



HAL
open science

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

Eugénie Marescaux, Anne Auger

► **To cite this version:**

Eugénie Marescaux, Anne Auger. Multiobjective Hypervolume Based ISOOMOO Algorithms Converge with At Least Sublinear Speed to the Entire Pareto Front. 2021. hal-03198414v1

HAL Id: hal-03198414

<https://hal.science/hal-03198414v1>

Preprint submitted on 13 Jul 2021 (v1), last revised 22 Apr 2022 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

4 EUGÉNIE MARESCAUX AND ANNE AUGER

5 **Abstract.** In multiobjective optimization, one is interested in finding a good approximation of the
6 Pareto set and the Pareto front, i.e the sets of best compromises in the decision and objective spaces,
7 respectively. In this context, we introduce a new algorithm framework, Incremental SingleObjective
8 Optimization for MultiObjective Optimization (ISOOMOO) for approximating the Pareto front with
9 an increasing number of points. We focus on HV-ISOOMOO, its instantiation with the hypervolume
10 indicator, a set-quality indicator which is widely used for algorithms design and performance assessment.
11 HV-ISOOMOO algorithms approximate the Pareto front by greedily maximizing the hypervolume. We
12 study the convergence to the entire Pareto front of HV-ISOOMOO coupled with a perfect singleobjective
13 optimizer. The convergence is defined as the convergence of the hypervolume of the sets of all meta-
14 iterations incumbents towards the hypervolume of the Pareto front. We prove tight lower bounds on the
15 convergence-speed for convex and bilipschitz Pareto fronts in $O(1/n^c)$ with n being the number of meta-
16 iterations and $c = 1$ and $c \leq 1$, respectively. For convex Pareto fronts, the convergence rate is exactly in
17 $\Theta(1/n)$, namely the highest convergence rate achievable by a biobjective optimization algorithm. These
18 are the first results on the convergence-speed of multiobjective optimization algorithms towards the entire
19 Pareto front. We also analyze theoretically the asymptotic convergence behavior.

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

21 **AMS subject classifications.** 90C29, 90C30

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

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

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

52 In this paper, we introduce a new anytime algorithm framework, Incremental Sin-
 53 gleObjective Optimization for MultiObjective Optimization (ISOOMOO). We focus on
 54 its instantiation with the hypervolume indicator, HV-ISOOMOO. HV-ISOOMOO al-
 55 gorithms try to greedily maximize the hypervolume by adding points with the highest
 56 hypervolume contribution. These points are obtained by running a singleobjective opti-
 57 mizer. Greediness is already used in the selection part of some multiobjective optimization
 58 evolutionary algorithms such as SMS-EMOA [5] to approximate the set of p points among
 59 the population with the highest hypervolume. The hypervolume of such discrete greedy
 60 approximation is proven to be at least $(e-1)/e$ times the one of the p -optimal distribution
 61 [16]. To our knowledge, we provide the first continuous equivalent of this result. We in-
 62 vestigate the rate of convergence of HV-ISOOMOO towards the whole Pareto front in the
 63 ideal case of perfect singleobjective optimization, measuring the convergence with the hy-
 64 pervolume. For convex and bilipschitz Pareto fronts, we prove that the convergence-speed
 65 is in $O(1/n^c)$ with $c = 1$ and $c \leq 1$, respectively, with n being the number of singleob-
 66 jective optimization runs performed. For convex Pareto fronts, since it has already been
 67 proven that no biobjective optimization algorithm can converge to the Pareto front faster
 68 than in $\Omega(1/n)$ [15], the convergence rate is exactly in $\Theta(1/n)$. Additionally, we prove
 69 that for bilipschitz and smooth enough Pareto fronts doubling the number of points in
 70 the approximation halves the optimality gap, asymptotically. In the proof process, we
 71 obtain bounds on the normalized maximum hypervolume and a geometric interpretation
 72 of optimality conditions.

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

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

84 **2. Background, algorithm framework and assumptions.** Here, we lay the
 85 foundations of the problem we analyze. First, we recall some classic concepts of mul-
 86 tiobjective optimization. Then, we introduce the ISOOMOO class of algorithms and
 87 its hypervolume based instantiation HV-ISOOMOO. We also formalize a mathematical
 88 abstraction of HV-ISOOMOO, the greedy set sequences. Finally, we examine our as-
 89 sumptions on the biobjective optimization problem.

90 **2.1. Biobjective optimization problems, the Pareto front and the hyper-**
 91 **volume indicator.** We consider a biobjective minimization problem:

92 (2.1)
$$\min_{x \in \Omega \subset \mathbb{R}^d} F(x)$$

94 with $F : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^2 : x \mapsto (F_1(x), F_2(x))$. We define two dominance relations for
 95 vectors in the objective space. We say that u weakly dominates v denoted by $u \preceq v$ if
 96 $u_1 \leq v_1$ and $u_2 \leq v_2$ and that u dominates v denoted by $u \prec v$ if $u \preceq v$ and $u \neq v$.
 97 A vector of the objective-space \mathbb{R}^2 is said *feasible* when it belongs to $F(\Omega)$. Solving the
 98 optimization problem consists in finding a good approximation of the *Pareto front*, the
 99 set of non-dominated feasible vectors, $\{F(X) : X \in \Omega, \forall Y \in \Omega, F(Y) \not\prec F(X)\}$. We
 100 restrict ourselves to Pareto fronts with an explicit representation:

101 (2.2)
$$\text{PF}_f = \{(x, f(x)) : x \in [x_{\min}, x_{\max}]\}$$

103 with $f : \mathbb{R} \mapsto \mathbb{R}$ decreasing. We denote by $u_{\min} := (x_{\min}, f(x_{\min}))$ and $u_{\max} :=$
 104 $(x_{\max}, f(x_{\max}))$ the extreme vectors of the Pareto front. Likewise, we denote by $\tilde{u}_{\min,r} :=$
 105 $(\tilde{x}_{\min,r}, f(\tilde{x}_{\min,r}))$ and $\tilde{u}_{\max,r} := (\tilde{x}_{\max,r}, f(\tilde{x}_{\max,r}))$ the extremes vectors of the part of
 106 the Pareto front dominating a reference point r , with $\tilde{x}_{\min,r} := \max(x_{\min}, f^{-1}(r_2))$ and
 107 $\tilde{x}_{\max,r} := \min(x_{\max}, r_1)$. The vector $(x_{\max}, f(x_{\min}))$ is called the *nadir* point. All these
 notations are illustrated in Figure 1.

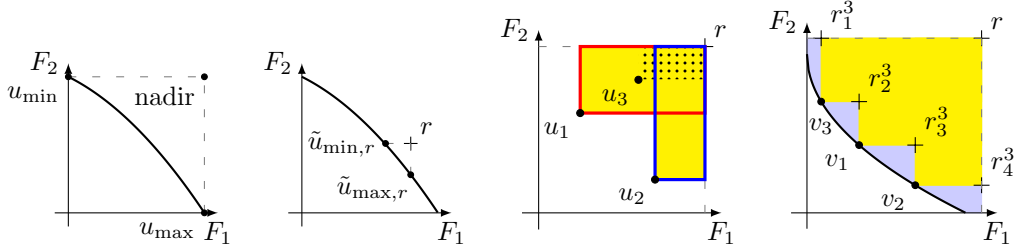


Fig. 1: Illustration of notations. The extreme vectors u_{\min} and u_{\max} and the nadir point $(x_{\max}, f(x_{\min}))$ (leftmost) ; the extreme vectors relative to the reference point r $\tilde{u}_{\min,r}$ and $\tilde{u}_{\max,r}$ (left) ; three vectors u_1 , u_2 and u_3 and the regions weakly dominated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

108
 109 The *hypervolume* with respect to a reference point r of an assessed set S of objective
 110 vectors, that we denote by $\text{HV}_r(S)$, is the Lebesgue measure of the region of the objective-
 111 space dominated by S and strictly dominating the reference point r . When no vector of
 112 the Pareto front dominates the reference point r , the hypervolume with respect to r
 113 of any set of feasible points of the objective space is null. Since this particular case is not
 114 interesting, we only consider reference points dominated by at least one vector of the
 115 Pareto front from now on. We refer to such reference points as *valid*.

