Multiobjective Hypervolume Based ISOOMOO Algorithms Converge with At Least Sublinear Speed to the Entire Pareto Front

Eugénie Marescaux, Anne Auger

MULTIOBJECTIVE HYPERVOLUME BASED ISOOMOO ALGORITHMS CONVERGE WITH AT LEAST SUBLINEAR SPEED TO THE ENTIRE PARETO FRONT

EUGÉNIE MARESCAUX AND ANNE AUGER

Abstract. In multiobjective optimization, one is interested in finding a good approximation of the Pareto set and the Pareto front, i.e the sets of best compromises in the decision and objective spaces, respectively. In this context, we introduce a new algorithm framework, Incremental SingleObjective Optimization for MultiObjective Optimization (ISOOMOO) for approximating the Pareto front with an increasing number of points. We focus on HV-ISOOMOO, its instanciation with the hypervolume indicator, a set-quality indicator which is widely used for algorithms design and performance assessment. HV-ISOOMOO algorithms approximate the Pareto front by greedily maximizing the hypervolume. We study the convergence to the entire Pareto front of HV-ISOOMOO coupled with a perfect singleobjective optimizer. The convergence is defined as the convergence of the hypervolume of the sets of all meta-iterations incumbents towards the hypervolume of the Pareto front. We prove tight lower bounds on the convergence-speed for convex and bilipschitz Pareto fronts in $O(1/n^c)$ with n being the number of meta-iterations and $c = 1$ and $c \leq 1$, respectively. For convex Pareto fronts, the convergence rate is exactly in $\Theta(1/n)$, namely the highest convergence rate achievable by a biobjective optimization algorithm. These are the first results on the convergence-speed of multiobjective optimization algorithms towards the entire Pareto front. We also analyze theoretically the asymptotic convergence behavior.

Key words. multiobjective optimization, convergence, hypervolume, pareto front

AMS subject classifications. 90C29, 90C30

1. Introduction. Real-world problems often involve the optimization of several conflicting objectives. The solution of such problems is the set of non-dominated decision vectors, the Pareto set defined as the set of solutions that cannot be improved along one objective without degrading along another one. Its image in the objective-space is the Pareto front. A decision maker is then often involved to choose, based on its preferences, a single best compromise. The shape of the Pareto front informs on the trade-off between objectives. Many algorithms such as evolutionary algorithms approximate the Pareto front with a number of points fixed in the beginning. But some algorithms, in particular Direct Multi-Search Methods such as MultiGLODS [9], DMS [10] and D-Multi-MADS [6] aim at approximating the entire Pareto set or Pareto front with as many well-distributed points as possible. Ideally, the quality of the Pareto front approximation increases with time without stagnating and such algorithms can be considered as *anytime* algorithms.

The convergence speed towards a critical decision vector or a vector of the Pareto front have been examined for many algorithms such as $(1+1)$ evolutionary multiobjective algorithms [4] or Newton's method [13]. Convergence rates are typically similar to the ones obtained for singleobjective optimization. Indeed, this is still a convergence towards a single point. Its analysis is sometimes reduced to the study of the convergence of a singleobjective optimization algorithm. The convergence of anytime algorithms towards the whole Pareto set or front is of a different kind because these are sets and not points. It has already been theoretically investigated for some algorithms [9] and more abstract frameworks [18], but analysis of the rate of convergence are missing. Additionally, empirical studies typically focus on determining which algorithm is faster and do not provide orders of convergence or precise convergence rate. Yet, while largely overlooked, studying convergence rates either theoretically or empirically is crucial. In this context, it has been

proven that convergence towards the whole Pareto front is always sublinear in the number of function evaluations, at least when measuring convergence with the hypervolume indicator [15] or the multiplicative ϵ-indicator [7], and thus much slower than typical speeds of convergence to a single point. The hypervolume is a set-quality indicator widely used in multiobjective optimization, both to guide algorithms and for performance assessment. It is at the core of all known stricly Pareto-compliant indicators [19].

In this paper, we introduce a new anytime algorithm framework, Incremental SingleObjective Optimization for MultiObjective Optimization (ISOOMOO). We focus on its instanciation with the hypervolume indicator, HV-ISOOMOO. HV-ISOOMOO algorithms try to greedily maximize the hypervolume by adding points with the highest hypervolume contribution. These points are obtained by running a singleobjective optimizer. Greediness is already used in the selection part of some multiobjective optimization evolutionary algorithms such as SMS-EMOA [5] to approximate the set of p points among the population with the highest hypervolume. The hypervolume of such discrete greedy approximation is proven to be at least $(e-1)/e$ times the one of the p-optimal distribution [16]. To our knowledge, we provide the first continuous equivalent of this result. We investigate the rate of convergence of HV-ISOOMOO towards the whole Pareto front in the ideal case of perfect singleobjective optimization, measuring the convergence with the hypervolume. For convex and bilipschitz Pareto fronts, we prove that the convergence-speed is in $O(1/n^c)$ with $c = 1$ and $c \leq 1$, respectively, with n being the number of singleobjective optimization runs performed. For convex Pareto fronts, since it has already been proven that no biobjective optimization algorithm can converge to the Pareto front faster than in $\Omega(1/n)$ [15], the convergence rate is exactly in $\Theta(1/n)$. Additionally, we prove that for bilipschitz and smooth enough Pareto fronts doubling the number of points in the approximation halves the optimality gap, asymptotically. In the proof process, we obtain bounds on the normalized maximum hypervolume and a geometric interpretation of optimality conditions.

The paper is organized as follows. In Section 2, we lay the foundations of the problem we investigate. In Section 3, we prove preliminary results later used to investigate convergence. In Section 4, we derive lower bounds on the convergence rate of the perfect version of HV-ISOOMOO for bilipschitz and convex Pareto fronts and an insight on the asymptotic convergence behavior.

Notations and conventions. For $a, b \in \mathbb{N}$, we note $[\![a; b]\!]$ the set $\{a, a+1, \dots, b-1, b\}$. For a vector $u \in \mathbb{R}^2$, we note u_1 and u_2 respectively its first and its second coordinate. If the vector notation already contains an index, we separate the two indices with a comma. For simplicity sake, we often replace the set $\{u\}$ by u in the notations. We say that a function $f : \mathbb{R} \to \mathbb{R}$ is decreasing (respectively strictly decreasing) when for all $x < y$, we have $f(x) \geq f(y)$ (respectively $f(x) > f(y)$).

2. Background, algorithm framework and assumptions. Here, we lay the foundations of the problem we analyze. First, we recall some classic concepts of multiobjective optimization. Then, we introduce the ISOOMOO class of algorithms and its hypervolume based instanciation HV-ISOOMOO. We also formalize a mathematical abstraction of HV-ISOOMOO, the greedy set sequences. Finally, we examine our assumptions on the biobjective optimization problem.

2.1. Biobjective optimization problems, the Pareto front and the hypervolume indicator. We consider a biobjective minimization problem:

$$\min_{x \in \Omega \subset \mathbb{R}^d} F(x) \tag{2.1}$$

with $F : \Omega \subset \mathbb{R}^d \to \mathbb{R}^2 : x \mapsto (F_1(x), F_2(x))$. We define two dominance relations for vectors in the objective space. We say that u weakly dominates v denoted by $u \preceq v$ if $u_1 \leq v_1$ and $u_2 \leq v_2$ and that u dominates v denoted by $u \prec v$ if $u \preceq v$ and $u \neq v$. A vector of the objective-space $\mathbb{R}^2$ is said *feasible* when it belongs to $F(\Omega)$. Solving the optimization problem consists in finding a good approximation of the *Pareto front*, the set of non-dominated feasible vectors, $\{F(X) : X \in \Omega,\ \forall Y \in \Omega, F(Y) \not\preceq F(X)\}$. We restrict ourselves to Pareto fronts with an explicit representation:

$$\mathrm{PF}_f = \{(x, f(x)) : x \in [x_{\min}, x_{\max}]\} \tag{2.2}$$

with $f : \mathbb{R} \mapsto \mathbb{R}$ decreasing. We denote by $u_{\min} := (x_{\min}, f(x_{\min}))$ and $u_{\max} := (x_{\max}, f(x_{\max}))$ the extreme vectors of the Pareto front. Likewise, we denote by $\tilde{u}_{\min,r} := (\tilde{x}_{\min,r}, f(\tilde{x}_{\min,r}))$ and $\tilde{u}_{\max,r} := (\tilde{x}_{\max,r}, f(\tilde{x}_{\max,r}))$ the extremes vectors of the part of the Pareto front dominating a reference point r, with $\tilde{x}_{\min,r} := \max(x_{\min}, f^{-1}(r_2))$ and $\tilde{x}_{\max,r} := \min(x_{\max}, r_1)$. The vector $(x_{\max}, f(x_{\min}))$ is called the *nadir* point. All these notations are illustrated in Figure 1.

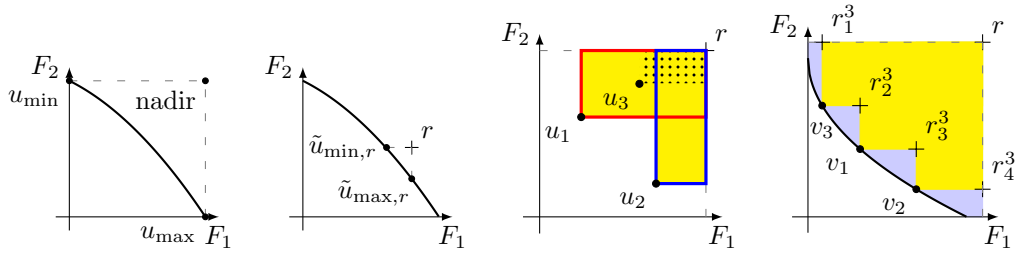

Fig. 1: Illustration of notations. The extreme vectors $u_{\min}$ and $u_{\max}$ and the nadir point $(x_{\max}, f(x_{\min}))$ (leftmost) ; the extreme vectors relative to the reference point r $\tilde{u}_{\min,r}$ and $\tilde{u}_{\max,r}$ (left) ; three vectors u_1, u_2 and u_3 and the regions weakly dominated by them and dominating r, $\mathcal{D}^r_{u_1}$, $\mathcal{D}^r_{u_2}$ and $\mathcal{D}^r_{u_3}$ (right) ; the greedy set $\mathcal{S}_3 = \{v_1, v_2, v_3\}$, its four gap regions and the associated reference points (rightmost).

The *hypervolume* with respect to a reference point r of an assessed set S of objective vectors, that we denote by $\mathrm{HV}_r(S)$, is the Lebesgue measure of the region of the objective-space dominated by S and strictly dominating the reference point r. When no vector of the Pareto front dominates the reference point r, the hypervolume with respect to r of any set of feasible points of the objective space is null. Since this particular case is not interesting, we only consider reference points dominated by at least one vector of the Pareto front from now on. We refer to such reference points as *valid*.

The region of the objective-space dominated by S and dominating r, see the righthand plot of Figure 1, is denoted by $\mathcal{D}^r_S$ and formally defined as:

$$\mathcal{D}^r_S = \{w \in \mathbb{R}^2 : \exists u \in S : u \preceq w \prec r\}. \tag{2.3}$$

The hypervolume of a set S relative to the reference point r equals $\lambda(\mathcal{D}_S^r)$ with $\lambda(.)$ being the Lebesgue measure. In this paper, we only consider two-dimensional objective spaces and refer to the Lebesgue measure of a set as its area. The set $\mathcal{D}_S^r$ is the union of the $\mathcal{D}_u^r$ for $u \in S$, $\mathcal{D}_u^r$ being the rectangle $[u_1, r_1] \times [u_2, r_2]$ when u dominates r and $\emptyset$ otherwise, see the righthand plot of Figure 1. Note that the $\mathcal{D}_u^r$ are not disjoints.

We use the hypervolume to characterize the convergence of a set S of objective vectors to the entire Pareto front. For a fixed valid reference point r, a set S is said to converge to the Pareto front when the hypervolume difference $\mathrm{HV}_r(\mathrm{PF}_f) - \mathrm{HV}_r(S)$ converges to 0. We define the *optimality gap* of S with respect to a valid reference point r as $\mathrm{HV}_r(\mathrm{PF}_f) - \mathrm{HV}_r(S)$.

Another quantity of interest is how much adding a vector to a set affects its hypervolume. The *hypervolume improvement* with respect to r of the vector u to the set S is $\mathrm{HVI}_r(u, S) = \mathrm{HV}_r(S \cup \{u\}) - \mathrm{HV}_r(S)$. We also use the term hypervolume improvement to refer to the hypervolume increase of an increasing sequence of sets. More precisely, the hypervolume improvement at iteration n of a monotone sequence $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ is $\mathrm{HV}_r(\mathcal{S}_{n+1}) - \mathrm{HV}_r(\mathcal{S}_n)$.

2.2. The ISOOMOO framework, its HV-ISOOMOO instanciation and the associated greedy set sequences. The Incremental SingleObjective Optimization for MultiObjective Optimization (ISOOMOO) framework builds incrementally a monotone[1] sequence $(\mathcal{I}_n)_{n\in\mathbb{N}^*}$ of sets of vectors of the objective space. The pseudo code of ISOOMOO is given in Algorithm 2.1, where the current value of $\mathcal{I}_n$ is denoted by $\mathcal{I}$. At each so-called meta-iteration, a generic singleobjective maximization algorithm SOOPTIMIZER (line 3 in Algorithm 2.1) is run on the criterion $X \in \Omega \subset \mathbb{R}^d \mapsto J(\mathcal{I}, X)$ and the resulting solution is added to $\mathcal{I}$ (line 4 in Algorithm 2.1). We use the term meta-iteration to separate between the (meta-)iterations of ISOOMOO and the iterations of SOOPTIMIZER. Since the set $\mathcal{I}$ is composed of the final objective incumbents of previous runs of SOOPTIMIZER and (ideally) provides an approximation of the Pareto front, we call it *final incumbents Pareto front approximation*.

The singleobjective optimization procedure may vary between meta-iterations. More precisely, the run of SOOPTIMIZER depends on data about precedent runs stored in D (line 4 in Algorithm 2.1). This allows to alternate between various singleobjective optimization algorithms with different features, but also to adapt the initialization. This could be done by storing in D an iteration index or the final search-space incumbents of SOOPTIMIZER runs.

