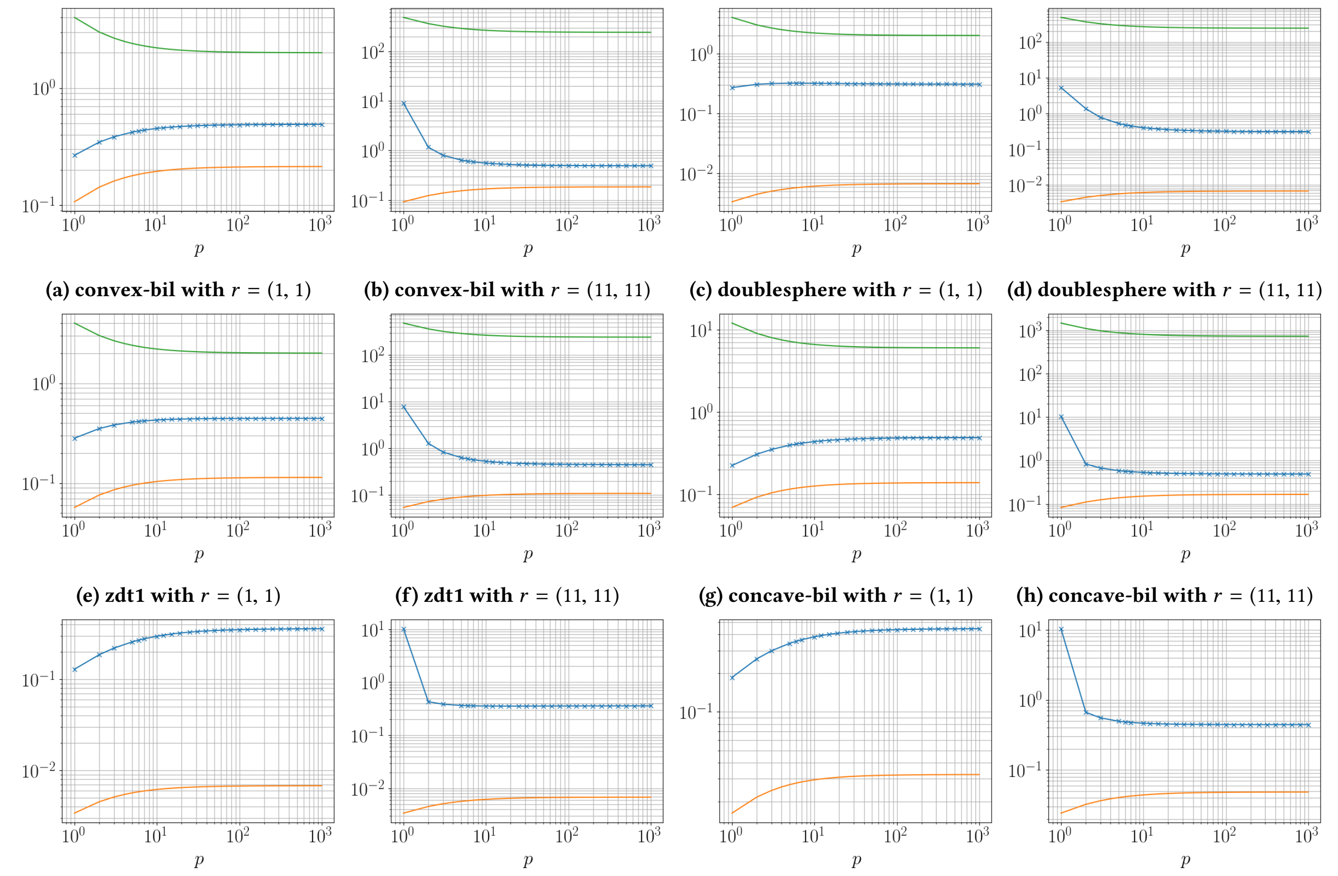
Several evolutionary algorithms achieve linear convergence on convex-quadratic functions [2]: their distance to the optimum stays close to $1/\alpha^p$ for some $\alpha > 1$ (typical convergence speeds depend on the search space dimension n and α rarely exceeds $1.2^{1/n}$). For random-search algorithms, the precision ϵ evolves on most functions as $\Theta(1/p^{1/n})$ with p being here the expected number of evaluations to reach the ball of radius ϵ, see [2, Theorem 10.8].