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

118 (2.3)
$$\mathcal{D}_S^r = \{w \in \mathbb{R}^2 : \exists u \in S : u \preceq w \prec r\}.$$

120 The hypervolume of a set S relative to the reference point r equals $\lambda(\mathcal{D}_S^r)$ with $\lambda(\cdot)$ being
 121 the Lebesgue measure. In this paper, we only consider two-dimensional objective spaces
 122 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
 123 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,
 124 see the righthand plot of [Figure 1](#). Note that the \mathcal{D}_u^r are not disjoint.

125 We use the hypervolume to characterize the convergence of a set S of objective vectors
 126 to the entire Pareto front. For a fixed valid reference point r , a set S is said to converge
 127 to the Pareto front when the hypervolume difference $\text{HV}_r(\text{PF}_f) - \text{HV}_r(S)$ converges
 128 to 0. We define the *optimality gap* of S with respect to a valid reference point r as
 129 $\text{HV}_r(\text{PF}_f) - \text{HV}_r(S)$.

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

136 **2.2. The ISOOMOO framework, its HV-ISOOMOO instantiation and the**
 137 **associated greedy set sequences.** The Incremental SingleObjective Optimization for
 138 MultiObjective Optimization (ISOOMOO) framework builds incrementally a monotone¹
 139 sequence $(\mathcal{I}_n)_{n \in \mathbb{N}^*}$ of sets of vectors of the objective space. The pseudo code of ISOOMOO
 140 is given in [Algorithm 2.1](#), where the current value of \mathcal{I}_n is denoted by \mathcal{I} . At each so-
 141 called meta-iteration, a generic singleobjective maximization algorithm SOOPTIMIZER
 142 (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
 143 solution is added to \mathcal{I} (line 4 in [Algorithm 2.1](#)). We use the term meta-iteration to separate
 144 between the (meta-)iterations of ISOOMOO and the iterations of SOOPTIMIZER. Since
 145 the set \mathcal{I} is composed of the final objective incumbents of previous runs of SOOPTIMIZER
 146 and (ideally) provides an approximation of the Pareto front, we call it *final incumbents*
 147 *Pareto front approximation*.

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

Algorithm 2.1 Incremental SingleObjective Optimization for MultiObjective Optimiza-
 tion (ISOOMOO)

```

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

```

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

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

156 is a class of algorithms derived from ISOOMOO for which the maximization of the cri-
 157 terion J is *compliant* with the maximization of the hypervolume improvement as defined
 158 below.

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

$$164 \quad (2.4) \quad \operatorname{argmax}_{X \in \mathbb{R}^d} J(\mathcal{I}, X) = \operatorname{argmax}_{X \in \mathbb{R}^d} \operatorname{HVI}_r(F(X), \mathcal{I}) .$$

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

167 **DEFINITION 2.2** (HV-ISOOMOO). *We define an HV-ISOOMOO algorithm as an*
 168 *ISOOMOO algorithm as described in [Algorithm 2.1](#) where the criterion J satisfies [As-](#)*
 169 *sumption 2.1.*

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

176 **DEFINITION 2.3.** *We define the convergence of HV-ISOOMOO as the convergence of*
 177 *$\operatorname{HV}_r(\mathcal{I}_n)$ towards $\operatorname{HV}_r(\operatorname{PF}_f)$.*

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

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

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

195 The assumption of perfect singleobjective optimization is reminiscent to the assump-
 196 tion of perfect line search which is common in the analysis of gradient based methods [11].
 197 Under this assumption, all choices of criterions verifying [Assumption 2.1](#) are equivalent.
 198 The convergence of HV-ISOOMOO under perfect singleobjective optimization is a neces-
 199 sary condition for the soundness of the approach. Additionally, we could obtain a lower

200 bound on the convergence-speed of a real instantiation of HV-ISOOMOO by combining
 201 such lower bounds on the convergence-speed of HV-ISOOMOO under [Assumption 2.4](#)
 202 with existing lower bounds on the convergence speed of singleobjective optimization al-
 203 gorithms.

204 We introduce below the notions of greedy sequence and greedy set sequence, which
 205 are mathematical abstractions of HV-ISOOMOO under [Assumption 2.4](#) of perfect sin-
 206 gleobjective optimization.

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

209 (2.5)
$$v_1 \in \arg \max_{v \in F(\Omega)} HV_r(v) \text{ and}$$

210 (2.6)
$$v_{n+1} \in \arg \max_{v \in F(\Omega)} HV_r(\{v_1, \dots, v_n, v\}) \text{ for all } n \geq 1 .$$

 211

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

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

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

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

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

232 *Proof.* If f is lower semi-continuous, then for any reference point r , the maximum of
 233 $HV_r(\cdot)$ exists, see [3, Theorem 1]. Therefore, there exists a vector verifying (2.5) and the
 234 problem of maximizing the maximum of a finite number of hypervolume functions defined
 235 in (3.4) admits a solution. Since the recurrence equation defining greedy sequences (2.6)
 236 is equivalent to (3.4) by [Lemma 3.5](#), a lemma proven in the next section, we can build a
 237 sequence $\{v_n, n \geq 1\}$ verifying (2.5) and (2.6), namely a greedy sequence. \square

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

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

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

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

254 Yet, in general, there exists more than one greedy sequence, and thus greedy set sequence.
 255 For example, there are infinitely many greedy sequences associated to any affine Pareto
 256 front with a reference point dominating the nadir point. This statement relies on the fact
 257 that the unique maximizer of the hypervolume relative to a reference point r dominating
 258 the nadir point is the middle of the section of the Pareto front dominating r , see [2,
 259 Theorem 5]. As a consequence, the middle of the section of the Pareto front dominating r
 260 is the only candidate for v_1 but v_2 can be either at 1/4 or at 3/4 of this section. Similarly,
 261 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
 262 section of the Pareto front dominating r . For any n , we can find an iteration m such
 263 that v_m can be placed at 2^n different points, whatever the $m - 1$ first terms of the greedy
 264 sequence are.

265 **2.3. Assumptions on the Pareto front and the objective functions.** We
 266 present and discuss here the assumptions on the function f describing the Pareto front
 267 under which we derive convergence results. We typically assume that the function f is
 268 bilipschitz, convex or simultaneously bilipschitz and with a Hölder continuous derivative,
 269 namely $\mathcal{C}^{1,\alpha}$. Under any of these three assumptions, f is continuous. For the sake of
 270 conciseness, we transfer the properties of f to the Pareto front. For example, we call
 271 *convex Pareto front* a Pareto front described by a convex function. We recall that a
 272 function g is Hölder continuous with exponent α when there exists $H \geq 0$ such that
 273 $|g(x) - g(y)| \leq H \times |x - y|^\alpha$ for all x, y [12]. We note $[g]_\alpha$ the *minimum Hölder coefficient*
 274 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,
 275 we detail the bilipschitz constants and say that a bilipschitz function f is (L_{\min}, L_{\max}) -
 276 bilipschitz if for all $x, y \in [x_{\min}, x_{\max}]$, we have $L_{\min} \times |x - y| \leq |f(x) - f(y)| \leq L_{\max} \times |x - y|$
 277 where $L_{\max} \geq L_{\min} > 0$. We also consider *affine Pareto fronts*, the simplest kind of Pareto
 278 front. As they form a line in the biobjective case, they are usually referred to as linear
 279 Pareto fronts. They provide good examples to illustrate a point and help to understand
 280 the results we prove on the asymptotic convergence behavior.

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

283 **PROPOSITION 2.8.** *Given a biobjective minimization problem as in (2.1) whose Pareto*
 284 *front is described by a function f . If F_1 and F_2 are respectively $(L_{\min,1}, L_{\max,1})$ -bilipschitz*
 285 *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.*

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

288 and F_2 are convex, then f is convex.