Algorithm 2.1 Incremental SingleObjective Optimization for MultiObjective Optimization (ISOOMOO)

1: **while** not **stopping criterion do**
2: $\quad Y, d \leftarrow$ SOOPTIMIZER$(X \mapsto J(\mathcal{I}, X),\ D)$
3: $\quad \mathcal{I} \leftarrow \mathcal{I} \cup \{F(Y)\}$ # update of the approximation of the Pareto front
4: $\quad D \leftarrow D \cup \{d\}$ # update of the data collected
5: **end while**

In this paper, we study HV-ISOOMOO, an instanciation of ISOOMOO for which the criterion $J(\mathcal{I}, .)$ relates to the hypervolume improvement to $\mathcal{I}$. Formally, HV-ISOOMOO

[1] A sequence of set $\{A_n, n \geq 0\}$ is monotone if the following inclusions $A_0 \subset A_1, \ldots \subset A_n \subset \ldots$ hold.

is a class of algorithms derived from ISOOMOO for which the maximization of the criterion J is *compliant* with the maximization of the hypervolume improvement as defined below.

Assumption 2.1. (Compliance to hypervolume improvement maximization) The maximization of a criterion J as in ISOOMOO is *compliant* with the maximization of the hypervolume improvement if for any set $\mathcal{I}$ of objective vectors, maximizing $J(\mathcal{I}, .)$ is equivalent to maximizing the hypervolume improvement with respect to a valid reference point r to the set $\mathcal{I}$:

$$\operatorname{argmax}_{X \in \mathbb{R}^d} J(\mathcal{I}, X) = \operatorname{argmax}_{X \in \mathbb{R}^d} \mathrm{HVI}_r(F(X), \mathcal{I}) \ . \tag{2.4}$$

We simply refer to an algorithm built from the HV-ISOOMOO framework as *an HV-ISOOMOO algorithm*.

Definition 2.2 (HV-ISOOMOO). *We define an HV-ISOOMOO algorithm as an ISOOMOO algorithm as described in Algorithm 2.1 where the criterion J satisfies Assumption 2.1.*

At each meta-iteration n, an HV-ISOOMOO algorithm seeks a feasible vector maximizing the hypervolume improvement to the final incumbents Pareto front approximation $\mathcal{I}_n$. Ideally, when n goes to infinity, the non-dominated subset of $(\mathcal{I}_n)_{n \in \mathbb{N}^*}$ converges to the (entire) Pareto front, namely the non-dominated set of feasible objective vectors, which maximizes the hypervolume. In other words, HV-ISOOMOO algorithms try to approximate the Pareto front with a greedy approach.

Definition 2.3. *We define the convergence of HV-ISOOMOO as the convergence of $HV_r(\mathcal{I}_n)$ towards $HV_r(PF_f)$.*

The performance of a specific HV-ISOOMOO algorithm depends crucially on the choice of the criterion J. In this respect, the hypervolume improvement to $\mathcal{I}$ is not a good candidate for $J(\mathcal{I}, .)$. Indeed, it is constant equal to zero in the region dominated by $\mathcal{I}$, which makes it difficult to optimize. A criterion whose maximization is compliant with the maximization of the hypervolume improvement and designed to be easier to optimize has already been introduced in [17] under the name *uncrowded hypervolume improvement* (UHVI). In the non-zero region of the hypervolume improvement to a set $\mathcal{I}$, namely for $F(X)$ not dominated by $\mathcal{I}$, the UHVI is simply the hypervolume improvement. Otherwise, in the region where the hypervolume improvement is null, the UHVI is negative and equals minus the distance to the empirical non-dominated front of the set $\mathcal{I}$ relative to r. It is easy to see that the UHVI satisfies (2.4).

The choice of SOOPTIMIZER also plays a key role in the performance of an HV-ISOOMOO algorithm. In this paper, we analyze HV-ISOOMOO under the assumption of perfect singleobjective optimization formalized below.

Assumption 2.4 (Perfect Singleobjective Optimization). At every meta-iteration n, for any final incumbents Pareto front approximation $\mathcal{I}_n$, the run of SOOPTIMIZER (line 2 in Algorithm 2.1) returns $Y \in \operatorname{argmax}_{X \in \Omega} J(\mathcal{I}_n, F(X))$.

The assumption of perfect singleobjective optimization is reminiscent to the assumption of perfect line search which is common in the analysis of gradient based methods [11]. Under this assumption, all choices of criterions verifying Assumption 2.1 are equivalent. The convergence of HV-ISOOMOO under perfect singleobjective optimization is a necessary condition for the soundness of the approach. Additionally, we could obtain a lower

bound on the convergence-speed of a real instanciation of HV-ISOOMOO by combining such lower bounds on the convergence-speed of HV-ISOOMOO under Assumption 2.4 with existing lower bounds on the convergence speed of singleobjective optimization algorithms.

We introduce below the notions of greedy sequence and greedy set sequence, which are mathematical abstractions of HV-ISOOMOO under Assumption 2.4 of perfect singleobjective optimization.

DEFINITION 2.5 (Greedy sequence and greedy set sequence). *Given a valid reference point r, we define as greedy sequence, a sequence $\{v_n, n \geq 1\}$ satisfying*

$$v_1 \in \arg\max_{v \in F(\Omega)} HV_r(v) \text{ and} \tag{2.5}$$

$$v_{n+1} \in \arg\max_{v \in F(\Omega)} HV_r(\{v_1, \cdots, v_n, v\}) \text{ for all } n \geq 1 \ . \tag{2.6}$$

The greedy set sequence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ associated to the greedy sequence $\{v_n, n \geq 1\}$ is composed of the greedy sets $\mathcal{S}_n := \{v_k, k \leq n\}$.

There is a bijection between greedy sequences and greedy set sequences. The n-th element of the greedy sequence $\{v_n, n \geq 1\}$ associated to a greedy set sequence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ is simply the unique element of $\mathcal{S}_n \setminus \mathcal{S}_{n-1}$ if $n > 1$ and of $\mathcal{S}_1$ if $n = 1$.

The recurrence relation of the greedy sequence (2.6) is equivalent to v_{n+1} belonging to $\arg\max_{v \in F(\Omega)} \mathrm{HVI}_r(v, \mathcal{S}_n)$ for all $n \geq 1$. It is immediate to see that under Assumption 2.4, the final incumbents generated by HV-ISOOMOO constitute *a* greedy sequence while the final incumbents Pareto front approximations form the associated greedy set sequence $(\mathcal{I}_n)_{n \in \mathbb{N}^*}$. The indices n of both greedy and greedy set sequences iterations correspond to HV-ISOOMOO meta-iterations. In this paper, we derive convergence results for greedy set sequences, which transfer to HV-ISOOMOO under Assumption 2.4.

As we will see in Subsection 3.1, the problem of maximizing the hypervolume improvement to a fixed set can be rewritten as the maximum of a finite number of hypervolume maximization problems. Therefore, we can infer from [3, Theorem 1] that as soon as the Pareto front is lower semi-continuous, there exists a greedy sequence and the associated greedy set sequence.

PROPOSITION 2.6. *If the Pareto front is described by a lower semi-continuous function f, then there exists a greedy sequence $\{v_n, n \geq 1\}$ associated to any valid reference point r.*

Proof. If f is lower semi-continuous, then for any reference point r, the maximum of $\mathrm{HV}_r(.)$ exists, see [3, Theorem 1]. Therefore, there exists a vector verifying (2.5) and the problem of maximizing the maximum of a finite number of hypervolume functions defined in (3.4) admits a solution. Since the recurrence equation defining greedy sequences (2.6) is equivalent to (3.4) by Lemma 3.5, a lemma proven in the next section, we can build a sequence $\{v_n, n \geq 1\}$ verifying (2.5) and (2.6), namely a greedy sequence. ∎

Additionally, since the hypervolume indicator associated to a valid reference point is strictly Pareto-compliant (see [14]), this sequence is composed of vectors of the Pareto front.

PROPOSITION 2.7. *If the Pareto front is described by a lower semi-continuous function f, then any vector of a greedy sequence associated to a valid reference point r belongs*

to the Pareto front. Consequently, for such Pareto front and reference point and under Assumption 2.4 of perfect singleobjective optimization, all final incumbents Pareto front approximations $\mathcal{I}_n$ of HV-ISOOMOO algorithms relative to r are subsets of the Pareto front.

Proof. Since for any valid reference point r, $\mathrm{HV}_r(.)$ is strictly Pareto-compliant [14], its maximum always belongs to the set of non-dominated feasible objective vectors, that is the Pareto front. Thus, in particular, a vector v_1 verifying (2.5) belongs to the Pareto front. Additionally, by Lemma 3.5, a lemma proven in the next section, every solution of the update equation of greedy sequences (2.6) verifies (3.4). As a consequence, such vectors are solution of at least one hypervolume maximization problem, and thus also belong to the Pareto front. ∎

Yet, in general, there exists more than one greedy sequence, and thus greedy set sequence. For example, there are infinitely many greedy sequences associated to any affine Pareto front with a reference point dominating the nadir point. This statement relies on the fact that the unique maximizer of the hypervolume relative to a reference point r dominating the nadir point is the middle of the section of the Pareto front dominating r, see [2, Theorem 5]. As a consequence, the middle of the section of the Pareto front dominating r is the only candidate for v_1 but v_2 can be either at 1/4 or at 3/4 of this section. Similarly, v_3 has to be in the position where v_2 is not but v_4 can be at 1/8, 3/8, 5/8 or 7/8 of the section of the Pareto front dominating r. For any n, we can find an iteration m such that v_m can be placed at 2^n different points, whatever the $m-1$ first terms of the greedy sequence are.

2.3. Assumptions on the Pareto front and the objective functions. We present and discuss here the assumptions on the function f describing the Pareto front under which we derive convergence results. We typically assume that the function f is bilipschitz, convex or simultaneously bilipschitz and with a Hölder continuous derivative, namely $\mathcal{C}^{1,\alpha}$. Under any of these three assumptions, f is continuous. For the sake of conciseness, we transfer the properties of f to the Pareto front. For example, we call *convex Pareto front* a Pareto front described by a convex function. We recall that a function g is Hölder continuous with exponent α when there exists $H \geq 0$ such that $|g(x)-g(y)| \leq H \times |x-y|^\alpha$ for all x, y [12]. We note $[g]_\alpha$ the *minimum Hölder coefficient* of g with respect to the exponent α, that is $[g]_\alpha := \sup_{x \neq y} \frac{|g(x)-g(y)|}{|x-y|^\alpha}$. When needed, we detail the bilipschitz constants and say that a bilipschitz function f is ($L_{\min}$,$L_{\max}$)-bilipschitz if for all $x, y \in [x_{\min}, x_{\max}]$, we have $L_{\min} \times |x-y| \leq |f(x)-f(y)| \leq L_{\max} \times |x-y|$ where $L_{\max} \geq L_{\min} > 0$. We also consider *affine Pareto fronts*, the simplest kind of Pareto front. As they form a line in the biobjective case, they are usually refered to as linear Pareto fronts. They provide good examples to illustrate a point and help to understand the results we prove on the asymptotic convergence behavior.

We remind below sufficient conditions on the search-space and on the objective functions which guarantee that f is convex and bilipschitz.

PROPOSITION 2.8. *Given a biobjective minimization problem as in (2.1) whose Pareto front is described by a function f. 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.*

PROPOSITION 2.9. *Given a biobjective minimization problem as in (2.1) whose Pareto front is described by a function f. If the search space Ω and the objective functions F_1*

and F_2 are convex, then f is convex.

The proofs of both propositions can be found for instance in [15]. The conditions on F_1, F_2 and Ω are sufficient but non-necessary conditions. Indeed, adding small discontinuity in the objective functions far from the Pareto set makes them non-convex and non-bilipschitz without modifying the Pareto front.

Representing F_1 values on the absciss and F_2 values on the ordinate instead of the converse is an arbitrary choice. When f is a bijection, had we chosen to represent the F_2 values on the abscissa instead of on the ordinate, we would have had another representation of the Pareto front : $\{(y, f^{-1}(y)) : y \in [f(x_{\max}); f(x_{\min})]\}$. If so, the inverse function f^{-1} would have played the role of f. It is interesting to notice that the choice of the objective function represented on the horizontal axis does not impact whether the function characterizing the Pareto front is bilipschitz or convex. Indeed, f being bilipschitz is equivalent to both f and f^{-1} being lipschitz. Additionally, we can prove that given that the function f is decreasing, f being convex is equivalent to its inverse f^{-1} being convex. The proof of this property is straightforward. If f is convex, then for all $x, y \in \mathbb{R}$, $f\left(\frac{f^{-1}(x)+f^{-1}(y)}{2}\right)$ is smaller than $\frac{f(f^{-1}(x))+f(f^{-1}(y))}{2}$, that is $\frac{x+y}{2}$. Since f and therefore f^{-1} are decreasing, by composing by f^{-1} each side of the inequality, we obtain a characterization of the convexity of f^{-1}: for all $x, y \in \mathbb{R}$, $\frac{f^{-1}(x)+f^{-1}(y)}{2}$ is higher than $f^{-1}(\frac{x+y}{2})$.

3. Preliminary results. In this section, we present preliminary results which are crucial for the analysis of the convergence of HV-ISOOMOO. While they are here exposed as tools for convergence analysis, they are also interesting for their own sake.

3.1. Decomposition of the optimality gap using gap regions. We introduced $\mathcal{D}_S^r$ in (2.3) as the region of the objective space dominating r and weakly dominated by S. Its Lebesgue measure is $\mathrm{HV}_r(S)$. We now introduce *total gap regions*, whose Lebesgue measure are optimality gaps.

DEFINITION 3.1. *The* total gap region *of S with respect to a fixed valid reference point r, $\mathcal{G}_S^r$, is defined as the region of the objective-space which dominates r and is weakly dominated by PF_f but not by S, namely $\mathcal{D}_{PF_f}^r \setminus \mathcal{D}_S^r$.*

When S is a subset of the Pareto front dominating the reference point r, the total gap region has a particular shape which can be visualized in the rightmost plot of Figure 1. The total gap region $\mathcal{G}_S^r$ can be decomposed into the disjoint union of $|S|+1$ sets of the form $\mathcal{D}_{S'}^{r'}$ that are formally defined below.