A convergence rate which would seem slow for a single-objective algorithm does not come from an inefficient algorithm or from hard to optimize objective functions. The slowness is inherent to the set-quality indicators used. The simplicity of the proof, the fact that the convergence rate is $\Omega(1/p)$ for both the hypervolume indicator and the multiplicative approximation ratio, and the very general (weak) assumptions suggest that this is a fundamental limitation on the convergence rate in multi-objective optimization. This is however not a very surprising result when considering that, contrary to single-objective optimization, the goal is not to approximate a single vector, but an entire set.

We empirically observed on six functions that p times the optimality gap converges rapidly to a constant, even for the two functions for which we have no upper bound. We suspect that the optimality gap is equivalent to a constant times $1/p$ on most if not all biobjective optimization problems with a partially continuous Pareto front.

In general, theoretical lower bounds are quite useful in algorithm development. Apart from designing algorithms that actually reach the lower bound, they can in particular avoid futile but long lasting attempts to improve algorithms that already reach the bound. This requires a non-asymptotic bound, as presented in this paper. Yet, the question arises how the lower bound on the optimality gap of the p-optimal distribution transfers to a practical algorithm. A practical algorithm faces at least two additional problems: it can only approximate any p-optimal distribution (but never reaches it), and the intersection between p-optimal distributions for different values of p is often small. Therefore, the presented bounds need to be carefully combined with bounds on the convergence speed towards points in the p-optimal distribution. We have currently no conjecture as to whether a convergence rate of $\Theta(1/p)$ is achievable by a real biobjective algorithm.

For finite discrete Pareto fronts, we cannot talk of a convergence rate in $\theta(1/p)$ since p is bounded by the size of the Pareto front, that we will denote N. In that case, it is trivial that for $p \neq N$ and for C_1 and C_2 respectively small and large enough, the optimality gap of the p-optimal distribution is between $C_1 \times 1/p$ and $C_2 \times 1/p$. However, we can still expect that for some discrete Pareto fronts the optimality gap of the p-optimal distribution will resemble $CONST \times 1/(p + CONST)$ for medium values of p. Indeed, we expect that for regular discretization of continous Pareto fronts, the impact on the optimality gap of the lack of precision of the approximation of a p-optimal distribution is negligable for $p \ll N$. Thus, the optimality gaps for the discrete approximation and for the original continuous Pareto front should be alike for these values of p. For example, consider a discrete Pareto front PF_d which is the regular discretization of a continuous linear Pareto front PF_c. If N is even, then the optimality gap for PF_d of any p-optimal distribution with p even is exactly $(1/(p+1) - 1/(N+1)) \times \mathrm{HV}_r(\mathrm{PF}_c)$. For $p \ll N$, it is close to $1/(p+1) \times \mathrm{HV}_r(\mathrm{PF}_c)$, the optimality gap in the continuous case.



(a) convex-bil with $r = (1, 1)$ **(b) convex-bil with $r = (11, 11)$** **(c) doublesphere with $r = (1, 1)$** **(d) doublesphere with $r = (11, 11)$**

(e) zdt1 with $r = (1, 1)$ **(f) zdt1 with $r = (11, 11)$** **(g) concave-bil with $r = (1, 1)$** **(h) concave-bil with $r = (11, 11)$**

(i) dtlz2 with $r = (1, 1)$ **(j) dtlz2 with $r = (11, 11)$** **(k) zdt1 with $r = (1, 1)$** **(l) zdt2 with $r = (11, 11)$**

Figure 9: Comparison between p times the empirical optimality gap $\mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S_r^p)$ and its theoretical bounds on the Pareto fronts described in Section 5.1. The empirical optimality gap, $\mathrm{HV}_r(\mathrm{PF}) - \mathrm{HV}_r(S_r^p)$ is represented in blue and the theoretical lower and upper bounds are represented in orange and green, respectively.

REFERENCES

[1] Anne Auger, Johannes Bader, Dimo Brockhoff, and Eckart Zitzler. 2009. Theory of the Hypervolume Indicator: Optimal μ-Distributions and the Choice of the Reference Point. In *Proceedings of the Tenth ACM SIGEVO Workshop on Foundations of Genetic Algorithms (FOGA '09)*. Association for Computing Machinery, New York, NY, USA, 87–102. https://doi.org/10.1145/1527125.1527138

[2] A. Auger and N. Hansen. 2011. Theory of Evolution Strategies: A New Perspective. In *Theory of Randomized Search Heuristics: Foundations and Recent Developments*, A. Auger and B. Doerr (Eds.). World Scientific Publishing, Singapore, Chapter 10, 289–325.