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

293 Representing F_1 values on the absciss and F_2 values on the ordinate instead of the
 294 converse is an arbitrary choice. When f is a bijection, had we chosen to represent the
 295 F_2 values on the absciss instead of on the ordinate, we would have had another repre-
 296 sentation of the Pareto front : $\{(y, f^{-1}(y)) : y \in [f(x_{\max}); f(x_{\min})]\}$. If so, the inverse
 297 function f^{-1} would have played the role of f . It is interesting to notice that the choice
 298 of the objective function represented on the horizontal axis does not impact whether the
 299 function characterizing the Pareto front is bilipschitz or convex. Indeed, f being bilip-
 300 schitz is equivalent to both f and f^{-1} being lipschitz. Additionally, we can prove that
 301 given that the function f is decreasing, f being convex is equivalent to its inverse f^{-1}
 302 being convex. The proof of this property is straightforward. If f is convex, then for all
 303 $x, y \in \mathbb{R}$, $f\left(\frac{f^{-1}(x)+f^{-1}(y)}{2}\right)$ is smaller than $\frac{f(f^{-1}(x))+f(f^{-1}(y))}{2}$, that is $\frac{x+y}{2}$. Since f and
 304 therefore f^{-1} are decreasing, by composing by f^{-1} each side of the inequality, we obtain
 305 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
 306 $f^{-1}\left(\frac{x+y}{2}\right)$.

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

310 **3.1. Decomposition of the optimality gap using gap regions.** We introduced
 311 \mathcal{D}_S^r in (2.3) as the region of the objective space dominating r and weakly dominated by
 312 S . Its Lebesgue measure is $\text{HV}_r(S)$. We now introduce *total gap regions*, whose Lebesgue
 313 measure are optimality gaps.

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

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

321 **DEFINITION 3.2** (Gap regions, gaps and associated reference points). *Let $S =$*
 322 *$\{v_1, \dots, v_n\}$ be a set of n distinct vectors of the Pareto front dominating a valid refer-*
 323 *ence point r . Let σ be the permutation ordering the v_i by increasing F_1 values: $v_{\sigma(1),1} <$
 324 $v_{\sigma(2),1} < \dots < v_{\sigma(n),1}$.*

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

$$327 \quad (3.1) \quad \begin{aligned} r_1 &= (v_{\sigma(1),1}, r_2), r_{n+1} = (r_1, v_{\sigma(n),2}) \text{ and} \\ r_i &= (v_{\sigma(i),1}, v_{\sigma(i-1),2}) \text{ for all } i \in \llbracket 2, n \rrbracket . \end{aligned}$$

328 • *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 ,*
 329 *respectively.*

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

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

$$341 \quad (3.2) \quad HV_r(PF_f) - HV_r(\mathcal{S}_n) = \sum_{i=1}^{n+1} HV_{r_i^n}(PF_f) .$$

343 *Proof.* The optimality gap at iteration n is the Lebesgue measure of the total gap
 344 region $\mathcal{G}_{\mathcal{S}_n}^r$, which is the disjoint union of the gap regions $\mathcal{G}_{\mathcal{S}_n,i}^r$ of \mathcal{S}_n . Since each gap
 345 region $\mathcal{G}_{\mathcal{S}_n,i}^r$ equals $\mathcal{D}_{PF_f}^{r_i^n}$, the optimality gap is equal to $\sum_{i=1}^{n+1} HV_{r_i^n}(PF_f)$. \square

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

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

$$352 \quad (3.3) \quad HVI_r(u, \mathcal{S}_n) = HV_{r_i^n}(u) .$$

354 *Proof.* The hypervolume improvement of any $u \in \mathcal{G}_{\mathcal{S}_n,i}^r$ is the Lebesgue-measure of
 355 the intersection between $\mathcal{G}_{\mathcal{S}_n,i}^r$ and \mathcal{D}_u^r . Therefore, it is equal to $\lambda(\mathcal{D}_u^{r_i^n})$, that is $HV_{r_i^n}(u)$. \square

356 We can now reformulate the recurrence relation defining a greedy sequence at iteration
 357 $n + 1$. Indeed, picking a vector maximizing the hypervolume improvement to \mathcal{S}_n is
 358 equivalent to pick a vector where the highest value of the maximum of the hypervolumes
 359 with respect to the r_i^n is reached.

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

$$362 \quad (3.4) \quad v_{n+1} \in \arg \max_{u \in PF_f} \max_{i \in [1, n+1]} HV_{r_i^n}(u).$$

364 *Proof.* The hypervolume improvement of any vector u to \mathcal{S}_n is $\max_{i \in [1, n+1]} HV_{r_i^n}(u)$.
 365 It is a consequence of [Lemma 3.4](#) and of the fact that the hypervolume with respect to r_i^n
 366 is null outside the i -th gap region of \mathcal{S}_n . Additionally, v_{n+1} belongs to the Pareto front
 367 by [Proposition 2.7](#). Thus, (2.6) is equivalent to (3.4). \square

368 Similarly, we can express the decrease of the optimality gap at iteration $n + 1$,
 369 $HV_r(\mathcal{S}_{n+1}) - HV_r(\mathcal{S}_n)$, as the maximum of $n + 1$ hypervolume maximization problems.

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

$$372 \quad (3.5) \quad HV_r(\mathcal{S}_{n+1}) - HV_r(\mathcal{S}_n) = \max_{u \in PF_f} \max_{i \in [1, n+1]} HV_{r_i^n}(u) .$$

374 *Proof.* The hypervolume improvement $HV_r(\mathcal{S}_{n+1}) - HV_r(\mathcal{S}_n)$ is the hypervolume
 375 improvement of v_{n+1} to \mathcal{S}_n . With the same arguments as in the proof of Lemma 3.5, we
 376 can prove that it equals $\max_{u \in PF_f} \max_{i \in [1, n+1]} HV_{r_i^n}(u)$. \square

377 **3.2. Lower bound of the normalized maximum hypervolume for convex**
 378 **Pareto fronts.** In this section and the next one, we provide bounds on the maximum
 379 hypervolume achievable by a single feasible vector normalized by the maximum hyper-
 380 volume of a feasible set. We refer to this ratio as the *normalized maximum hypervolume*
 381 with respect to r :

$$382 \quad (3.6) \quad \frac{\max_{u \in PF_f} HV_r(u)}{HV_r(PF_f)} .$$

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

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

390 PROPOSITION 3.7. Let $x^* \in]x_{\min}, x_{\max}[$ such that $u^* := (x^*, f(x^*))$ maximizes the
 391 hypervolume with respect to a valid reference point $r = (r_1, r_2)$. If the function f de-
 392 scribing the Pareto front admits left and right derivatives in x^* , respectively $f'_-(x^*)$ and
 393 $f'_+(x^*)$, then

$$394 \quad (3.7) \quad -f'_+(x^*) \leq \frac{r_2 - f(x^*)}{r_1 - x^*} \leq -f'_-(x^*) .$$

396 *Proof.* We define the function $HV_{x,r}(\cdot)$ as $x \mapsto HV_r((x, f(x)))$. If x^* maximizes
 397 $HV_{x,r}(\cdot)$, then the left and the right derivatives of $HV_{x,r}(\cdot)$ are positive and negative,
 398 respectively. By replacing the left and right derivatives of $HV_{x,r}(\cdot)$ by their explicit
 399 formulas and reorganizing the terms we obtain (3.7). \square

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

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

$$409 \quad (3.8) \quad -f'(x^*) = \frac{r_2 - f(x^*)}{r_1 - x^*} .$$

411 *Proof.* It is a direct consequence of [Proposition 3.7](#) □

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

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

$$422 \quad (3.9) \quad \frac{\max_{u \in \text{PF}_f} \text{HV}_r(u)}{\text{HV}_r(\text{PF}_f)} \geq \frac{1}{2}$$

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

426 *Proof.* The function f being convex, it has left and right derivatives and thus, so does
 427 $\text{HV}_{x,r}(\cdot) : x \mapsto \text{HV}_r((x, f(x)))$. Thus, by [Proposition 3.7](#), (3.7) holds. Since f is convex,
 428 (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
 429 of f . Therefore, $\text{PF}_g := \{g(x) : x \in [x_{\min}, x_{\max}]\}$ dominates PF_f , and thus has a higher
 430 hypervolume.