DEFINITION 3.2 (Gap regions, gaps and associated reference points). *Let $S = \{v_1, ..., v_n\}$ be a set of n distinct vectors of the Pareto front dominating a valid reference point r. Let σ be the permutation ordering the v_i by increasing F_1 values: $v_{\sigma(1),1} < v_{\sigma(2),1} < \ldots < v_{\sigma(n),1}$.*

- *For all $i \in [\![1, n+1]\!]$, the i-th gap region of the set S, $\mathcal{G}_{S,i}^r$, is the set $\mathcal{D}_{PF_f}^{r_i}$ with the* associated reference points *r_i being*

$$\begin{aligned} &r_1 = (v_{\sigma(1),1}, r_2), r_{n+1} = (r_1, v_{\sigma(n),2}) \text{ and} \\ &r_i = (v_{\sigma(i),1}, v_{\sigma(i-1),2}) \text{ for all } i \in [\![2, n]\!] \ . \end{aligned} \tag{3.1}$$

- *We refer to $\mathcal{G}_{S,1}^r$ and $\mathcal{G}_{S,n+1}^r$ as the left and the right extreme gap region of S, respectively.*

The left and the right extreme gap regions are empty when the left and the right extreme vectors of the Pareto front belong to S. Non-extreme gap regions are never empty. The total gap region is the disjoint union of the gap regions: $\mathcal{G}^r_{\mathcal{S}_n} = \dot{\cup}_{i=1}^{n+1} \mathcal{G}^r_{\mathcal{S}_n,i}$. This decomposition of the total gap region, and thus of the optimality gap, is the cornerstone of the convergence analysis. The area of a gap region $\mathcal{G}^r_{\mathcal{S}_n,i}$ is the hypervolume of the Pareto front with respect to the reference point r^n_i. As a consequence, we can write the optimality gap as the sum of $n+1$ hypervolumes of the Pareto front with respect to different reference points.

LEMMA 3.3. *At any iteration n, the optimality gap of a greedy set sequence with respect to a valid reference point can be decomposed as the sum of $n+1$ hypervolumes of the Pareto front:*

$$HV_r(PF_f) - HV_r(\mathcal{S}_n) = \sum_{i=1}^{n+1} HV_{r^n_i}(PF_f) \ . \tag{3.2}$$

Proof. The optimality gap at iteration n is the Lebesgue measure of the total gap region $\mathcal{G}^r_{\mathcal{S}_n}$, which is the disjoint union of the gap regions $\mathcal{G}^r_{\mathcal{S}_n,i}$ of $\mathcal{S}_n$. Since each gap region $\mathcal{G}^r_{\mathcal{S}_n,i}$ equals $\mathcal{D}^{r^n_i}_{\mathrm{PF}_f}$, the optimality gap is equal to $\sum_{i=1}^{n+1} \mathrm{HV}_{r^n_i}(\mathrm{PF}_f)$. ∎

Additionally, we can express the hypervolume improvement of any vector to $\mathcal{S}_n$ as an hypervolume. It is immediate for vectors which do not dominate $\mathcal{S}_n$. For other vectors, the reference point depends on the gap region to which the vector belongs.

LEMMA 3.4. *Let $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ be a greedy set sequence relative to a valid reference point r. At any iteration n, the hypervolume improvement to $\mathcal{S}_n$ of any u belonging to the i-th gap region of $\mathcal{S}_n$, $\mathcal{G}^r_{\mathcal{S}_n,i}$, satisfies*

$$HVI_r(u, \mathcal{S}_n) = HV_{r^n_i}(u) \ . \tag{3.3}$$

Proof. The hypervolume improvement of any $u \in \mathcal{G}^r_{\mathcal{S}_n,i}$ is the Lebesgue-measure of the intersection between $\mathcal{G}^r_{\mathcal{S}_n,i}$ and $\mathcal{D}^r_u$. Therefore, it is equal to $\lambda(\mathcal{D}^{r^n_i}_u)$, that is $\mathrm{HV}_{r^n_i}(u)$. ∎

We can now reformulate the recurrence relation defining a greedy sequence at iteration $n+1$. Indeed, picking a vector maximizing the hypervolume improvement to $\mathcal{S}_n$ is equivalent to pick a vector where the highest value of the maximum of the hypervolumes with respect to the r^n_i is reached.

LEMMA 3.5. *At any iteration n, the recurrence formula satisfied by v_{n+1}, i.e. (2.6), can be reformulated as*

$$v_{n+1} \in \arg\max_{u\in PF_f} \max_{i\in[\![1,n+1]\!]} HV_{r^n_i}(u). \tag{3.4}$$

Proof. The hypervolume improvement of any vector u to $\mathcal{S}_n$ is $\max_{i\in[\![1,n+1]\!]} \mathrm{HV}_{r^n_i}(u)$. It is a consequence of Lemma 3.4 and of the fact that the hypervolume with respect to r^n_i is null outside the i-th gap region of $\mathcal{S}_n$. Additionally, v_{n+1} belongs to the Pareto front by Proposition 2.7. Thus, (2.6) is equivalent to (3.4). ∎

Similarly, we can express the decrease of the optimality gap at iteration $n+1$, $\mathrm{HV}_r(\mathcal{S}_{n+1}) - \mathrm{HV}_r(\mathcal{S}_n)$, as the maximum of $n+1$ hypervolume maximization problems.

LEMMA 3.6. *Let $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ be a greedy set sequence relative to a valid reference point r. The hypervolume improvement at iteration $n+1$ equals*

$$HV_r(\mathcal{S}_{n+1}) - HV_r(\mathcal{S}_n) = \max_{u\in PF_f} \max_{i\in[\![1,n+1]\!]} HV_{r_i^n}(u) \ . \tag{3.5}$$

Proof. The hypervolume improvement $\mathrm{HV}_r(\mathcal{S}_{n+1}) - \mathrm{HV}_r(\mathcal{S}_n)$ is the hypervolume improvement of v_{n+1} to $\mathcal{S}_n$. With the same arguments as in the proof of Lemma 3.5, we can prove that it equals $\max_{u\in\mathrm{PF}_f} \max_{i\in[\![1,n+1]\!]} \mathrm{HV}_{r_i^n}(u)$. ∎

3.2. Lower bound of the normalized maximum hypervolume for convex Pareto fronts. In this section and the next one, we provide bounds on the maximum hypervolume achievable by a single feasible vector normalized by the maximum hypervolume of a feasible set. We refer to this ratio as the *normalized maximum hypervolume* with respect to r:

$$\frac{\max_{u\in\mathrm{PF}_f} \mathrm{HV}_r(u)}{\mathrm{HV}_r(\mathrm{PF}_f)}. \tag{3.6}$$

Bounds on the normalized maximum hypervolume are exploited in Section 4 to provide bounds on the speed of convergence of the greedy set sequence towards the Pareto front.

The hypervolume relative to a reference point $r = (r_1, r_2)$ of a vector $u = (x, f(x))$ of the Pareto front is $\mathrm{HV}_r(u) = (r_1 - x) \times (r_2 - f(x))$. From this simple formula, we derive in the next proposition necessary conditions for a vector of the Pareto front $u^* = (x^*, f(x^*))$ to be an hypervolume maximizer when f has at least left and right derivatives in x^*.

PROPOSITION 3.7. *Let $x^* \in]x_{\min}, x_{\max}[$ such that $u^* := (x^*, f(x^*))$ maximizes the hypervolume with respect to a valid reference point $r = (r_1, r_2)$. If the function f describing the Pareto front admits left and right derivatives in x^*, respectively $f'_-(x^*)$ and $f'_+(x^*)$, then*

$$-f'_+(x^*) \leq \frac{r_2 - f(x^*)}{r_1 - x^*} \leq -f'_-(x^*) \ . \tag{3.7}$$

Proof. We define the function $\mathrm{HV}_{x,r}(.)$ as $x \mapsto \mathrm{HV}_r((x, f(x)))$. If x^* maximizes $\mathrm{HV}_{x,r}(.)$, then the left and the right derivatives of $\mathrm{HV}_{x,r}(.)$ are positive and negative, respectively. By replacing the left and right derivatives of $\mathrm{HV}_{x,r}(.)$ by their explicit formulas and reorganizing the terms we obtain (3.7). ∎

Equation (3.7) states that the slope of the diagonal of the rectangle $\mathcal{D}^r_{u^*}$ is between the absolute values of the slopes of the right and the left derivatives of f at x^*, see the middle plot of Figure 2. To the best of our knowledge, this geometric interpretation is new. It becomes simpler when f is differentiable. As soon as u^* is a non-extreme vector, the absolute value of the slope of the tangent of the front at u^* is equal to the slope of the diagonal of the rectangle $\mathcal{D}^r_{u^*}$, see the lefthand plot of Figure 2.

COROLLARY 3.8. *Let $x^* \in]x_{\min}, x_{\max}[$ be such that $u^* := (x^*, f(x^*))$ maximizes the hypervolume with respect to a valid reference point $r = (r_1, r_2)$. If the Pareto front is described by a differentiable function f in x^*, then $f'(x^*)$ satisfies*

$$-f'(x^*) = \frac{r_2 - f(x^*)}{r_1 - x^*} \ . \tag{3.8}$$

Proof. It is a direct consequence of Proposition 3.7 ∎

A convex function may not be differentiable, but it always has left and right derivatives. It is also above its left and right tangent lines respectively on the left and on the right of x^*. Therefore, (3.7) implies that the affine function $g : x \mapsto f(x^*) - \frac{r_2 - f(x^*)}{r_1 - x^*} \times (x - x^*)$ is a minorant of f, and thus that the hypervolume of PF_f is smaller than the hypervolume of $\mathrm{PF}_g := \{g(x) : x \in [x_{\min}, x_{\max}]\}$. This upper bound on $\mathrm{HV}_r(\mathrm{PF}_f)$ involves the lengths of the rectangle $\mathcal{D}^r_{u^*}$, whose area is $\mathrm{HV}_r(u^*)$. It is the key idea of the proof of the following lower bound on the normalized maximum hypervolume.

PROPOSITION 3.9. *If the Pareto front is described by a convex function f, then the following lower bound on the normalized maximum hypervolume with respect to any valid reference point r holds:*

$$\frac{\max_{u \in PF_f} HV_r(u)}{HV_r(PF_f)} \geq \frac{1}{2} \tag{3.9}$$

where the inequality is an equality if and only if the Pareto front is affine and r dominates the nadir point.

Proof. The function f being convex, it has left and right derivatives and thus, so does $\mathrm{HV}_{x,r}(.) : x \mapsto \mathrm{HV}_r((x, f(x)))$. Thus, by Proposition 3.7, (3.7) holds. Since f is convex, (3.7) implies that the affine function $g : x \mapsto f(x^*) - \frac{r_2 - f(x^*)}{r_1 - x^*} \times (x - x^*)$ is a minorant of f. Therefore, $\mathrm{PF}_g := \{g(x) : x \in [x_{\min}, x_{\max}]\}$ dominates PF_f, and thus has a higher hypervolume.

We note $L_1 := r_1 - \tilde{x}_{\min,r}$ and $L_2 := r_2 - f(\tilde{x}_{\max,r})$ the lengths of the rectangle $\mathcal{R} := [\tilde{x}_{\min,r}, r_1] \times [f(\tilde{x}_{\max,r}), r_2]$. We note $l_1 := r_1 - x^*$ and $l_2 := r_2 - f(x^*)$ the lengths of the rectangle $\mathcal{D}^r_{u^*}$. The region of $\mathcal{R}$ which dominates PF_g is a right-angled triangle. Additionally, by definition, the slope of its hypotenuse is $\frac{l_2}{l_1}$, and thus the lengths of the other sides are $L_1 - l_1 + (L_2 - l_2) \times \frac{l_1}{l_2}$ and $L_2 - l_2 + (L_1 - l_1) \times \frac{l_2}{l_1}$ (see the middle plot of Figure 2). Therefore

$$\begin{aligned}
\mathrm{HV}_r(\mathrm{PF}_g) &= \lambda(\mathcal{R}) - \lambda(\{u \in \mathbb{R}^2 : u \in \mathcal{R}, u \preceq \mathrm{PF}_g\}) \\
&= L_1 L_2 - \frac{1}{2} \times (L_1 - l_1 + (L_2 - l_2) \times \frac{l_1}{l_2}) \times (L_2 - l_2 + (L_1 - l_1) \times \frac{l_2}{l_1}) \\
&= l_1 l_2 \times \Big[-2 + 2 \times \frac{L_2}{l_2} - \frac{1}{2} \times \Big(\frac{L_2}{l_2}\Big)^2 + 2 \times \frac{L_1}{l_1} - \frac{1}{2} \times \Big(\frac{L_1}{l_1}\Big)^2 \Big] \ .
\end{aligned}$$

For all x, we have $(x-2)^2 \geq 0$ and thus $2x - \frac{1}{2}x^2 \leq 2$. Therefore, we can conclude that $\mathrm{HV}_r(\mathrm{PF}_g)$, and thus $\mathrm{HV}_r(\mathrm{PF}_f)$ is lower than $2 \times l_1 l_2$, that is $2 \times \mathrm{HV}_r(u^*)$. If either $L_1/l_1 \neq 2$ or $L_2/l_2 \neq 2$, the inequality is strict. Thus, when the inequality is an equality, the center of $\mathcal{R}$ belongs to the Pareto front. Since f is convex, it requires from f to be affine and from the reference point r to dominate the nadir point $(x_{\max}, f(x_{\min}))$. Conversely, if the Pareto front is affine and the reference point r dominates the nadir point, we know that the optimum is in the middle of the Pareto front and that we have the equality (see [2, Theorem 5]). ∎

We just proved that one half is a tight lower bound on the normalized maximum hypervolume for convex Pareto fronts. However, except for the trivial upper bound 1, there is no upper bound valid for every convex Pareto front, even when r dominates the

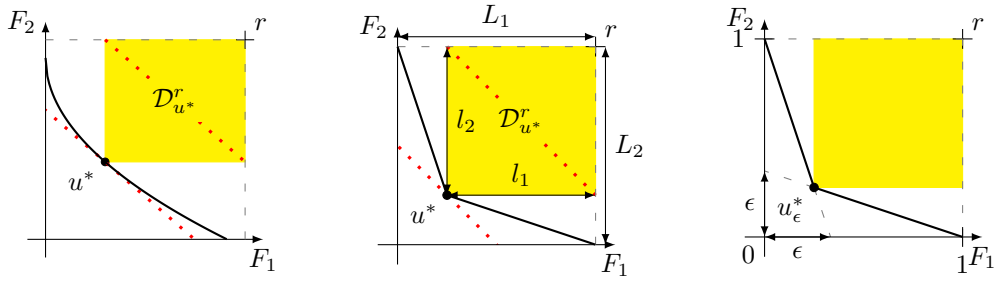

Fig. 2: Left and middle: Two convex Pareto fronts and their respective hypervolume maximizers u^*, one differentiable (left) and one non-differentiable (middle). The slopes of the two dotted lines, namely PF_g and the diagonal of $\mathcal{D}^r_{u^*}$, are equal. Right: The Pareto front PF_ϵ and the hypervolume maximizer u^*_ϵ for $\epsilon = 1/4$ and $r = (1, 1)$.

nadir point. Here is a simple example which illustrates this. Let consider the convex Pareto front $\mathrm{PF}_\epsilon := \{\max(1 - \frac{x}{\epsilon}, \epsilon - \epsilon \times x) : x \in [0, 1]\}$ represented in the righthand plot of Figure 2 and the reference point $r = (1, 1)$. When $\epsilon \leq 1$, PF_ϵ is convex and (3.8) implies that $u^*_\epsilon = (\epsilon \times (1 - \epsilon), \epsilon \times (1 - \epsilon))$ is the unique hypervolume minimizer. Thus, the normalized maximum hypervolume of PF_ϵ for this reference point is equal to $\frac{(1-\epsilon+\epsilon^2)^2}{1-\epsilon\times(1-\epsilon)^2+(\epsilon-\epsilon^2)^2}$ and converges to 1 when ϵ goes to 0.