[3] Nicola Beume, Marco Laumanns, and Günter Rudolph. 2010. Convergence Rates of (1+1) Evolutionary Multiobjective Optimization Algorithms. In *Parallel Problem Solving from Nature, PPSN XI (Lecture Notes in Computer Science)*, Robert Schaefer, Carlos Cotta, Joanna Kołodziej, and Günter Rudolph (Eds.). Springer, Berlin, Heidelberg, 597–606. https://doi.org/10.1007/978-3-642-15844-5_60

[4] Karl Bringmann and Tobias Friedrich. 2010. The Maximum Hypervolume Set Yields Near-Optimal Approximation. In *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation - GECCO '10*. ACM Press, Portland, Oregon, USA, 511. https://doi.org/10.1145/1830483.1830576

[5] A. L. Custódio and J. F. A. Madeira. 2018. MultiGLODS: Global and Local Multiobjective Optimization Using Direct Search. *J Glob Optim* 72, 2 (Oct. 2018), 323–345. https://doi.org/10.1007/s10898-018-0618-1

[6] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. 2002. Scalable Multi-Objective Optimization Test Problems. In *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*, Vol. 1. IEEE, Honolulu, HI, USA, USA, 825–830 vol.1. https://doi.org/10.1109/CEC.2002.1007032

[7] Matthias Ehrgott. 2005. *Multicriteria Optimization* (second ed.). Springer-Verlag, Berlin Heidelberg. https://doi.org/10.1007/3-540-27659-9

[8] J. Fliege, L. M. Graña Drummond, and B. F. Svaiter. 2009. Newton's Method for Multiobjective Optimization. *SIAM J. Optim.* 20, 2 (Jan. 2009), 602–626. https://doi.org/10.1137/08071692X

[9] Ellen H. Fukuda and L. M. Graña Drummond. 2011. On the Convergence of the Projected Gradient Method for Vector Optimization. *Optimization* 60, 8-9 (Aug. 2011), 1009–1021. https://doi.org/10.1080/02331934.2010.522710

[10] Eugénie Marescaux and Anne Auger. 2021. Multiobjective Hypervolume Based ISOOMOO Algorithms Converge with At Least Sublinear Speed to the Entire Pareto Front. (2021). hal-03198414.

[11] Cheikh Toure, Anne Auger, Dimo Brockhoff, and Nikolaus Hansen. 2019. On Bi-Objective Convex-Quadratic Problems. In *Evolutionary Multi-Criterion Optimization (Lecture Notes in Computer Science)*, Kalyanmoy Deb, Erik Goodman, Carlos A. Coello Coello, Kathrin Klamroth, Kaisa Miettinen, Sanaz Mostaghim, and Patrick Reed (Eds.). Springer International Publishing, Cham, 3–14. https://doi.org/10.1007/978-3-030-12598-1_1

[12] David A Van Veldhuizen and Gary B Lamont. 1998. Evolutionary Computation and Convergence to a Pareto Front. In *Proceedings of the Third Annual Conference on Genetic Programming*. Stanford University Bookstore, San Francisco, 221–228.

[13] Eckart Zitzler, Dimo Brockhoff, and Lothar Thiele. 2007. The Hypervolume Indicator Revisited: On the Design of Pareto-Compliant Indicators Via Weighted Integration. In *Evolutionary Multi-Criterion Optimization*, Shigeru Obayashi, Kalyanmoy Deb, Carlo Poloni, Tomoyuki Hiroyasu, and Tadahiko Murata (Eds.). Vol. 4403. Springer Berlin Heidelberg, Berlin, Heidelberg, 862–876. https://doi.org/10.1007/978-3-540-70928-2_64

[14] Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. 2000. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation* 8, 2 (June 2000), 173–195. https://doi.org/10.1162/106365600568202

[15] Eckart Zitzler, Joshua Knowles, and Lothar Thiele. 2008. Quality Assessment of Pareto Set Approximations. In *Multiobjective Optimization: Interactive and Evolutionary Approaches*, Jürgen Branke, Kalyanmoy Deb, Kaisa Miettinen, and Roman Słowiński (Eds.). Springer, Berlin, Heidelberg, 373–404. https://doi.org/10.1007/978-3-540-88908-3_14

[16] E. Zitzler, L. Thiele, M. Laumanns, C.M. Fonseca, and V.G. da Fonseca. 2003. Performance Assessment of Multiobjective Optimizers: An Analysis and Review. *IEEE Transactions on Evolutionary Computation* 7, 2 (April 2003), 117–132. https://doi.org/10.1109/TEVC.2003.810758