431 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
 432 $\mathcal{R} := [\tilde{x}_{\min,r}, r_1] \times [f(\tilde{x}_{\max,r}), r_2]$. We note $l_1 := r_1 - x^*$ and $l_2 := r_2 - f(x^*)$ the lengths
 433 of the rectangle $\mathcal{D}_{u^*}^r$. The region of \mathcal{R} which dominates PF_g is a right-angled triangle.
 434 Additionally, by definition, the slope of its hypotenuse is $\frac{l_2}{l_1}$, and thus the lengths of the
 435 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
 436 of [Figure 2](#)). Therefore

$$437 \quad \begin{aligned} \text{HV}_r(\text{PF}_g) &= \lambda(\mathcal{R}) - \lambda(\{u \in \mathbb{R}^2 : u \in \mathcal{R}, u \preceq \text{PF}_g\}) \\ 438 \quad &= L_1 L_2 - \frac{1}{2} \times (L_1 - l_1 + (L_2 - l_2) \times \frac{l_1}{l_2}) \times (L_2 - l_2 + (L_1 - l_1) \times \frac{l_2}{l_1}) \\ 439 \quad &= l_1 l_2 \times \left[-2 + 2 \times \frac{L_2}{l_2} - \frac{1}{2} \times \left(\frac{L_2}{l_2}\right)^2 + 2 \times \frac{L_1}{l_1} - \frac{1}{2} \times \left(\frac{L_1}{l_1}\right)^2 \right]. \end{aligned}$$

441 For all x , we have $(x - 2)^2 \geq 0$ and thus $2x - \frac{1}{2}x^2 \leq 2$. Therefore, we can conclude
 442 that $\text{HV}_r(\text{PF}_g)$, and thus $\text{HV}_r(\text{PF}_f)$ is lower than $2 \times l_1 l_2$, that is $2 \times \text{HV}_r(u^*)$. If either
 443 $L_1/l_1 \neq 2$ or $L_2/l_2 \neq 2$, the inequality is strict. Thus, when the inequality is an equality,
 444 the center of \mathcal{R} belongs to the Pareto front. Since f is convex, it requires from f to
 445 be affine and from the reference point r to dominate the nadir point $(x_{\max}, f(x_{\min}))$.
 446 Conversely, if the Pareto front is affine and the reference point r dominates the nadir
 447 point, we know that the optimum is in the middle of the Pareto front and that we have
 448 the equality (see [[2](#), Theorem 5]). □

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

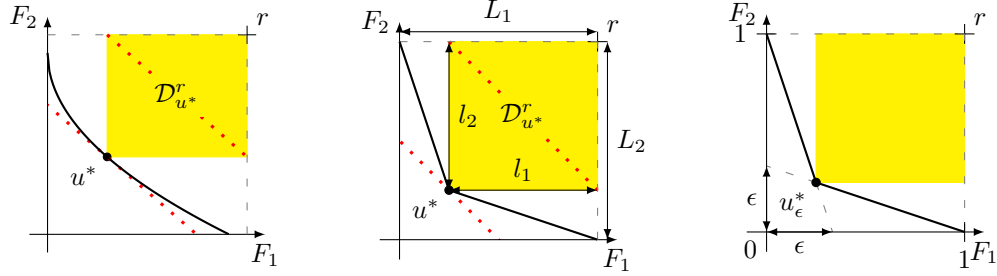


Fig. 2: Left and middle: Two convex Pareto fronts and their respective hypervolume maximizers u^* , one differentiable (left) and one non-differentiable (middle). The slopes of the two dotted lines, namely PF_g and the diagonal of $\mathcal{D}_{u^*}^r$, are equal. Right: The Pareto front PF_ϵ and the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

452 nadir point. Here is a simple example which illustrates this. Let consider the convex
 453 Pareto front $\text{PF}_\epsilon := \{\max(1 - \frac{x}{\epsilon}, \epsilon - \epsilon \times x) : x \in [0, 1]\}$ represented in the righthand
 454 plot of Figure 2 and the reference point $r = (1, 1)$. When $\epsilon \leq 1$, PF_ϵ is convex and
 455 (3.8) implies that $u_\epsilon^* = (\epsilon \times (1 - \epsilon), \epsilon \times (1 - \epsilon))$ is the unique hypervolume minimizer.
 456 Thus, the normalized maximum hypervolume of PF_ϵ for this reference point is equal to
 457 $\frac{(1-\epsilon+\epsilon^2)^2}{1-\epsilon \times (1-\epsilon)^2 + (\epsilon-\epsilon^2)^2}$ and converges to 1 when ϵ goes to 0.

458 **3.3. Lower and upper bounds of the normalized maximum hypervolume**
 459 **for bilipschitz Pareto fronts.** In this section, we examine lower and upper bounds on
 460 the normalized maximum hypervolume in the case of bilipschitz Pareto fronts.

461 We consider two affine fronts with the same left extreme vector as PF_f and slopes
 462 $-L_{\min}$ and $-L_{\max}$, see Figure 3. We call them PF_{\min} and PF_{\max} , respectively. Formally:

463 (3.10)
$$\text{PF}_{\max} := \{(x, f_{\max}(x)) : x \in [x_{\min}, x_{\max}]\},$$

464 (3.11)
$$\text{PF}_{\min} := \{(x, f_{\min}(x)) : x \in [x_{\min}, x_{\max}]\}$$

466 with $f_{\min}(x) = f(x_{\min}) - (x - x_{\min}) \times L_{\min}$ and $f_{\max}(x) = f(x_{\min}) - (x - x_{\min}) \times$
 467 L_{\max} . For a (L_{\min}, L_{\max}) -bilipschitz function f , $f_{\min}(x) \leq f(x) \leq f_{\max}(x)$ for $x \in$
 468 $[x_{\min}, x_{\max}]$, and thus the Pareto front is dominated by PF_{\max} and dominates PF_{\min} .
 469 These two affine fronts provide bounds on both the hypervolume of the Pareto front
 470 and the highest hypervolume of a vector on the Pareto front. They are key to prove
 471 the following lower bound on the normalized maximum hypervolume of a (L_{\min}, L_{\max}) -
 472 bilipschitz Pareto front.

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

476 (3.12)
$$\frac{\max_{u \in \text{PF}_f} \text{HV}_r(u)}{\text{HV}_r(\text{PF}_f)} \geq \frac{1}{2} \times \frac{L_{\min}}{L_{\max}} .$$

478 *Proof.* The fronts PF_{\max} and PF_{\min} are defined respectively in (3.10) and (3.11).
 479 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 $V :=$

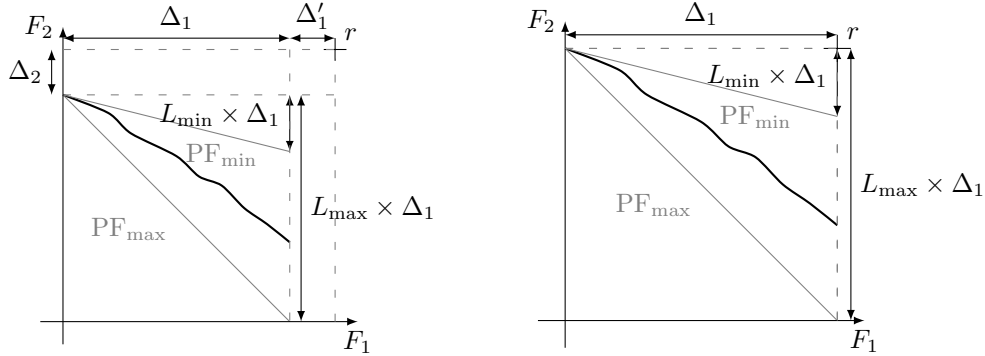


Fig. 3: The Pareto front PF_f surrounded by PF_{\max} and PF_{\min} in the case where the reference point r does not dominate the nadir point (left) and in the case where it does (right).