3.3. Lower and upper bounds of the normalized maximum hypervolume for bilipschitz Pareto fronts. In this section, we examine lower and upper bounds on the normalized maximum hypervolume in the case of bilipschitz Pareto fronts.

We consider two affine fronts with the same left extreme vector as PF_f and slopes $-L_{\min}$ and $-L_{\max}$, see Figure 3. We call them $\mathrm{PF}_{\min}$ and $\mathrm{PF}_{\max}$, respectively. Formally:

$$\mathrm{PF}_{\max} := \{(x, f_{\max}(x)) : x \in [x_{\min}, x_{\max}]\}, \tag{3.10}$$

$$\mathrm{PF}_{\min} := \{(x, f_{\min}(x) : x \in [x_{\min}, x_{\max}]\} \tag{3.11}$$

with $f_{\min}(x) = f(x_{\min}) - (x - x_{\min}) \times L_{\min}$ and $f_{\max}(x) = f(x_{\min}) - (x - x_{\min}) \times L_{\max}$. For a ($L_{\min}$,$L_{\max}$)-bilipschitz function f, $f_{\min}(x) \leq f(x) \leq f_{\max}(x)$ for $x \in [x_{\min}, x_{\max}]$, and thus the Pareto front is dominated by $\mathrm{PF}_{\max}$ and dominates $\mathrm{PF}_{\min}$. These two affine fronts provide bounds on both the hypervolume of the Pareto front and the highest hypervolume of a vector on the Pareto front. They are key to prove the following lower bound on the normalized maximum hypervolume of a ($L_{\min}$,$L_{\max}$)-bilipschitz Pareto front.

PROPOSITION 3.10. *If the Pareto front is described by a ($L_{\min}$,$L_{\max}$)-bilipschitz function f, then the following bound on the normalized maximum hypervolume with respect to any valid reference point r holds*

$$\frac{\max_{u \in PF_f} HV_r(u)}{HV_r(PF_f)} \geq \frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \ . \tag{3.12}$$

Proof. The fronts $\mathrm{PF}_{\max}$ and $\mathrm{PF}_{\min}$ are defined respectively in (3.10) and (3.11). We note $\Delta_1 := \tilde{x}_{\max,r} - \tilde{x}_{\min,r}$, $\Delta'_1 := r_1 - \tilde{x}_{\max,r}$, $\Delta_2 := r_2 - f(\tilde{x}_{\min,r})$ and $V :=$

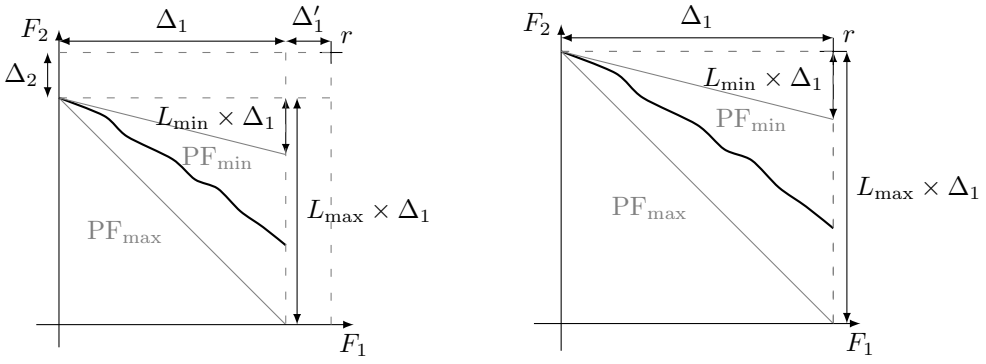

Fig. 3: The Pareto front PF_f surrounded by $\mathrm{PF}_{\max}$ and $\mathrm{PF}_{\min}$ in the case where the reference point r does not dominate the nadir point (left) and in the case where it does (right).

$\Delta_2 \times (r_1 - \tilde{x}_{\min,r})$, see the lefthand plot of Figure 3. Since the front $\mathrm{PF}_{\max}$ dominates the Pareto front, the hypervolume of PF_f is smaller than the hypervolume of $\mathrm{PF}_{\max}$, $V + L_{\max} \times \Delta_1 \times \Delta_1' + \frac{1}{2} \times L_{\max} \times \Delta_1^2$. Since each vector of $\mathrm{PF}_{\min}$ is dominated by a vector of PF_f, the maximum hypervolume of a vector of PF_f is higher than the maximum hypervolume of a vector of $\mathrm{PF}_{\min}$. The front $\mathrm{PF}_{\min}$ being an affine and therefore convex front, we know by Proposition 3.9 that the maximum hypervolume of a vector of $\mathrm{PF}_{\min}$ is higher than half of $\mathrm{HV}_r(\mathrm{PF}_{\min})$, which is equal to $\frac{1}{2} \times (V + L_{\min} \times \Delta_1 \times \Delta_1' + \frac{1}{2} \times L_{\min} \times \Delta_1^2)$. To summarize, the maximum hypervolume of a vector of PF_f is higher than $\frac{1}{2} \times (V + L_{\min} \times \Delta_1 \times \Delta_1' + \frac{1}{2} \times L_{\min} \times \Delta_1^2)$. Combining the upper bound on the hypervolume of PF_f and the lower bound on the maximum hypervolume of a vector of PF_f, the normalized maximum hypervolume is higher than $\frac{\frac{1}{2} \times (V + L_{\min} \times \Delta_1 \times \Delta_1' + \frac{1}{2} \times L_{\min} \times \Delta_1^2)}{V + L_{\max} \times \Delta_1 \times \Delta_1' + \frac{1}{2} \times L_{\max} \times \Delta_1^2}$. This quantity is itself larger than $\frac{1}{2} \times \frac{L_{\min} \times \Delta_1 \times \Delta_1' + \frac{1}{2} \times L_{\min} \times \Delta_1^2}{L_{\max} \times \Delta_1 \times \Delta_1' + \frac{1}{2} \times L_{\max} \times \Delta_1^2}$. As $V \geq 0$ and $0 < \frac{L_{\min} \times \Delta_1 \times \Delta_1' + \frac{1}{2} \times L_{\min} \times \Delta_1^2}{L_{\max} \times \Delta_1 \times \Delta_1' + \frac{1}{2} \times L_{\max} \times \Delta_1^2} < 1$, we conclude that the normalized maximum hypervolume is higher than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}}$. ∎

We cannot guarantee any upper bound strictly inferior to 1 on the normalized maximum hypervolume without adding an assumption on the reference point. Indeed, for a given bounded Pareto front, it is easy to show that the normalized maximum hypervolume goes to 1 when $r \to \infty$. However, if f is $(L_{\min}, L_{\max})$-bilipschitz and r dominates the nadir point, we can prove that the normalized maximum hypervolume is higher than $\frac{1}{2} \times \frac{L_{\max}}{L_{\min}}$. The proof relies on the fact that if the reference point dominates the nadir point, the vector of an affine front with the highest hypervolume is its middle (see [2, Theorem 5]), whose hypervolume is half of the hypervolume of the entire front.

PROPOSITION 3.11. *If the Pareto front is described by a $(L_{\min}, L_{\max})$-bilipschitz function f and the reference point r is valid and dominates the nadir point, the following upper-bound on the normalized maximum hypervolume with respect to r holds*

$$\frac{\max_{u \in PF_f} HV_r(u)}{HV_r(PF_f)} \leq \frac{1}{2} \times \frac{L_{\max}}{L_{\min}} \quad . \tag{3.13}$$

Proof. We use the same notations as in the proof of Proposition 3.10. Since r dominates the nadir point, both Δ_1', Δ_2 and V equal 0, and thus the hypervolumes of $\mathrm{PF}_{\max}$ and $\mathrm{PF}_{\min}$ equal $\frac{1}{2} \times L_{\max} \times \Delta_1^2$ and $\frac{1}{2} \times L_{\min} \times \Delta_1^2$, respectively. The domination of $\mathrm{PF}_{\min}$ by PF_f implies that the hypervolume of the Pareto front is below $\frac{1}{2} \times L_{\min} \times \Delta_1^2$. Since $\mathrm{PF}_{\max}$ is an affine front whose extremes dominate r, its middle is the unique hypervolume maximizer (see [1, Theorem 5]) with an hypervolume equal to $\frac{1}{4} \times L_{\max} \times \Delta_1$.The domination of PF_f by $\mathrm{PF}_{\max}$ implies that the maximum hypervolume of a vector of PF_f is higher than $\frac{1}{4} \times L_{\max} \times \Delta_1^2$. Gathering the lower bound on $\mathrm{HV}_r(\mathrm{PF}_f)$ and the upper bound on the maximum hypervolume of a vector of PF_f, we retrieve (3.13). ∎

This upper bound is only relevant for $L_{\max}/L_{\min} < 2$. The bound is the tightest for $L_{\max} = L_{\min}$, where it achieves the value $1/2$. In this paper, we use this upper bound for $L_{\max}/L_{\min}$ close to 1 to analyze the asymptotic convergence behavior of HV-ISOOMOO.

4. Convergence of HV-ISOOMOO under perfect singleobjective optimization. We prove in this section various convergence results for HV-ISOOMOO algorithms under the assumption of perfect singleobjective optimization. We first prove that when the Pareto front is either convex or bilipschitz, these algorithms converge to the entire Pareto front with guaranteed convergence-speed. We transform the bounds on the normalized maximum hypervolume proven in Section 3 into lower bounds on the convergence speed. Second, we analyze the asymptotic convergence behavior when the Pareto front is bilipschitz with a Hölder continuous derivatives.

To analyze the decrease of the optimality gap with respect to n, we need to be able to track in which gap regions the vectors of the greedy sequence are inserted over multiple iterations. Naturally, a gap region of $\mathcal{S}_n$ persists in being a gap region in the following iterations, as long as no greedy vector is added in this specific gap region. The greedy vector v_{n+1} is said to fill the gap region of $\mathcal{S}_n$ to which it belongs. At iteration $n+1$, this gap region disappears, replaced by two gap regions that we call its *children*. More generally, we say that a gap region is a *descendant* of another gap region when it is a proper subset of this gap region.

4.1. Convergence of HV-ISOOMOO with guaranteed convergence speed. We prove some upper bounds on the relation between the optimality gap at iteration $2n+1$ and at iteration n. These bounds translate into lower bounds on the speed of convergence of HV-ISOOMOO with perfect singleobjective optimization. The proof relies on inequalities of the form

$$\max_{u \in \mathrm{PF}_f} \mathrm{HV}_{r'}(u) \geq C \times \mathrm{HV}_{r'}(\mathrm{PF}_f) \tag{4.1}$$

stated in Propositions 3.9 and 3.10 and on the reformulation of optimality gaps, areas of gap regions and hypervolume improvement done in Subsection 3.1. A consequence of (4.1) being true for any reference point r' is that the optimality gap at iteration $2n+1$ is at most $(1-C)$ times the optimality gap at iteration n.

We sketch the proof idea in the simple case where each of the v_k ($k \in [\![n+1, 2n+1]\!]$) is inserted in a distinct gap region of $\mathcal{S}_n$, see the lefthand plot of Figure 4. Inserting v_k in a gap region leads to an hypervolume improvement larger than C times the area of this gap region by (4.1). Thus, the hypervolume improvement from iteration n to $2n+1$ is larger than C times the area of the union of all gap regions of $\mathcal{S}_n$, namely the optimality gap at iteration n. A detailed proof is presented after the theorem statement.

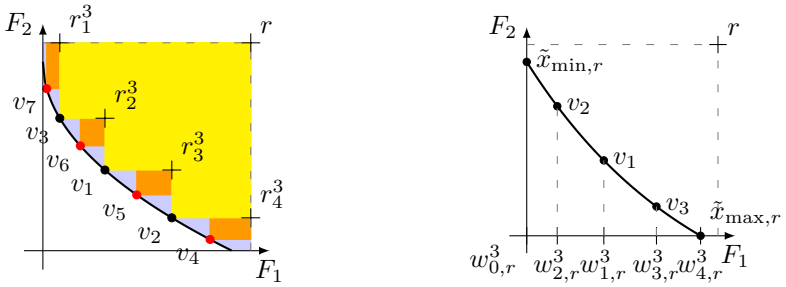

Fig. 4: Left: A Pareto front where each of the gap regions of $\mathcal{S}_3$ is filled by one of the greedy vectors v_k for $k \in [\![4,7]\!]$. It is described by $f(x) = 1 - \sqrt{x}$ for $x \in [0,1]$. We represent the region $\mathcal{D}^r_{\mathcal{S}_3}$, the regions corresponding to $\mathrm{HVI}_r(v_k, \mathcal{S}_{k-1})$ for $k \in [\![4,7]\!]$ and the gap regions of $\mathcal{S}_3$. Right: The ordered greedy set along F_1-values $w^n_{i,r}$ corresponding to the greedy set $\mathcal{S}_3$. The Pareto front is described by $f(x) = \frac{e}{e-1} \times e^{-x} + 1 - \frac{e}{e-1}$ for $x \in [0,1]$.

PROPOSITION 4.1. *Consider a biobjective optimization problem where the Pareto front is described by a function f. Any greedy set sequence $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ associated to a valid reference point r satisfies for all n*

$$\frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n+1})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \leq 1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \quad \textit{if } f \textit{ is } (L_{\min}, L_{\max})\textit{-bilipschitz and} \tag{4.2}$$

$$\frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n+1})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \leq \frac{1}{2} \quad \textit{if } f \textit{ is convex.} \tag{4.3}$$

Proof. Fix $n \geq 1$. We note σ a permutation of $[\![1, n+1]\!]$ such that $n + \sigma(i)$ is the index of the first greedy vector v_k inserted in $\mathcal{G}^r_{\mathcal{S}_n,i}$ when possible. With this choice of σ, the i-th gap region of $\mathcal{S}_n$ is a gap region of $\mathcal{S}_{n+\sigma(i)-1}$. As a consequence, the hypervolume improvement to $\mathcal{S}_{n+\sigma(i)-1}$ of any vector u belonging to the i-th gap region of $\mathcal{S}_n$ is equal to $\mathrm{HV}_{r_i^n}(u)$ by (3.3). The hypervolume improvement of the greedy vector $v_{n+\sigma(i)}$ to $\mathcal{S}_{n+\sigma(i)-1}$ being maximal, it is in particular larger than the one of any vector of $\mathcal{G}^r_{\mathcal{S}_n,i}$, the i-th gap region of $\mathcal{S}_n$, and thus than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times \mathrm{HV}_{r_i^n}(\mathrm{PF}_f)$ by Proposition 3.10. In other words, the hypervolume improvement at any iteration $n + \sigma(i)$ is higher than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times \mathrm{HV}_{r_i^n}(\mathrm{PF}_f)$. By adding these inequations for all $i \in [\![1, n+1]\!]$, we deduce that the hypervolume improvement from iteration n to $2n+1$ is larger than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times \sum_{i=1}^{n+1} \mathrm{HV}_{r_i^n}(\mathrm{PF}_f)$. Since the sum of the $\mathrm{HV}_{r_i^n}(\mathrm{PF}_f)$ is the optimality gap at iteration n, we have (4.2). If f is convex instead of bilipschitz, we use Proposition 3.9 instead of Proposition 3.10 and obtain (4.3). ∎

Since the optimality gaps form a decreasing sequence, such lower bounds on the relation between the optimality gaps at iteration $2n+1$ and at iteration n imply that the optimality gap associated to a greedy set sequence converges asymptotically to 0. Equivalently, HV-ISOOMOO algorithms converge to the entire Pareto front under Assumption 2.4, as stated formally below.

THEOREM 4.2. *Consider a biobjective optimization problem with a Pareto front described by a bilipschitz or convex function f.*

The hypervolume of a greedy set sequence associated to a valid reference point r converges to the hypervolume of the entire Pareto front, i.e. $HV_r(\mathcal{S}_n) \xrightarrow[n\to\infty]{} HV_r(PF_f)$.

Equivalently, for such Pareto fronts and under Assumption 2.4 of perfect singleobjective optimization, HV-ISOOMOO algorithms associated to a valid reference point r converge to the Pareto front in the sense of Definition 2.3.

From the lower bounds on the relation between the optimality gaps at iteration $2n+1$ and at iteration n, we deduce the following upper bounds on the normalized optimality gap at any iteration.

COROLLARY 4.3. *Consider a biobjective optimization problem with a Pareto front described by a ($L_{\min}$,$L_{\max}$)-bilipschitz function. A greedy set sequence $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ associated to a valid reference point r satisfies for all n*

$$\frac{HV_r(PF_f) - HV_r(\mathcal{S}_n)}{HV_r(PF_f)} \leq \left(1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}}\right)^{\lfloor \log_2(n+1) \rfloor} \leq (2n+2)^{\log_2(1-\frac{1}{2}\times\frac{L_{\min}}{L_{\max}})} \quad . \tag{4.4}$$

If the function f is convex, then any greedy set sequence associated to a valid reference point r satisfies for all n

$$\frac{HV_r(PF_f) - HV_r(\mathcal{S}_n)}{HV_r(PF_f)} \leq \left(\frac{1}{2}\right)^{\lfloor \log_2(n+1) \rfloor} \leq \frac{1}{2n+2} \quad . \tag{4.5}$$

Consequently, under Assumption 2.4 of perfect singleobjective optimization, HV-ISOOMOO algorithms relative to a valid reference point r satisfy (4.4) if f is ($L_{\min}$,$L_{\max}$)-bilipschitz and (4.5) if f is convex where $\mathcal{S}_n$ is replaced by $\mathcal{I}_n$, the final incumbents Pareto front approximation at iteration n.

Proof. The k-th term of the sequence defined by $u_0 = 1$ and $u_{n+1} = 2 \times u_n + 1$ for all $n \geq 1$ is $2^k - 1$. Thus, (4.2) and (4.3) imply that when f is ($L_{\min}$,$L_{\max}$)-bilipschitz or convex, the normalized optimality gap at iteration $2^k - 1$ is inferior to $(1-C)^k$ with C equal to $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}}$ and $\frac{1}{2}$, respectively. Since the hypervolume of the greedy set increases with n, and thus the optimality gap decreases with n, we deduce the first inequalities in (4.4) and (4.5) via the change of variable $k = \lfloor \log_2(n+1) \rfloor$.

Additionally, for every n, $\lfloor \log_2(n+1) \rfloor$ is smaller than $\log_2(n+1)+1$, that is $\log_2(2n+2)$. For every C, $\log_2(2n+2)$ equals $\log_C(2n+2) \times \log_2(C)$, and thus $C^{\log_2(2n+2)}$ equals $(2n+2)^{\log_2(C)}$. Therefore, we can infer that $(2n+2)^{\log_2(C)}$ is an upper bound of the normalized optimality gap with $C := 1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}}$ and $C := \frac{1}{2}$ when f is ($L_{\min}$,$L_{\max}$)-bilipschitz and convex, respectively. ∎

We focus here on the relation between the optimality gap at iteration n and at iteration $2n+1$. We could similarly examine the relation between the optimality gap at iteration n and at any later iteration. For example, we could prove that if f is ($L_{\min}$,$L_{\max}$)-bilipschitz, then for all n, for all $k \leq n+1$, $\frac{\mathrm{HV}_r(\mathrm{PF}_f) - \mathrm{HV}_r(\mathcal{S}_{n+k})}{\mathrm{HV}_r(\mathrm{PF}_f) - \mathrm{HV}_r(\mathcal{S}_n)}$ is lower than $1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times \frac{k}{n+1}$.

Sketch of proof. We consider the k gap regions of $\mathcal{S}_n$ with the highest areas. The hypervolume improvement from iteration n to $n+k$ is at least $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}}$ times the area of the union of these gap regions, which is at least $\frac{k}{n+1}$ times the optimality gap at iteration n.

4.2. Asymptotical behavior of the convergence of $\mathbf{HV}_r(\mathcal{S}_n)$ to $\mathbf{HV}_r(\mathbf{PF}_f)$. In this section, we analyze the asymptotic convergence behavior for a Pareto front described by a bilipschitz function with a Hölder continuous derivative. We prove that, in this case, doubling the number of vectors in the greedy set divides the optimality gap by a factor which converges asymptotically to two as stated in Theorem 4.10. This asymptotic limit corresponds to the case of affine Pareto fronts with a reference point dominating the nadir point. Indeed, for such Pareto fronts and reference points, the optimality gap is always halved when the number of vectors in the greedy set goes from n to $2n+1$, see Figure 5.

First, we study the properties of the part of the Pareto front corresponding to a specific gap region of $\mathcal{S}_n$. For all n, let note σ_n the permutation of $[\![1,n]\!]$ which orders the vectors of $\mathcal{S}_n$ by increasing F_1-values and the so-called *ordered greedy set along F_1-values*:

$$w^n_{i,r} := v_{\sigma_n(i),1} \text{ for } i \in [\![1,n]\!] \tag{4.6}$$

$$w^n_{0,r} := \tilde{x}_{\min,r} \text{ and } w^n_{n+1,r} := \tilde{x}_{\max,r} \ . \tag{4.7}$$

Naturally, we have $w^n_{0,r} \leq w^n_{1,r} \leq \ldots \leq w^n_{n+1,r}$, and the intervals $[w^n_{i-1,r}, w^n_{i,r}[$ for $i \in [\![1,n+1]\!]$ form a partition of $[\tilde{x}_{\min,r}, \tilde{x}_{\max,r}[$, see the righthand plot of Figure 4. The interval $[w^n_{i-1,r}, w^n_{i,r}]$ corresponds to the part of the Pareto front dominating r^n_i, the reference point associated to the i-th gap region of $\mathcal{S}_n$. When the Pareto front is bilipschitz, the lengths of these intervals converge asymptotically to 0 as stated in the next lemma. It is a direct consequence of the convergence of $\mathrm{HV}_r(\mathcal{S}_n)$ to $\mathrm{HV}_r(\mathrm{PF}_f)$ stated in Theorem 4.2.

LEMMA 4.4. *If the Pareto front is described by a bilipschitz function f and the greedy set sequence is associated to a valid reference point r, then the ordered greedy set along F_1-values satisfy* $\max_{i\in[\![1,n+1]\!]} w^n_{i,r} - w^n_{i-1,r} \xrightarrow[n\to\infty]{} 0$ *with the $w^n_{i,r}$ defined in (4.7).*

Proof. Let $L_{\min}$ and $L_{\max}$ be constants such that f is ($L_{\min}$,$L_{\max}$)-bilipschitz. The area of the i-th gap region of $\mathcal{S}_n$ is $\int_{w^n_{i-1,r}}^{w^n_{i,r}} (f(x) - f(w^n_{i,r}))dx$. This is larger than $\int_{w^n_{i-1,r}}^{w^n_{i,r}} L_{\min} \times (w^n_{i,r} - x)dx$, which equals $\frac{1}{2} \times L_{\min} \times (w^n_{i,r} - w^n_{i-1,r})^2$. Since the area of any gap region of $\mathcal{S}_n$ is inferior to the optimality gap at iteration n, this implies that the difference $w^n_{i,r} - w^n_{i-1,r}$ is inferior to $\sqrt{2 \times (\mathrm{HV}_r(\mathrm{PF}_f) - \mathrm{HV}_r(\mathcal{S}_n))}$ for all n, for all $i \in [\![1,n+1]\!]$. Therefore, the convergence of $\mathrm{HV}_r(\mathcal{S}_n)$ to $\mathrm{HV}_r(\mathrm{PF}_f)$ stated in Theorem 4.2 implies that the maximum over i of $w^n_{i,r} - w^n_{i-1,r}$ converges to 0. ∎

We prove in the next lemma that if the Pareto front is described by a bilipschitz function with a Hölder continuous derivative, then the restriction of this function associated to a gap region of $\mathcal{S}_n$ is bilipschitz for some constants whose ratio converges asymptotically to 1. Linear functions being the only functions to be ($L_{\min}$,$L_{\max}$)-bilipschitz with $L_{\min}/L_{\max} = 1$, it supports the interpretation that for such Pareto fronts, the convergence of the greedy set sequence is asymptotically similar as if they were affine Pareto fronts.

When the function describing the Pareto front is bilipschitz, its restriction to the part

of the Pareto front dominating r_i^n, that is $[w_{i-1,r}^n, w_{i,r}^n]$, is $(L_{\min}^{i,n}, L_{\max}^{i,n})$-bilipschitz with

$$
\begin{aligned}
L_{\min}^{i,n} &:= \inf\left\{\left|\frac{f(x)-f(y)}{x-y}\right|, x, y \in [w_{i-1,r}^n, w_{i,r}^n], x \neq y\right\} \text{ and} \\
L_{\max}^{i,n} &:= \sup\left\{\left|\frac{f(x)-f(y)}{x-y}\right|, x, y \in [w_{i-1,r}^n, w_{i,r}^n], x \neq y\right\}.
\end{aligned} \tag{4.8}
$$

At iteration n, the ratio between $L_{\max}^{i,n}$ and $L_{\min}^{i,n}$, the bilipschitz constants on the i-th gap region of $\mathcal{S}_n$, is by definition smaller than

$$
q_n := \max\left\{\frac{L_{\max}^{i,n}}{L_{\min}^{i,n}}, i \in [\![1, n+1]\!] : [w_{i-1,r}^n, w_{i,r}^n] \neq \emptyset\right\} \quad . \tag{4.9}
$$

The proof of the convergence of q_n to 1 relies on the fact that a derivable function can be approximated locally by an affine function. The quality of this approximation is guaranteed by the Hölder continuity of the derivative.