480 $\Delta_2 \times (r_1 - \tilde{x}_{\min,r})$, see the lefthand plot of [Figure 3](#). Since the front PF_{\max} dominates
481 the Pareto front, the hypervolume of PF_f is smaller than the hypervolume of PF_{\max} ,
482 $V + L_{\max} \times \Delta_1 \times \Delta'_1 + \frac{1}{2} \times L_{\max} \times \Delta_1^2$. Since each vector of PF_f is dominated by a
483 vector of PF_{\min} , the maximum hypervolume of a vector of PF_f is higher than the maximum
484 hypervolume of a vector of PF_{\min} . The front PF_{\min} being an affine and therefore convex
485 front, we know by [Proposition 3.9](#) that the maximum hypervolume of a vector of PF_{\min} is
486 higher than half of $HV_r(PF_{\min})$, which is equal to $\frac{1}{2} \times (V + L_{\min} \times \Delta_1 \times \Delta'_1 + \frac{1}{2} \times L_{\min} \times \Delta_1^2)$.
487 To summarize, the maximum hypervolume of a vector of PF_f is higher than $\frac{1}{2} \times (V + L_{\min} \times$
488 $\Delta_1 \times \Delta'_1 + \frac{1}{2} \times L_{\min} \times \Delta_1^2)$. Combining the upper bound on the hypervolume of PF_f and the
489 lower bound on the maximum hypervolume of a vector of PF_f , the normalized maximum
490 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
491 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
492 conclude that the normalized maximum hypervolume is higher than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}}$. \square

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

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

$$504 \quad (3.13) \quad \frac{\max_{u \in PF_f} HV_r(u)}{HV_r(PF_f)} \leq \frac{1}{2} \times \frac{L_{\max}}{L_{\min}} .$$

505

506 *Proof.* We use the same notations as in the proof of [Proposition 3.10](#). Since r domi-
507 nates the nadir point, both Δ'_1 , Δ_2 and V equal 0, and thus the hypervolumes of PF_{\max}
508 and 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
509 PF_{\min} by PF_f implies that the hypervolume of the Pareto front is below $\frac{1}{2} \times L_{\min} \times \Delta_1^2$.
510 Since PF_{\max} is an affine front whose extremes dominate r , its middle is the unique hyper-
511 volume maximizer (see [\[1, Theorem 5\]](#)) with an hypervolume equal to $\frac{1}{4} \times L_{\max} \times \Delta_1$. The
512 domination of PF_f by PF_{\max} implies that the maximum hypervolume of a vector of PF_f
513 is higher than $\frac{1}{4} \times L_{\max} \times \Delta_1^2$. Gathering the lower bound on $\text{HV}_r(\text{PF}_f)$ and the upper
514 bound on the maximum hypervolume of a vector of PF_f , we retrieve [\(3.13\)](#). \square

515 This upper bound is only relevant for $L_{\max}/L_{\min} < 2$. The bound is the tightest for
516 $L_{\max} = L_{\min}$, where it achieves the value $1/2$. In this paper, we use this upper bound for
517 L_{\max}/L_{\min} close to 1 to analyze the asymptotic convergence behavior of HV-ISOOMOO.

518 **4. Convergence of HV-ISOOMOO under perfect singleobjective optimiza-**
519 **tion.** We prove in this section various convergence results for HV-ISOOMOO algorithms
520 under the assumption of perfect singleobjective optimization. We first prove that when
521 the Pareto front is either convex or bilipschitz, these algorithms converge to the entire
522 Pareto front with guaranteed convergence-speed. We transform the bounds on the nor-
523 malized maximum hypervolume proven in [Section 3](#) into lower bounds on the convergence
524 speed. Second, we analyze the asymptotic convergence behavior when the Pareto front is
525 bilipschitz with a Hölder continuous derivatives.

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

534 **4.1. Convergence of HV-ISOOMOO with guaranteed convergence speed.**

535 We prove some upper bounds on the relation between the optimality gap at iteration
536 $2n + 1$ and at iteration n . These bounds translate into lower bounds on the speed of
537 convergence of HV-ISOOMOO with perfect singleobjective optimization. The proof relies
538 on inequalities of the form

$$539 \quad (4.1) \quad \max_{u \in \text{PF}_f} \text{HV}_{r'}(u) \geq C \times \text{HV}_{r'}(\text{PF}_f)$$

541 stated in [Propositions 3.9](#) and [3.10](#) and on the reformulation of optimality gaps, areas
542 of gap regions and hypervolume improvement done in [Subsection 3.1](#). A consequence of
543 [\(4.1\)](#) being true for any reference point r' is that the optimality gap at iteration $2n + 1$
544 is at most $(1 - C)$ times the optimality gap at iteration n .

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

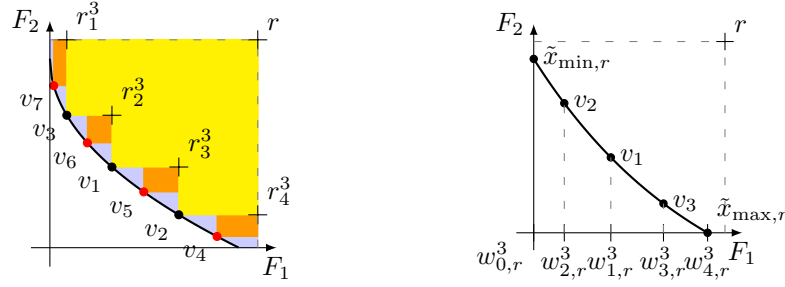


Fig. 4: Left: A Pareto front where each of the gap regions of \mathcal{S}_3 is filled by one of the greedy vectors v_k for $k \in \llbracket 4, 7 \rrbracket$. It is described by $f(x) = 1 - \sqrt{x}$ for $x \in [0, 1]$. We represent the region $\mathcal{D}_{\mathcal{S}_3}^r$ ■, the regions corresponding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

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

$$554 \quad (4.2) \quad \frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n+1})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \leq 1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \text{ if } f \text{ is } (L_{\min}, L_{\max})\text{-bilipschitz and}$$

$$555 \quad (4.3) \quad \frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n+1})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \leq \frac{1}{2} \text{ if } f \text{ is convex.}$$

557 *Proof.* Fix $n \geq 1$. We note σ a permutation of $\llbracket 1, n+1 \rrbracket$ such that $n + \sigma(i)$ is the
558 index of the first greedy vector v_k inserted in $\mathcal{G}_{\mathcal{S}_n, i}^r$ when possible. With this choice of σ ,
559 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
560 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
561 to $\text{HV}_{r_i^n}(u)$ by (3.3). The hypervolume improvement of the greedy vector $v_{n+\sigma(i)}$ to
562 $\mathcal{S}_{n+\sigma(i)-1}$ being maximal, it is in particular larger than the one of any vector of $\mathcal{G}_{\mathcal{S}_n, i}^r$,
563 the i -th gap region of \mathcal{S}_n , and thus than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times \text{HV}_{r_i^n}(PF_f)$ by Proposition 3.10.
564 In other words, the hypervolume improvement at any iteration $n + \sigma(i)$ is higher than
565 $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times \text{HV}_{r_i^n}(PF_f)$. By adding these inequations for all $i \in \llbracket 1, n+1 \rrbracket$, we deduce
566 that the hypervolume improvement from iteration n to $2n+1$ is larger than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times$
567 $\sum_{i=1}^{n+1} \text{HV}_{r_i^n}(PF_f)$. Since the sum of the $\text{HV}_{r_i^n}(PF_f)$ is the optimality gap at iteration
568 n , we have (4.2). If f is convex instead of bilipschitz, we use Proposition 3.9 instead of
569 Proposition 3.10 and obtain (4.3). □

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

575 THEOREM 4.2. Consider a biobjective optimization problem with a Pareto front de-
 576 scribed by a bilipschitz or convex function f .

577 The hypervolume of a greedy set sequence associated to a valid reference point r con-
 578 verges to the hypervolume of the entire Pareto front, i.e. $HV_r(\mathcal{S}_n) \xrightarrow{n \rightarrow \infty} HV_r(PF_f)$.

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

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

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

$$588 \quad (4.4) \quad \frac{HV_r(PF_f) - HV_r(\mathcal{S}_n)}{HV_r(PF_f)} \leq \left(1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}}\right)^{\lfloor \log_2(n+1) \rfloor} \leq (2n+2)^{\log_2(1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}})} .$$

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

$$592 \quad (4.5) \quad \frac{HV_r(PF_f) - HV_r(\mathcal{S}_n)}{HV_r(PF_f)} \leq \left(\frac{1}{2}\right)^{\lfloor \log_2(n+1) \rfloor} \leq \frac{1}{2n+2} .$$

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

598 *Proof.* The k -th term of the sequence defined by $u_0 = 1$ and $u_{n+1} = 2 \times u_n + 1$ for
 599 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
 600 convex, the normalized optimality gap at iteration $2^k - 1$ is inferior to $(1 - C)^k$ with C
 601 equal to $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}}$ and $\frac{1}{2}$, respectively. Since the hypervolume of the greedy set increases
 602 with n , and thus the optimality gap decreases with n , we deduce the first inequalities in
 603 (4.4) and (4.5) via the change of variable $k = \lfloor \log_2(n+1) \rfloor$.

604 Additionally, for every n , $\lfloor \log_2(n+1) \rfloor$ is smaller than $\log_2(n+1) + 1$, that is $\log_2(2n +$
 605 $2)$. For every C , $\log_2(2n + 2)$ equals $\log_C(2n + 2) \times \log_2(C)$, and thus $C^{\log_2(2n+2)}$ equals
 606 $(2n + 2)^{\log_2(C)}$. Therefore, we can infer that $(2n + 2)^{\log_2(C)}$ is an upper bound of the
 607 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}) -
 608 bilipschitz and convex, respectively. \square

609 We focus here on the relation between the optimality gap at iteration n and at
 610 iteration $2n + 1$. We could similarly examine the relation between the optimality gap at
 611 iteration n and at any later iteration. For example, we could prove that if f is (L_{\min}, L_{\max}) -
 612 bilipschitz, then for all n , for all $k \leq n + 1$, $\frac{HV_r(PF_f) - HV_r(\mathcal{S}_{n+k})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)}$ is lower than $1 - \frac{1}{2} \times$
 613 $\frac{L_{\min}}{L_{\max}} \times \frac{k}{n+1}$.

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

618 **4.2. Asymptotical behavior of the convergence of $\text{HV}_r(\mathcal{S}_n)$ to $\text{HV}_r(\text{PF}_f)$.**

619 In this section, we analyze the asymptotic convergence behavior for a Pareto front de-
 620 scribed by a bilipschitz function with a Hölder continuous derivative. We prove that, in
 621 this case, doubling the number of vectors in the greedy set divides the optimality gap by a
 622 factor which converges asymptotically to two as stated in [Theorem 4.10](#). This asymptotic
 623 limit corresponds to the case of affine Pareto fronts with a reference point dominating
 624 the nadir point. Indeed, for such Pareto fronts and reference points, the optimality gap
 625 is always halved when the number of vectors in the greedy set goes from n to $2n + 1$, see
 626 [Figure 5](#).

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

630 (4.6)
$$w_{i,r}^n := v_{\sigma_n(i),1} \text{ for } i \in \llbracket 1, n \rrbracket$$

631 (4.7)
$$w_{0,r}^n := \tilde{x}_{\min,r} \text{ and } w_{n+1,r}^n := \tilde{x}_{\max,r} .$$

633 Naturally, we have $w_{0,r}^n \leq w_{1,r}^n \leq \dots \leq w_{n+1,r}^n$, and the intervals $[w_{i-1,r}^n, w_{i,r}^n[$ for
 634 $i \in \llbracket 1, n + 1 \rrbracket$ form a partition of $[\tilde{x}_{\min,r}, \tilde{x}_{\max,r}[$, see the righthand plot of [Figure 4](#).
 635 The interval $[w_{i-1,r}^n, w_{i,r}^n[$ corresponds to the part of the Pareto front dominating r_i^n , the
 636 reference point associated to the i -th gap region of \mathcal{S}_n . When the Pareto front is bilip-
 637 schitz, the lengths of these intervals converge asymptotically to 0 as stated in the next
 638 lemma. It is a direct consequence of the convergence of $\text{HV}_r(\mathcal{S}_n)$ to $\text{HV}_r(\text{PF}_f)$ stated in
 639 [Theorem 4.2](#).

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

643 *Proof.* Let L_{\min} and L_{\max} be constants such that f is (L_{\min}, L_{\max}) -bilipschitz. The
 644 area of the i -th gap region of \mathcal{S}_n is $\int_{w_{i-1,r}^n}^{w_{i,r}^n} (f(x) - f(w_{i,r}^n))dx$. This is larger than
 645 $\int_{w_{i-1,r}^n}^{w_{i,r}^n} L_{\min} \times (w_{i,r}^n - x)dx$, which equals $\frac{1}{2} \times L_{\min} \times (w_{i,r}^n - w_{i-1,r}^n)^2$. Since the area
 646 of any gap region of \mathcal{S}_n is inferior to the optimality gap at iteration n , this implies that
 647 the difference $w_{i,r}^n - w_{i-1,r}^n$ is inferior to $\sqrt{2} \times (\text{HV}_r(\text{PF}_f) - \text{HV}_r(\mathcal{S}_n))$ for all n , for all
 648 $i \in \llbracket 1, n + 1 \rrbracket$. Therefore, the convergence of $\text{HV}_r(\mathcal{S}_n)$ to $\text{HV}_r(\text{PF}_f)$ stated in [Theorem 4.2](#)
 649 implies that the maximum over i of $w_{i,r}^n - w_{i-1,r}^n$ converges to 0. \square

650 We prove in the next lemma that if the Pareto front is described by a bilipschitz func-
 651 tion with a Hölder continuous derivative, then the restriction of this function associated
 652 to a gap region of \mathcal{S}_n is bilipschitz for some constants whose ratio converges asymptoti-
 653 cally to 1. Linear functions being the only functions to be (L_{\min}, L_{\max}) -bilipschitz with
 654 $L_{\min}/L_{\max} = 1$, it supports the interpretation that for such Pareto fronts, the conver-
 655 gence of the greedy set sequence is asymptotically similar as if they were affine Pareto
 656 fronts.

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

658 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

$$659 \quad (4.8) \quad \begin{aligned} L_{\min}^{i,n} &:= \inf \left\{ \left| \frac{f(x) - f(y)}{x - y} \right|, x, y \in [w_{i-1,r}^n, w_{i,r}^n], x \neq y \right\} \text{ and} \\ L_{\max}^{i,n} &:= \sup \left\{ \left| \frac{f(x) - f(y)}{x - y} \right|, x, y \in [w_{i-1,r}^n, w_{i,r}^n], x \neq y \right\}. \end{aligned}$$

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

$$662 \quad (4.9) \quad q_n := \max \left\{ \frac{L_{\max}^{i,n}}{L_{\min}^{i,n}}, i \in \llbracket 1, n+1 \rrbracket : [w_{i-1,r}^n, w_{i,r}^n] \neq \emptyset \right\}.$$

663

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

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

670 *Proof.* We take α such that f' is Hölder continuous with exponent α , i.e f is $\mathcal{C}^{1,\alpha}$,
671 and $L_{\min}, L_{\max} > 0$ such that the function f describing the Pareto front is (L_{\min}, L_{\max}) -
672 bilipschitz. We recall that f is decreasing, and thus for all $x < y$, we have $f(x) - f(y) \geq 0$.
673 Since f is $\mathcal{C}^{1,\alpha}$ and therefore \mathcal{C}^1 , the Taylor formula with Lagrange remainder states that
674 for all $x < y$, there exists $\xi \in [x, y]$ such that $f(y) = f(x) + (y-x) \times f'(\xi)$. Since f is $\mathcal{C}^{1,\alpha}$,
675 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,
676 $\frac{f(y)-f(x)}{x-y}$ is lower than $-f'(x) + [f']_{\mathcal{C}^\alpha} \times (y-x)^\alpha$. We now restrict ourselves to x and
677 y belonging to the non-empty interval $[w_{i-1,r}^n, w_{i,r}^n]$. Our goal is to find an upper bound
678 depending on i but not on either x or y . Since f is $\mathcal{C}^{1,\alpha}$, the difference between $-f'(x)$
679 and $-f'(w_{i-1,r}^n)$ is lower than $[f']_{\mathcal{C}^\alpha} \times (x - w_{i-1,r}^n)^\alpha$, and thus $[f']_{\mathcal{C}^\alpha} \times (w_{i,r}^n - w_{i-1,r}^n)^\alpha$.
680 Additionally, the difference between x and y is lower than $w_{i,r}^n - w_{i-1,r}^n$. We conclude that
681 for $x, y \in [w_{i-1,r}^n, w_{i,r}^n]$, $\frac{f(y)-f(x)}{x-y}$ is lower than $-f'(w_{i-1,r}^n) + 2[f']_{\mathcal{C}^\alpha} \times (w_{i,r}^n - w_{i-1,r}^n)^\alpha$,
682 and thus so is $L_{\max}^{i,n}$ defined in (4.8).