LEMMA 4.5. *We consider a greedy set sequence $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ relative to a valid reference point r. If the Pareto front is described by a bilipschitz function with a Hölder continuous derivative, then q_n defined in (4.9) converges asymptotically to 1.*

Proof. We take α such that f' is Hölder continuous with exponent α, i.e f is $\mathcal{C}^{1,\alpha}$, and $L_{\min}, L_{\max} > 0$ such that the function f describing the Pareto front is $(L_{\min}, L_{\max})$-bilipschitz. We recall that f is decreasing, and thus for all $x < y$, we have $f(x) - f(y) \geq 0$. Since f is $\mathcal{C}^{1,\alpha}$ and therefore $\mathcal{C}^1$, the Taylor formula with Lagrange remainder states that for all $x < y$, there exists $\xi \in [x, y]$ such that $f(y) = f(x) + (y-x) \times f'(\xi)$. Since f is $\mathcal{C}^{1,\alpha}$, this implies that for all $x < y$, $|f(y) - f(x) - (y-x) \times f'(x)| \leq (y-x)^{1+\alpha} \times [f']_{\mathcal{C}^\alpha}$. Thus, $\frac{f(y)-f(x)}{x-y}$ is lower than $-f'(x) + [f']_{\mathcal{C}^\alpha} \times (y-x)^\alpha$. We now restrict ourselves to x and y belonging to the non-empty interval $[w_{i-1,r}^n, w_{i,r}^n]$. Our goal is to find an upper bound depending on i but not on either x or y. Since f is $\mathcal{C}^{1,\alpha}$, the difference between $-f'(x)$ and $-f'(w_{i-1,r}^n)$ is lower than $[f']_{\mathcal{C}^\alpha} \times (x - w_{i-1,1,r}^n)^\alpha$, and thus $[f']_{\mathcal{C}^\alpha} \times (w_{i,r}^n - w_{i-1,r}^n)^\alpha$. Additionally, the difference between x and y is lower than $w_{i,r}^n - w_{i-1,r}^n$. We conclude that for $x, y \in [w_{i-1,r}^n, w_{i,r}^n]$, $\frac{f(y)-f(x)}{x-y}$ is lower than $-f'(w_{i-1,1,r}^n) + 2[f']_{\mathcal{C}^\alpha} \times (w_{i,r}^n - w_{i-1,r}^n)^\alpha$, and thus so is $L_{\max}^{i,n}$ defined in (4.8).

Following the same approach, we can also infer that $L_{\max}^{i,n}$ defined in (4.8) is greater than the symmetric quantity $-f'(w_{i-1,1,r}^n) - 2[f']_{\mathcal{C}^\alpha} \times (w_{i,r}^n - w_{i-1,r}^n)^\alpha$. The quantity $-f'(w_{i-1,1,r}^n)$ is greater than $L_{\min}$ and $(w_{i,r}^n - w_{i-1,r}^n)^\alpha$ is smaller than $\max_{i\in[\![1,n+1]\!]}(w_{i,r}^n - w_{i-1,r}^n)^\alpha$. As a consequence, q_n is lower than $\frac{L_{\min} + 2[f']_{\mathcal{C}^\alpha} \times \max_{i\in[\![1,n+1]\!]}(w_{i,r}^n - w_{i-1,r}^n)^\alpha}{L_{\min} - 2[f']_{\mathcal{C}^\alpha} \times \max_{i\in[\![1,n+1]\!]}(w_{i,r}^n - w_{i-1,r}^n)^\alpha}$. By Lemma 4.4, $\max_{i\in[\![1,n+1]\!]} w_{i,r}^n - w_{i-1,r}^n$ converges to 0 and thus, this upper bound on q_n converges to 1. By definition, q_n is always higher than 1, and thus converges to 1. ∎

A consequence of the previous lemma is that the bounds on the normalized hypervolume improvement of v_{n+1} to $\mathcal{S}_n$ that we can infer from Propositions 3.10 and 3.11 converge asymptotically to 1/2, see (4.10). Similarly, the bounds on the normalized area of the child of a gap region that we can infer from Lemma A.2 converge to 1/4, see (4.11). These asymptotic values correspond to the case of an affine Pareto front with a reference point dominating the nadir point, see Figure 5.

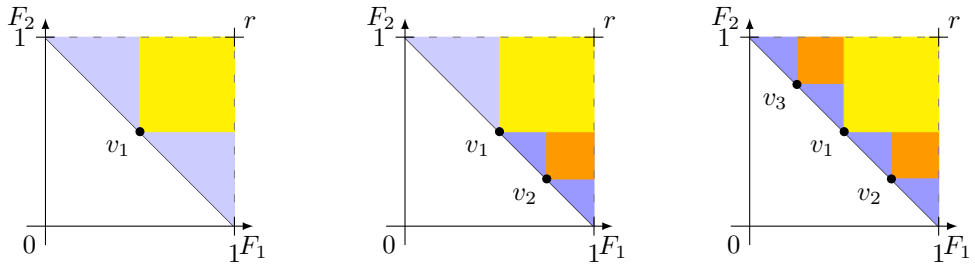

Fig. 5: The three greedy sets $\mathcal{S}_1$ (left), $\mathcal{S}_2$ (middle) and $\mathcal{S}_3$ (right) and their gap regions for an affine Pareto front with a reference point r dominating the nadir point $(1,1)$. The area of the gap regions of $\mathcal{S}_1$ are half of $\mathrm{HV}_r(\mathrm{PF}_f)$ (left). The area of the new gap regions of $\mathcal{S}_2$ is a quarter of the area of their parents (middle). The optimality gap of $\mathcal{S}_3$ (right) is half of the optimality gap of $\mathcal{S}_1$.

LEMMA 4.6. *We consider a greedy set sequence $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ relative to a valid reference point r. If the Pareto front is described by a bilipschitz function f with a Hölder continuous derivative, then for all $\epsilon > 0$, for n large enough, for every non-empty gap region $\mathcal{G}^r_{\mathcal{S}_n,i}$ and every child $\mathcal{G}^r_{\mathcal{S}_m,j}$ of $\mathcal{G}^r_{\mathcal{S}_n,i}$, we have*

$$\frac{1}{2}\times(1-\epsilon) \leq \frac{\max_{u\in\mathcal{G}^r_{\mathcal{S}_n,i}} HVI_r(u,\mathcal{S}_n)}{\lambda(\mathcal{G}^r_{\mathcal{S}_n,i})} \leq \frac{1}{2}\times(1+\epsilon) \text{ and} \tag{4.10}$$

$$\frac{1}{4\times(1+\epsilon)} \leq \frac{\lambda(\mathcal{G}^r_{\mathcal{S}_m,j})}{\lambda(\mathcal{G}^r_{\mathcal{S}_n,i})} \leq \frac{1}{4\times(1-\epsilon)} \ . \tag{4.11}$$

Proof. The set of the first coordinates of the vectors of the Pareto front which dominate r^n_i is the interval $[w^n_{i-1,r}, w^n_{i,r}]$. The restriction to $[w^n_{i-1,r}, w^n_{i,r}]$ of the function f describing the Pareto front is $(L_{\min},L_{\max})$-bilipschitz for some $L_{\min}$ and $L_{\max}$ such that $\frac{L_{\max}}{L_{\min}} = q_n$ with q_n defined in (4.9). Additionally, as stated in Proposition B.2, for n large enough, all the r^n_i corresponding to non-empty gap regions dominate the nadir point. It allows us to apply both Lemma A.2 and Proposition 3.11 to such gap regions. By Propositions 3.10 and 3.11, $\frac{\max_{u\in\mathrm{PF}_f}\mathrm{HV}_{r^n_i}(u)}{\mathrm{HV}_{r^n_i}(\mathrm{PF}_f)}$ is between $\frac{1}{2}\times\frac{1}{q_n}$ and $\frac{1}{2}\times q_n$.

Additionally, by Lemma A.2, $\frac{\lambda(\mathcal{G}^r_{\mathcal{S}_m,j})}{\mathrm{HV}_{r^n_i}(\mathrm{PF}_f)}$ is between $\frac{1-\frac{1}{2}\times q_n}{1+q_n^2}$ and $\frac{1-\frac{1}{2}\times\frac{1}{q_n}}{1+\frac{1}{q_n^2}}$. The maximum over the vectors u belonging to the Pareto front of $\mathrm{HV}_{r^n_i}(u)$ is equal to the maximum over u belonging to the i-th gap region of $\mathcal{S}_n$ of $\mathrm{HVI}_r(u,\mathcal{S}_n)$. Indeed, $\mathrm{HV}_{r^n_i}(.)$ is null for vectors outside the i-th gap region of $\mathcal{S}_n$ while it is nonnegative, equal to $\mathrm{HVI}_r(.,\mathcal{S}_n)$, otherwise. Additionally, $\mathrm{HV}_{r^n_i}(\mathrm{PF}_f)$ equals $\lambda(\mathcal{G}^r_{\mathcal{S}_n,i})$. The convergence of q_n to 1 stated in Lemma 4.5 imply that the bounds proven so far converge to a half and a quarter, respectively. Thus, we have (4.10) and (4.11) for n large enough. ∎

The following lemma states that for n large enough, the area of two non-empty gap regions relative to the same greedy set cannot be too different. More precisely, the area of any gap region of $\mathcal{S}_n$ cannot be more than $4\times(1+o(\epsilon))$ times greater than the area of another gap region of $\mathcal{S}_n$. The proof relies on considering the parents of the gap regions.

LEMMA 4.7. *We consider a greedy set sequence $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ relative to a valid reference point r. If the Pareto front is described by a bilipschitz function with a Hölder continuous derivative, then for all $\epsilon > 0$, for n large enough and for any non-empty gap regions of $\mathcal{S}_n$, $\mathcal{G}^r_{\mathcal{S}_n,i}$ and $\mathcal{G}^r_{\mathcal{S}_n,j}$ with $i, j \in [\![1, n+1]\!]$, we have*

$$\frac{\lambda(\mathcal{G}^r_{\mathcal{S}_n,i})}{\lambda(\mathcal{G}^r_{\mathcal{S}_n,j})} \leq 4 \times \frac{(1+\epsilon)^2}{1-\epsilon}. \tag{4.12}$$

Proof. Fix $\epsilon > 0$. By Lemma 4.6, there exists $N_1 \in \mathbb{N}^*$ such that for all n greater than N_1, (4.10) and (4.11) are verified for any non-empty gap region of $\mathcal{S}_n$ and its children. Since $\max_{i\in[\![1,n+1]\!]} w^n_{i,r} - w^n_{i-1,r}$ converges to 0 by Lemma 4.4, every non-empty gap region is filled at some point. Take N_2 such that all the non-empty gap regions of $\mathcal{S}_{N_1}$ are filled at iteration N_2. For all n greater than N_2, (4.10) and (4.11) are true for any non-empty gap region of $\mathcal{S}_n$ and its children, but also for its parents.

Take $n \geq N_2$. We note $\mathcal{G}_1 := \mathcal{G}^r_{\mathcal{S}_n,i}$ and $\mathcal{G}_2 := \mathcal{G}^r_{\mathcal{S}_n,j}$ two distinct non-empty gap regions of $\mathcal{S}_n$, and $\mathcal{P}_1$ and $\mathcal{P}_2$ their respective parents. When two sets correspond to gap regions relative to the same greedy set $\mathcal{S}_m$, we say that they cohabit at iteration m. Since only one vector is added to $\mathcal{S}_n$ at a time, the cohabitation of $\mathcal{G}_1$ and $\mathcal{G}_2$ implies that either $\mathcal{G}_1$ and $\mathcal{P}_2$ or $\mathcal{G}_2$ and $\mathcal{P}_1$ cohabit at some earlier iteration. In the first case, there necessarily exists $m \geq N_2$ such that $\mathcal{P}_2$ and $\mathcal{G}_1$ are gap regions relative to $\mathcal{S}_m$ and v_{m+1} belongs to $\mathcal{P}_2$, otherwise, $\mathcal{G}_1$ and $\mathcal{G}_2$ would not cohabit. By (4.10), the maximum hypervolume improvement to $\mathcal{S}_m$ of a vector of $\mathcal{G}_1$ and of a vector of $\mathcal{P}_2$ are at least $\frac{1}{2} \times (1-\epsilon) \times \lambda(\mathcal{G}_1)$ and at most $\frac{1}{2} \times (1+\epsilon) \times \lambda(\mathcal{P}_2)$, respectively. Since a vector of $\mathcal{P}_2$, v_{m+1}, maximizes the hypervolume improvement to $\mathcal{S}_m$, we have $\lambda(\mathcal{G}_1) \times \frac{1}{2} \times (1-\epsilon) \leq \lambda(\mathcal{P}_2) \times \frac{1}{2} \times (1+\epsilon)$. Since $\lambda(\mathcal{P}_2)$ is lower than $4 \times (1+\epsilon)$ times the area of its child $\lambda(\mathcal{G}_2)$ by (4.11), this inequality implies (4.12). In the second case, $\mathcal{P}_2$ is filled before $\mathcal{P}_1$. Thus, there exists $m \geq N_2$ such that $\mathcal{P}_1$ and $\mathcal{P}_2$ cohabit at iteration m and v_{m+1} belongs to $\mathcal{P}_2$. Since the area of $\mathcal{P}_1$ is higher than the one of its child $\mathcal{G}_1$, the hypervolume improvement of v_{m+1} to $\mathcal{S}_m$ is still higher than $\frac{1}{2} \times (1-\epsilon) \times \lambda(\mathcal{G}_1)$. The rest of the argumentation remains valid. ∎

We now have all the results needed to analyze the asymptotic impact of doubling the number of points in the greedy set. To prove the following asymptotic upper bound, we rely on similar arguments as for its nonasymptotic counterpart, Proposition 4.1. The previous lemma guarantees that the impact of doubling the number of points in the greedy set is asymptotically similar to the impact of passing from n points to $2n+1$.

PROPOSITION 4.8. *Let $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ be a greedy set sequence relative to valid reference point r. If the Pareto front is described by a bilipschitz function f with a Hölder continuous derivative, then for all $\epsilon > 0$, we have for n large enough*

$$\frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \leq \frac{1}{2} + o(\epsilon). \tag{4.13}$$

Proof. Fix $\epsilon > 0$. Fix n large enough to verify (4.10) and (4.12) for this particular ϵ.

Let σ be a permutation of $[\![1, n+1]\!]$ such that the i-th gap region of $\mathcal{S}_n$ is filled by $v_{n+\sigma(i)}$ when it is filled before iteration $2n+1$. With this choice of permutation, $\mathcal{G}^r_{\mathcal{S}_n,i}$ is always a gap region of $\mathcal{S}_{n+\sigma(i)-1}$. Thus, $\mathrm{HVI}_r(v_{n+\sigma(i)}, \mathcal{S}_{n+\sigma(i)-1})$ is superior to the maximum hypervolume improvement of a vector of $\mathcal{G}^r_{\mathcal{S}_n,i}$ to $\mathcal{S}_{n+\sigma(i)-1}$, which is superior to $\frac{1}{2} \times (1-\epsilon) \times \lambda(\mathcal{G}^r_{\mathcal{S}_n,i})$ by (4.10). It is equivalent to say that the hypervolume improvement

at iteration $n+\sigma(i)$ is higher than $\frac{1}{2}\times(1-\epsilon)\times\lambda(\mathcal{G}^r_{\mathcal{S}_n,i})$. Summing over $i\in[\![1,n+1]\!]$, we obtain that the hypervolume improvement between iteration n and $2n+1$ is higher than the sum over i of $\frac{1}{2}\times(1-\epsilon)\times\lambda(\mathcal{G}^r_{\mathcal{S}_n,i})$, that is $\frac{1}{2}\times(1-\epsilon)$ times the optimality gap at iteration n.

Now, we need to bound the hypervolume improvement at iteration $2n+1$, that is $\mathrm{HVI}_r(v_{2n+1},\mathcal{S}_{2n})$. It is lower than $\frac{1}{2}\times(1+\epsilon)\times\max_{i\in[\![1,2n+1]\!]}\lambda(\mathcal{G}^r_{\mathcal{S}_{2n},i})$ by (3.4) and (4.10). Since the area of a gap region is lower than the one of its parent, the maximum area of a gap region is lower at iteration $2n$ than at iteration n. The maximum area of one of the more than $n-1$ gap regions of $\mathcal{S}_n$ is lower than $\frac{1}{n-1}\times\frac{4\times(1+\epsilon)^2}{1-\epsilon}$ times the optimality gap at iteration n by (4.12). To summarize, the hypervolume improvement at iteration $2n+1$ is lower than $\frac{2\times(1+\epsilon)^3}{(n-1)\times(1-\epsilon)}$ times the optimality gap at iteration n.

We conclude that the relation between the optimality gap at iteration $2n$ and at iteration n is lower than $1-\frac{1}{2}\times(1-\epsilon)+\frac{1-\epsilon}{2\times(n-1)}$. ∎

We broadly follow the same approach to obtain the following asymptotic lower bound on the impact of doubling the number of points in the greedy set. Lemmas 4.6 and 4.7 are key to prove a lower bound on the hypervolume improvement at iteration k. They allow to prove that filling a gap region of $\mathcal{S}_n$ more than once gives, up to a factor $1+o(\epsilon)$, a lower hypervolume improvement than filling a gap region which was not filled. Indeed, the area of a descendant of a gap region of $\mathcal{S}_n$ is at most $\frac{1}{4}+o(\epsilon)$ times the area of its parent by Lemma 4.6, which is itself at most $4+o(\epsilon)$ times the area of any other gap region of $\mathcal{S}_n$ by Lemma 4.7.

PROPOSITION 4.9. *Let $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ be a greedy set sequence relative to a valid reference point r. If the Pareto front is desribed by a bilipschitz function f with a Hölder continuous derivative, then for all $\epsilon>0$, we have for n large enough*

$$\frac{HV_r(PF_f)-HV_r(\mathcal{S}_{2n})}{HV_r(PF_f)-HV_r(\mathcal{S}_n)}\geq\frac{1}{2}+o(\epsilon). \tag{4.14}$$

Proof. Fix $\epsilon>0$. Fix n large enough to verify (4.10), (4.11) and (4.12) for this particular ϵ. Let $\delta\in\{-1,0,1\}$ be such that $\mathcal{S}_n$ has $n+\delta$ non-empty gap regions. Let $i_0:=1$ when the left extreme gap region is empty and $i_0:=0$ otherwise.

Let σ be a permutation of $[\![1,n+\delta]\!]$ such that the i-th non-empty gap region of $\mathcal{S}_n$, $\mathcal{G}^r_{\mathcal{S}_n,i_0+i}$, is filled by the vector $v_{n+\sigma(i)}$ when it is filled before iteration $2n+\delta$. We distinguish two cases. In the first case, $v_{n+\sigma(i)}$ is the child of the i-th non-empty gap region of $\mathcal{S}_n$, and consequently its hypervolume improvement to $\mathcal{S}_{n+\sigma(i)-1}$ is at most $\frac{1}{2}\times(1+\epsilon)\times\lambda(\mathcal{G}^r_{\mathcal{S}_n,i_0+i})$ by (4.10). In the second case, $v_{n+\sigma(i)}$ belongs to $\mathcal{G}^r_{\mathcal{S}_n,i_0+j}$, the j-th non-empty gap region of $\mathcal{S}_n$, with $j\neq i$ and, by definition of σ, fills a descendant of this gap region not $\mathcal{G}^r_{\mathcal{S}_n,i_0+j}$ itself. By (4.10), the hypervolume improvement of $v_{n+\sigma(i)}$ to $\mathcal{S}_{n+\sigma(i)-1}$ is still at most $\frac{1}{2}\times(1+\epsilon)$ times the area of the gap region it fills. By (4.11), the area of a descendant of $\mathcal{G}^r_{\mathcal{S}_n,i_0+j}$ is smaller than $\frac{1}{4\times(1-\epsilon)}$ times the area of its ancestor. By (4.12), we also know that the area of the i-th non-empty gap region of $\mathcal{S}_n$ is at most $4\times\frac{(1+\epsilon)^2}{1-\epsilon}$ times the area of any other gap region of $\mathcal{S}_n$, in particular its i-th non-empty gap region. We conclude that the hypervolume improvement of $v_{n+\sigma(i)}$ to $\mathcal{S}_{n+\sigma(i)-1}$ is lower than $\frac{1}{2}\times\frac{(1+\epsilon)^3}{(1-\epsilon)^2}\times\lambda(\mathcal{G}^r_{\mathcal{S}_n,i_0+i})$. To summarize, since $1+\epsilon$ is lower than $\frac{(1+\epsilon)^3}{(1-\epsilon)^2}$, the hypervolume improvement at any iteration $n+\sigma(i)$ is lower than $\frac{1}{2}\times\frac{(1+\epsilon)^3}{(1-\epsilon)^2}\times\lambda(\mathcal{G}^r_{\mathcal{S}_n,i})$.

Summing over $i \in [\![1, n+\delta]\!]$, the hypervolume improvement from iteration n to $2n+\delta$ is lower than $\frac{1}{2} \times \frac{(1+\epsilon)^3}{(1-\epsilon)^2}$ times the sum over i of $\lambda(\mathcal{G}^r_{\mathcal{S}_n,i})$, that is the optimality gap at iteration n.

Now, it is left to prove an upper bound on $\mathrm{HV}_r(\mathcal{S}_{2n}) - \mathrm{HV}_r(\mathcal{S}_{2n+\delta})$. This quantity is the highest for $\delta = -1$, where it is simply the hypervolume improvement at iteration $2n$. As in the previous proof, it is lower than $\frac{1+\epsilon}{2\times(n-1)}$ times the optimality gap at iteration n. Therefore, the relation between the optimality gap at iteration $2n$ and at iteration n is higher than $1 - \frac{1}{2} \times \frac{(1+\epsilon)^3}{(1-\epsilon)^2} - \frac{1+\epsilon}{2\times(n-1)}$. ∎

We combine the lower and upper asymptotic bounds to obtain the following theorem.

THEOREM 4.10. *Consider a biobjective optimization problem and a greedy set sequence $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ relative to a valid reference point r. If the Pareto front is described by a bilipschitz function f with a Hölder continuous derivative, we have*

$$\frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \xrightarrow[n\to\infty]{} \frac{1}{2}. \tag{4.15}$$

Consequently, for such Pareto front and reference point and under Assumption 2.4 of perfect singleobjective optimization, HV-ISOOMOO algorithms relative to r satisfy (4.15) where $\mathcal{S}_n$ is replaced by $\mathcal{I}_n$, the final incumbents Pareto front approximation at iteration n.

5. Conclusion. We prove that HV-ISOOMOO algorithms coupled with a singleobjective optimizer converge in $O(1/n)$ on convex Pareto fronts and in $O(1/n^c)$ on bilipschitz Pareto fronts with $c \leq 1$ depending on the bilipschitz constants where n is the number of meta-iterations. Each meta-iteration corresponds to a singleobjective optimization run. Both bounds are tight over the class of Pareto fronts and reference points considered. Indeed, they are reached for affine Pareto fronts and reference points dominating the nadir point. On convex Pareto fronts, the highest achievable convergence-speed is reached [15]. It shows that greedily adding points maximizing the hypervolume contribution as in HV-ISOOMOO algorithms is an effective way to quickly increase the hypervolume. Additionally, we prove that for bilipschitz Pareto fronts with a Hölder continuous derivative, asymptotically, doubling the number of meta-iterations halves the optimality gap. This asymptotic behavior resembles what we would observe with an affine Pareto front and a reference point dominating the nadir point. Beware that this does not guarantee a rate of convergence in $\Theta(1/n)$. For example, the sequences $(\frac{\log(n)}{n})_{n\in\mathbb{N}^*}$ and $(\frac{1}{n\times\log(n)})_{n\in\mathbb{N}^*}$ both verify this property. The convergence rate on nonconvex Pareto fronts could theoretically be slower than $\Theta(1/n)$, but not faster by [15].

Appendix A. Normalized areas of the gap regions relative to an hypervolume maximizer. The goal of this section is to prove bounds on the normalized areas of the gap regions $\mathcal{G}^{u^*}_{\text{left}}$ and $\mathcal{G}^{u^*}_{\text{right}}$ relative to an hypervolume maximizer $u^* = (x^*, f(x^*))$ (see the lefthand plot of Figure 6) in the case of a bilipschitz Pareto front and of a reference point r dominating the nadir point. These bounds are stated in Lemma A.2.

We exploit the bounds on the normalized maximum hypervolume proven in Subsection 3.3 and the following lower and upper bounds on the relation between $\lambda(\mathcal{G}^{u^*}_{\text{left}})$ and $\lambda(\mathcal{G}^{u^*}_{\text{right}})$.

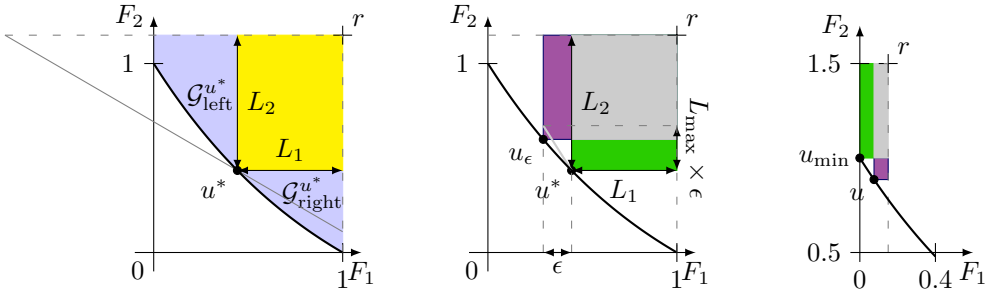

Fig. 6: Illustration of elements of the proofs of Proposition A.1 in the case $r_1 \leq x_{\max}$ (left and middle) and of Lemma B.1 (right). The Pareto front is described by $f(x) = \frac{e}{e-1} \times e^{-x} + 1 - \frac{e}{e-1}$ for $x \in [0, 1]$. Left: the gap regions $\mathcal{G}^{u^*}_{\text{left}}$ and $\mathcal{G}^{u^*}_{\text{right}}$ with a segment of slope $-L_{\min}$ passing through u^*. Middle: the hypervolume improvements $\text{HVI}_r(u^*, u_\epsilon)$ and $\text{HVI}_r(u_\epsilon, u^*)$. Right: the hypervolume improvement $\text{HVI}_r(u_{\min}, u)$ and its counterpart $\text{HVI}_r(u, u_{\min})$ where u is a vector of the Pareto front which dominates r.

PROPOSITION A.1. *We assume that the Pareto front is described by a ($L_{\min}$,$L_{\max}$)-bilipschitz function f. Let u^* be a non-extreme vector of the Pareto front which maximizes the hypervolume with respect to a valid reference point r. If $r_1 \leq x_{\max}$, we have $\lambda(\mathcal{G}^{u^*}_{right}) \geq \frac{L^2_{\min}}{L^2_{\max}} \times \lambda(\mathcal{G}^{u^*}_{left})$. If $r_2 \leq f(x_{\min})$, we have $\lambda(\mathcal{G}^{u^*}_{left}) \geq \frac{L^2_{\min}}{L^2_{\max}} \times \lambda(\mathcal{G}^{u^*}_{right})$.*

Proof. We consider the case where $r_1 \leq x_{\max}$. We note $L_1 := r_1 - x^*$ and $L_2 := r_2 - f(x^*)$ the lengths of the sides of the rectangle $\mathcal{D}^r_{u^*}$. For all $x, y \in [x_{\min}, x_{\max}]$, we have $|f(x) - f(y)| \geq L_{\min} \times |x - y|$. Additionally, since $r_1 \leq x_{\max}$, the segment $[x^*, x^* + L_1]$ is included in $[x_{\min}, x_{\max}]$. As a consequence, the section of the Pareto front on the right of u^* dominates the segment between u^* and $u^* + L_1 \times (1, -L_{\min})$, see the lefthand plot of Figure 6. Therefore, $\lambda(\mathcal{G}^{u^*}_{\text{right}})$ is larger than the area of the region of the objective space dominated by this segment, not dominated by u^* and dominating r, that is $\frac{1}{2} \times L_{\min} \times L_1^2$. For all $x, y \in [x_{\min}, x_{\max}]$, we also have $|f(x) - f(y)| \leq L_{\max} \times |x - y|$. Therefore, the part of the Pareto front on the left of u^* is dominated by the segment between u^* and $u^* + L_2 \times (-\frac{1}{L_{\min}}, 1)$, and $\lambda(\mathcal{G}^{u^*}_{\text{left}})$ is lower than $\frac{1}{2} \times \frac{1}{L_{\min}} \times L_2^2$. We have yet to prove a lower bound on $\frac{L_1}{L_2}$. The vector u^* being different from $u_{\min}$, for $\epsilon > 0$ small enough, the vector $u_\epsilon := (x^* - \epsilon, f(x^* - \epsilon))$ belongs to the Pareto front. As we can see in the middle plot of Figure 6, $\text{HVI}_r(u^*, u_\epsilon)$ is lower than $L_1 \times L_{\max} \times \epsilon$ and $\text{HVI}_r(u_\epsilon, u^*)$ is higher than $\epsilon \times (L_2 - \epsilon \times L_{\max})$. Additionally, u^* being an hypervolume maximizer, $\text{HVI}_r(u^*, u_\epsilon)$ is higher than $\text{HVI}_r(u_\epsilon, u^*)$, and thus $L_1 \times L_{\max} \geq L_2 - \epsilon \times L_{\max}$ for all $\epsilon > 0$. Taking the limit of this inequality when $\epsilon \to 0$, we obtain that $L_1 \times L_{\max} \geq L_2$. Combining the bounds on $\lambda(\mathcal{G}^{u^*}_{\text{left}})$ and $\lambda(\mathcal{G}^{u^*}_{\text{right}})$ with the lower-bound on $\frac{L_1}{L_2}$, we obtain the desired lower bound on $\lambda(\mathcal{G}^{u^*}_{\text{right}})$. We can obtain the symmetric inequality when $r_2 \geq f(x_{\min})$ by following the same approach. ∎

In particular, when f is bilipschitz and r dominates the nadir point, both bounds hold. We are now able to prove the desired bounds on the normalized area of the gap regions $\mathcal{G}^{u^*}_{\text{left}}$ and $\mathcal{G}^{u^*}_{\text{right}}$.