683 Following the same approach, we can also infer that $L_{\max}^{i,n}$ defined in (4.8) is greater
684 than the symmetric quantity $-f'(w_{i-1,r}^n) - 2[f']_{\mathcal{C}^\alpha} \times (w_{i,r}^n - w_{i-1,r}^n)^\alpha$. The quantity
685 $-f'(w_{i-1,r}^n)$ is greater than L_{\min} and $(w_{i,r}^n - w_{i-1,r}^n)^\alpha$ is smaller than $\max_{i \in \llbracket 1, n+1 \rrbracket} (w_{i,r}^n -$
686 $w_{i-1,r}^n)^\alpha$. As a consequence, q_n is lower than $\frac{L_{\min} + 2[f']_{\mathcal{C}^\alpha} \times \max_{i \in \llbracket 1, n+1 \rrbracket} (w_{i,r}^n - w_{i-1,r}^n)^\alpha}{L_{\min} - 2[f']_{\mathcal{C}^\alpha} \times \max_{i \in \llbracket 1, n+1 \rrbracket} (w_{i,r}^n - w_{i-1,r}^n)^\alpha}$. By
687 **Lemma 4.4**, $\max_{i \in \llbracket 1, n+1 \rrbracket} w_{i,r}^n - w_{i-1,r}^n$ converges to 0 and thus, this upper bound on q_n
688 converges to 1. By definition, q_n is always higher than 1, and thus converges to 1. \square

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

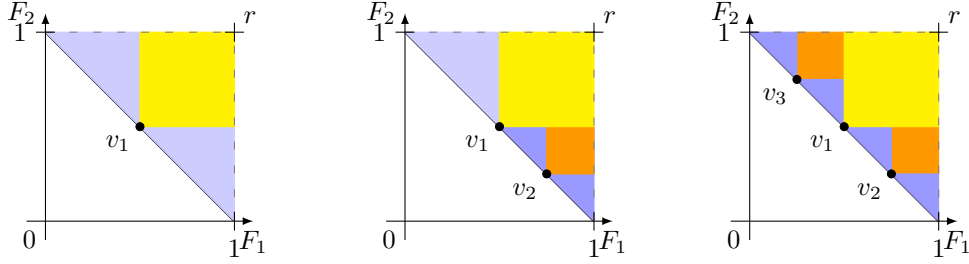


Fig. 5: The three greedy sets \mathcal{S}_1 (left), \mathcal{S}_2 (middle) and \mathcal{S}_3 (right) and their gap regions for an affine Pareto front with a reference point r dominating the nadir point $(1, 1)$. The area of the gap regions of \mathcal{S}_1 are half of $\text{HV}_r(\text{PF}_f)$ (left).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

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

699 (4.10)
$$\frac{1}{2} \times (1 - \epsilon) \leq \frac{\max_{u \in \mathcal{G}_{\mathcal{S}_n, i}^r} \text{HVI}_r(u, \mathcal{S}_n)}{\lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)} \leq \frac{1}{2} \times (1 + \epsilon) \text{ and}$$

700 (4.11)
$$\frac{1}{4 \times (1 + \epsilon)} \leq \frac{\lambda(\mathcal{G}_{\mathcal{S}_m, j}^r)}{\lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)} \leq \frac{1}{4 \times (1 - \epsilon)} .$$

701

702 *Proof.* The set of the first coordinates of the vectors of the Pareto front which domi-
703 nate r_i^n is the interval $[w_{i-1, r}^n, w_{i, r}^n]$. The restriction to $[w_{i-1, r}^n, w_{i, r}^n]$ of the function f
704 describing the Pareto front is (L_{\min}, L_{\max}) -bilipschitz for some L_{\min} and L_{\max} such that
705 $\frac{L_{\max}}{L_{\min}} = q_n$ with q_n defined in (4.9). Additionally, as stated in Proposition B.2, for n large
706 enough, all the r_i^n corresponding to non-empty gap regions dominate the nadir point.
707 It allows us to apply both Lemma A.2 and Proposition 3.11 to such gap regions. By
708 Propositions 3.10 and 3.11, $\frac{\max_{u \in \text{PF}_f} \text{HV}_{r_i^n}(u)}{\text{HV}_{r_i^n}(\text{PF}_f)}$ is between $\frac{1}{2} \times \frac{1}{q_n}$ and $\frac{1}{2} \times q_n$.

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

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

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

$$724 \quad (4.12) \quad \frac{\lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)}{\lambda(\mathcal{G}_{\mathcal{S}_n, j}^r)} \leq 4 \times \frac{(1 + \epsilon)^2}{1 - \epsilon}.$$

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

732 Take $n \geq N_2$. We note $\mathcal{G}_1 := \mathcal{G}_{\mathcal{S}_n, i}^r$ and $\mathcal{G}_2 := \mathcal{G}_{\mathcal{S}_n, j}^r$ two distinct non-empty gap regions
733 of \mathcal{S}_n , and \mathcal{P}_1 and \mathcal{P}_2 their respective parents. When two sets correspond to gap regions
734 relative to the same greedy set \mathcal{S}_m , we say that they cohabit at iteration m . Since only
735 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
736 and \mathcal{P}_2 or \mathcal{G}_2 and \mathcal{P}_1 cohabit at some earlier iteration. In the first case, there necessarily
737 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
738 to \mathcal{P}_2 , otherwise, \mathcal{G}_1 and \mathcal{G}_2 would not cohabit. By (4.10), the maximum hypervolume
739 improvement to \mathcal{S}_m of a vector of \mathcal{G}_1 and of a vector of \mathcal{P}_2 are at least $\frac{1}{2} \times (1 - \epsilon) \times \lambda(\mathcal{G}_1)$
740 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
741 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
742 $\lambda(\mathcal{P}_2)$ is lower than $4 \times (1 + \epsilon)$ times the area of its child $\lambda(\mathcal{G}_2)$ by (4.11), this inequality
743 implies (4.12). In the second case, \mathcal{P}_2 is filled before \mathcal{P}_1 . Thus, there exists $m \geq N_2$ such
744 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
745 higher than the one of its child \mathcal{G}_1 , the hypervolume improvement of v_{m+1} to \mathcal{S}_m is still
746 higher than $\frac{1}{2} \times (1 - \epsilon) \times \lambda(\mathcal{G}_1)$. The rest of the argumentation remains valid. \square

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

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

$$755 \quad (4.13) \quad \frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \leq \frac{1}{2} + o(\epsilon).$$

757 *Proof.* Fix $\epsilon > 0$. Fix n large enough to verify (4.10) and (4.12) for this particular ϵ .
758 Let σ be a permutation of $\llbracket 1, n+1 \rrbracket$ such that the i -th gap region of \mathcal{S}_n is filled by
759 $v_{n+\sigma(i)}$ when it is filled before iteration $2n + 1$. With this choice of permutation, $\mathcal{G}_{\mathcal{S}_n, i}^r$
760 is always a gap region of $\mathcal{S}_{n+\sigma(i)-1}$. Thus, $HVI_r(v_{n+\sigma(i)}, \mathcal{S}_{n+\sigma(i)-1})$ is superior to the
761 maximum hypervolume improvement of a vector of $\mathcal{G}_{\mathcal{S}_n, i}^r$ to $\mathcal{S}_{n+\sigma(i)-1}$, which is superior
762 to $\frac{1}{2} \times (1 - \epsilon) \times \lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)$ by (4.10). It is equivalent to say that the hypervolume improvement

763 at iteration $n + \sigma(i)$ is higher than $\frac{1}{2} \times (1 - \epsilon) \times \lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)$. Summing over $i \in \llbracket 1, n + 1 \rrbracket$,
764 we obtain that the hypervolume improvement between iteration n and $2n + 1$ is higher
765 than the sum over i of $\frac{1}{2} \times (1 - \epsilon) \times \lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)$, that is $\frac{1}{2} \times (1 - \epsilon)$ times the optimality gap
766 at iteration n .

767 Now, we need to bound the hypervolume improvement at iteration $2n + 1$, that is
768 $\text{HVI}_r(v_{2n+1}, \mathcal{S}_{2n})$. It is lower than $\frac{1}{2} \times (1 + \epsilon) \times \max_{i \in \llbracket 1, 2n+1 \rrbracket} \lambda(\mathcal{G}_{\mathcal{S}_{2n}, i}^r)$ by (3.4) and (4.10).
769 Since the area of a gap region is lower than the one of its parent, the maximum area of
770 a gap region is lower at iteration $2n$ than at iteration n . The maximum area of one of
771 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
772 gap at iteration n by (4.12). To summarize, the hypervolume improvement at iteration
773 $2n + 1$ is lower than $\frac{2 \times (1+\epsilon)^3}{(n-1) \times (1-\epsilon)}$ times the optimality gap at iteration n .