LEMMA A.2. *Let u^* be a vector which maximizes the hypervolume with respect to a valid reference point r. If the Pareto front is described by a ($L_{\min}$,$L_{\max}$)-bilipschitz function f and the reference point r dominates the nadir point, both $\mathcal{G}_{left}^{u^*}$ and $\mathcal{G}_{right}^{u^*}$ are between $(1-\frac{1}{2}\times\frac{L_{\max}}{L_{\min}})/(1+\frac{L_{\max}^2}{L_{\min}^2})$ and $(1-\frac{1}{2}\times\frac{L_{\min}}{L_{\max}})/(1+\frac{L_{\min}^2}{L_{\max}^2})$.*

Proof. Let note arbitrarily $\mathcal{G}_1$ and $\mathcal{G}_2$ the two gap regions of the set $S=\{u^*\}$. By Proposition A.1, $\lambda(\mathcal{G}_2)$ is between $\frac{L_{\min}^2}{L_{\max}^2}\times\lambda(\mathcal{G}_1)$ and $\frac{L_{\max}^2}{L_{\min}^2}\times\lambda(\mathcal{G}_1)$. Additionally, Propositions 3.10 and 3.11, the normalized maximum hypervolume $\max_{u\in\mathrm{PF}_f}\mathrm{HV}_r(u)$ over $\mathrm{HV}_r(\mathrm{PF}_f)$ is between $\frac{1}{2}\times\frac{L_{\min}}{L_{\max}}$ and $\frac{1}{2}\times\frac{L_{\max}}{L_{\min}}$. These bounds can be transformed into bounds on $\mathrm{HV}_r(\mathrm{PF}_f)-\max_{u\in\mathrm{PF}_f}\mathrm{HV}_r(u)$, that is $\lambda(\mathcal{G}_1)+\lambda(\mathcal{G}_2)$. As a consequence, $\lambda(\mathcal{G}_1)$ is between $(1-\frac{1}{2}\times\frac{L_{\max}}{L_{\min}})\times\mathrm{HV}_r(\mathrm{PF}_f)-\frac{L_{\max}^2}{L_{\min}^2}\times\lambda(\mathcal{G}_1)$ and $(1-\frac{1}{2}\times\frac{L_{\min}}{L_{\max}})\times\mathrm{HV}_r(\mathrm{PF}_f)-\frac{L_{\min}^2}{L_{\max}^2}\lambda(\mathcal{G}_1)$. Moving all the $\lambda(\mathcal{G}_1)$ terms on the same side and re-normalizing this side, we obtain the desired bounds for $\mathcal{G}_1$, which can be chosen to be either $\mathcal{G}_{\text{left}}^{u^*}$ or $\mathcal{G}_{\text{right}}^{u^*}$. ∎

Appendix B. The nadir point is dominated by all the r_i^n corresponding to non-empty gap regions for n large. We show in this section that for bilipschitz Pareto fronts, the nadir point is dominated by all the r_i^n corresponding to non-empty gap regions, for n large enough. This result is stated in Proposition B.2 and used in Subsection 4.2. It is equivalent to prove that the extreme vectors which dominate the reference point belong to the greedy set for n large enough.

First, we prove in the next proposition that if $r_1>x_{\max}$ (resp. $r_2>f(x_{\min})$), then for r_2 (resp. r_1) close enough to $f(x_{\max})$ (resp. $x_{\min}$) the extreme vector $u_{\max}$ (resp. $u_{\min}$) is the only hypervolume maximizer, see the righthand plot of Figure 6. There are similar statements in [8] for the set of μ points maximizing the hypervolume, but they only apply to $\mu\geq 2$.

LEMMA B.1. *We assume that the Pareto front is described by a function f which is ($L_{\min}$,$L_{\max}$)-bilipschitz. For any valid reference point $r=(r_1,r_2)$ such that $r_1>x_{\max}$ and $f(x_{\max})<r_2<f(x_{\max})+L_{\min}\times(r_1-x_{\max})$, the right extreme of the Pareto front $u_{\max}$ is the only maximizer of $HV_r(.)$. Additionally, for any valid reference point $r=(r_1,r_2)$ such that $r_2>f(x_{\min})$ and $x_{\min}<r_1<x_{\min}+\frac{r_2-f(x_{\min})}{L_{\max}}$, the vector $u_{\min}=(x_{\min},f(x_{\min}))$ is the only maximizer of $HV_r(.)$.*

Proof. This proof is illustrated in the righthand plot of Figure 6. Let r be a reference point such that $r_2>f(x_{\min})$ and $x_{\min}<r_1<x_{\min}+\frac{r_2-f(x_{\min})}{L_{\max}}$. Let $u=(x,f(x))\neq u_{\min}$ be a vector of the Pareto front which dominates r. The hypervolume improvement of $u_{\min}$ to $\{u\}$ is $(r_2-f(x_{\min}))\times(x-x_{\min})$. The hypervolume improvement of u to $\{u_{\min}\}$ is equal to $(f(x_{\min})-f(x))\times(r_1-x)$, which is smaller than $L_{\max}\times(x-x_{\min})\times(r_1-x_{\min})$ since u dominates r and f is ($L_{\min}$,$L_{\max}$)-bilipschitz. Since we assume that $L_{\max}\times(r_1-x_{\min})<r_2-f(x_{\min})$, the upper bound on $\mathrm{HVI}_r(u,u_{\min})$ is strictly lower than $\mathrm{HVI}_r(u_{\min},u)$. As a consequence, the hypervolume of $u_{\min}$ is strictly higher than the one of u. We conclude that $u_{\min}$ is the unique hypervolume maximizer. The symmetric result can be obtained with the same approach. ∎

It is left to prove that when $r_1>x_{\max}$ (resp. $r_2>f(x_{\min})$), the second coordinate of r_{n+1}^n (resp. the first coordinate of r_0^n) indeed converge to $f(x_{\max})$ (resp. $x_{\min}$). It is a straightforward consequence of Lemma 4.4. Therefore, we are able to conclude.

PROPOSITION B.2. *We assume that the the Pareto front is described by a bilipschitz function. Let $(\mathcal{S}_n)_{n\in\mathbb{N}^*}$ be a greedy set sequence associated to a valid reference point r. For n large enough, every reference point r_i^n corresponding to a non-empty gap region $\mathcal{G}^r_{\mathcal{S}_n,i}$ dominates the nadir point.*

Proof. By Lemma 4.4, $w^n_{n,r}$ converges to $x_{\max}$, and thus the right extreme reference point $r^n_{n+1} := (r_1, f(w^n_{n,r}))$ converges to $(r_1, f(x_{\max}))$ by continuity of f. Therefore, if r_1 is strictly higher than $x_{\max}$, then there exists N such that for all $n \geq N$, r^n_{n+1} verifies the assumptions on the reference point of Lemma B.1 which guarantee that $u_{\max}$ is the unique maximizer of $\mathrm{HV}_r(.)$ over the right extreme gap region $\mathcal{G}^r_{\mathcal{S}_n,n+1}$. Let assume that $u_{\max}$ does not belong to $\mathcal{S}_n$. Then, $w^N_{N,r} \neq x_{\max}$, and since $w^n_{n,r}$ converges to $x_{\max}$, the left extreme gap region $\mathcal{G}^r_{\mathcal{S}_n,i}$ is necessarily filled at some later iteration. When the right extreme gap region is filled, $u_{\max}$, the unique minimizer of $\mathrm{HV}_r(.)$ over this gap region, is added to the greedy set. To summarize, if $r_1 > x_{\max}$, then for n large enough $\mathcal{S}_n$ contains $u_{\max}$, and thus the right extreme gap region is empty. We can prove with the same approach that for $r_2 > f(x_{\min})$, $\mathcal{S}_n$ contains $u_{\min}$ for n large enough.

At any iteration, the non-extreme reference points dominate the nadir point. Additionally, we proved that either $r_1 < x_{\max}$ (resp. $r_2 < f(x_{\min})$), and thus the left (resp. right) extreme reference point dominates the nadir point or for n large enough, the left (resp. right) extreme gap region is empty. ∎

REFERENCES

[1] A. Auger, J. Bader, and D. Brockhoff, *Theoretically Investigating Optimal μ-Distributions for the Hypervolume Indicator: First Results for Three Objectives*, in Parallel Problem Solving from Nature, PPSN XI, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, R. Schaefer, C. Cotta, J. Kołodziej, and G. Rudolph, eds., vol. 6238, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 586–596, https://doi.org/10.1007/978-3-642-15844-5_59.

[2] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler, *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, New York, NY, USA, Jan. 2009, Association for Computing Machinery, pp. 87–102, https://doi.org/10.1145/1527125.1527138.

[3] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler, *Hypervolume-based multiobjective optimization: Theoretical foundations and practical implications*, Theoretical Computer Science, 425 (2012), pp. 75–103, https://doi.org/10.1016/j.tcs.2011.03.012.

[4] N. Beume, M. Laumanns, and G. Rudolph, *Convergence Rates of (1+1) Evolutionary Multiobjective Optimization Algorithms*, in Parallel Problem Solving from Nature, PPSN XI, R. Schaefer, C. Cotta, J. Kołodziej, and G. Rudolph, eds., Lecture Notes in Computer Science, Berlin, Heidelberg, 2010, Springer, pp. 597–606, https://doi.org/10.1007/978-3-642-15844-5_60.

[5] N. Beume, B. Naujoks, and M. Emmerich, *SMS-EMOA: Multiobjective selection based on dominated hypervolume*, European Journal of Operational Research, 181 (2007), pp. 1653–1669, https://doi.org/10.1016/j.ejor.2006.08.008.

[6] J. Bigeon, S. L. Digabel, and L. Salomon, *DMulti-MADS: Mesh adaptive direct multisearch for blackbox multiobjective optimization*, p. 30.

[7] K. Bringmann and T. Friedrich, *The maximum hypervolume set yields near-optimal approximation*, in Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation - GECCO '10, Portland, Oregon, USA, 2010, ACM Press, p. 511, https://doi.org/10.1145/1830483.1830576.

[8] D. Brockhoff, *Optimal μ-Distributions for the Hypervolume Indicator for Problems with Linear Bi-objective Fronts: Exact and Exhaustive Results*, in Simulated Evolution and Learning, K. Deb, A. Bhattacharya, N. Chakraborti, P. Chakroborty, S. Das, J. Dutta, S. K. Gupta,

A. Jain, V. Aggarwal, J. Branke, S. J. Louis, and K. C. Tan, eds., vol. 6457, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 24–34, https://doi.org/10.1007/978-3-642-17298-4_2.

[9] A. L. Custódio and J. F. A. Madeira, *MultiGLODS: Global and local multiobjective optimization using direct search*, J Glob Optim, 72 (2018), pp. 323–345, https://doi.org/10.1007/s10898-018-0618-1.

[10] A. L. Custódio, J. F. A. Madeira, A. I. F. Vaz, and L. N. Vicente, *Direct Multisearch for Multiobjective Optimization*, SIAM J. Optim., 21 (2011), pp. 1109–1140, https://doi.org/10.1137/10079731X.

[11] E. de Klerk, F. Glineur, and A. B. Taylor, *On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions*, arXiv:1606.09365 [math], (2016), https://arxiv.org/abs/1606.09365.

[12] R. Fiorenza, *Hölder and locally Hölder Continuous Functions, and Open Sets of Class C^k, $C^{k,\lambda}$*, Frontiers in Mathematics, Birkhäuser Basel, 2016, https://doi.org/10.1007/978-3-319-47940-8.

[13] E. H. Fukuda and L. M. G. Drummond, *On the convergence of the projected gradient method for vector optimization*, Optimization, 60 (2011), pp. 1009–1021, https://doi.org/10.1080/02331934.2010.522710.

[14] J. Knowles and D. Corne, *On metrics for comparing nondominated sets*, in Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600), vol. 1, May 2002, pp. 711–716 vol.1, https://doi.org/10.1109/CEC.2002.1007013.

[15] E. Marescaux and N. Hansen, *Hypervolume in biobjective optimization cannot converge faster than $\Omega(1/p)$*, in GECCO 2021 - The Genetic and Evolutionary Computation Conference, Lille / Virtual, France, July 2021.

[16] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, *An analysis of approximations for maximizing submodular set functions—I*, Mathematical Programming, 14 (1978), pp. 265–294, https://doi.org/10.1007/BF01588971.

[17] C. Touré, N. Hansen, A. Auger, and D. Brockhoff, *Uncrowded Hypervolume Improvement: COMO-CMA-ES and the Sofomore framework*, arXiv:1904.08823 [cs, math], (2019), https://doi.org/10.1145/3321707.3321852, https://arxiv.org/abs/1904.08823.

[18] D. A. V. Veldhuizen and G. B. Lamont, *Evolutionary Computation and Convergence to a Pareto Front*, in Proceedings of the Third Annual Conference on Genetic Programming, San Francisco, 1998, Stanford University Bookstore, pp. 221–228.

[19] E. Zitzler, J. Knowles, and L. Thiele, *Quality Assessment of Pareto Set Approximations*, in Multiobjective Optimization: Interactive and Evolutionary Approaches, J. Branke, K. Deb, K. Miettinen, and R. Słowiński, eds., Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2008, pp. 373–404, https://doi.org/10.1007/978-3-540-88908-3_14.