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

776 We broadly follow the same approach to obtain the following asymptotic lower bound on
777 the impact of doubling the number of points in the greedy set. Lemmas 4.6 and 4.7 are
778 key to prove a lower bound on the hypervolume improvement at iteration k . They allow
779 to prove that filling a gap region of \mathcal{S}_n more than once gives, up to a factor $1 + o(\epsilon)$, a
780 lower hypervolume improvement than filling a gap region which was not filled. Indeed,
781 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
782 parent by Lemma 4.6, which is itself at most $4 + o(\epsilon)$ times the area of any other gap
783 region of \mathcal{S}_n by Lemma 4.7.

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

$$787 \quad (4.14) \quad \frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \geq \frac{1}{2} + o(\epsilon).$$

789 *Proof.* Fix $\epsilon > 0$. Fix n large enough to verify (4.10), (4.11) and (4.12) for this
790 particular ϵ . Let $\delta \in \{-1, 0, 1\}$ be such that \mathcal{S}_n has $n + \delta$ non-empty gap regions. Let
791 $i_0 := 1$ when the left extreme gap region is empty and $i_0 := 0$ otherwise.

792 Let σ be a permutation of $\llbracket 1, n + \delta \rrbracket$ such that the i -th non-empty gap region of
793 \mathcal{S}_n , $\mathcal{G}_{\mathcal{S}_n, i_0+i}^r$, is filled by the vector $v_{n+\sigma(i)}$ when it is filled before iteration $2n + \delta$. We
794 distinguish two cases. In the first case, $v_{n+\sigma(i)}$ is the child of the i -th non-empty gap
795 region of \mathcal{S}_n , and consequently its hypervolume improvement to $\mathcal{S}_{n+\sigma(i)-1}$ is at most
796 $\frac{1}{2} \times (1 + \epsilon) \times \lambda(\mathcal{G}_{\mathcal{S}_n, i_0+i}^r)$ by (4.10). In the second case, $v_{n+\sigma(i)}$ belongs to $\mathcal{G}_{\mathcal{S}_n, i_0+j}^r$, the
797 j -th non-empty gap region of \mathcal{S}_n , with $j \neq i$ and, by definition of σ , fills a descendant of
798 this gap region not $\mathcal{G}_{\mathcal{S}_n, i_0+j}^r$ itself. By (4.10), the hypervolume improvement of $v_{n+\sigma(i)}$ to
799 $\mathcal{S}_{n+\sigma(i)-1}$ is still at most $\frac{1}{2} \times (1 + \epsilon)$ times the area of the gap region it fills. By (4.11),
800 the area of a descendant of $\mathcal{G}_{\mathcal{S}_n, i_0+j}^r$ is smaller than $\frac{1}{4 \times (1-\epsilon)}$ times the area of its ancestor.
801 By (4.12), we also know that the area of the i -th non-empty gap region of \mathcal{S}_n is at most
802 $4 \times \frac{(1+\epsilon)^2}{1-\epsilon}$ times the area of any other gap region of \mathcal{S}_n , in particular its i -th non-empty
803 gap region. We conclude that the hypervolume improvement of $v_{n+\sigma(i)}$ to $\mathcal{S}_{n+\sigma(i)-1}$ is
804 lower than $\frac{1}{2} \times \frac{(1+\epsilon)^3}{(1-\epsilon)^2} \times \lambda(\mathcal{G}_{\mathcal{S}_n, i_0+i}^r)$. To summarize, since $1 + \epsilon$ is lower than $\frac{(1+\epsilon)^3}{(1-\epsilon)^2}$, the
805 hypervolume improvement at any iteration $n + \sigma(i)$ is lower than $\frac{1}{2} \times \frac{(1+\epsilon)^3}{(1-\epsilon)^2} \times \lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)$.

806 Summing over $i \in \llbracket 1, n + \delta \rrbracket$, the hypervolume improvement from iteration n to $2n + \delta$
807 is lower than $\frac{1}{2} \times \frac{(1+\epsilon)^3}{(1-\epsilon)^2}$ times the sum over i of $\lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)$, that is the optimality gap at
808 iteration n .

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

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

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

$$818 \quad (4.15) \quad \frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \xrightarrow{n \rightarrow \infty} \frac{1}{2}.$$

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

824 **5. Conclusion.** We prove that HV-ISOOMOO algorithms coupled with a singleob-
825 jective optimizer converge in $O(1/n)$ on convex Pareto fronts and in $O(1/n^c)$ on bilipschitz
826 Pareto fronts with $c \leq 1$ depending on the bilipschitz constants where n is the number of
827 meta-iterations. Each meta-iteration corresponds to a singleobjective optimization run.
828 Both bounds are tight over the class of Pareto fronts and reference points considered. In-
829 deed, they are reached for affine Pareto fronts and reference points dominating the nadir
830 point. On convex Pareto fronts, the highest achievable convergence-speed is reached
831 [15]. It shows that greedily adding points maximizing the hypervolume contribution as in
832 HV-ISOOMOO algorithms is an effective way to quickly increase the hypervolume. Ad-
833 ditionally, we prove that for bilipschitz Pareto fronts with a Hölder continuous derivative,
834 asymptotically, doubling the number of meta-iterations halves the optimality gap. This
835 asymptotic behavior resembles what we would observe with an affine Pareto front and a
836 reference point dominating the nadir point. Beware that this does not guarantee a rate of
837 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
838 verify this property. The convergence rate on nonconvex Pareto fronts could theoretically
839 be slower than $\Theta(1/n)$, but not faster by [15].

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

845 We exploit the bounds on the normalized maximum hypervolume proven in [Subsec-](#)
846 [tion 3.3](#) and the following lower and upper bounds on the relation between $\lambda(\mathcal{G}_{\text{left}}^{u^*})$ and
847 $\lambda(\mathcal{G}_{\text{right}}^{u^*})$.

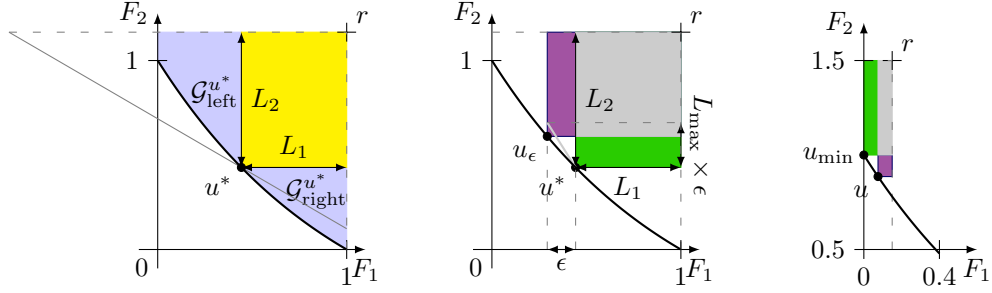


Fig. 6: Illustration of elements of the proofs of Proposition A.1 in the case $r_1 \leq x_{\max}$ (left and middle) and of Lemma B.1 (right). The Pareto front is described by $f(x) = \frac{e}{e-1} \times e^{-x} + 1 - \frac{e}{e-1}$ for $x \in [0, 1]$. Left: the gap regions $\mathcal{G}_{\text{left}}^{u^*}$ and $\mathcal{G}_{\text{right}}^{u^*}$ 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 .

848 PROPOSITION A.1. We assume that the Pareto front is described by a (L_{\min}, L_{\max}) -
849 bilipschitz function f . Let u^* be a non-extreme vector of the Pareto front which maximizes
850 the hypervolume with respect to a valid reference point r . If $r_1 \leq x_{\max}$, we have $\lambda(\mathcal{G}_{\text{right}}^{u^*}) \geq$
851 $\frac{L_{\min}^2}{L_{\max}^2} \times \lambda(\mathcal{G}_{\text{left}}^{u^*})$. If $r_2 \leq f(x_{\min})$, we have $\lambda(\mathcal{G}_{\text{left}}^{u^*}) \geq \frac{L_{\min}^2}{L_{\max}^2} \times \lambda(\mathcal{G}_{\text{right}}^{u^*})$.

852 *Proof.* We consider the case where $r_1 \leq x_{\max}$. We note $L_1 := r_1 - x^*$ and $L_2 :=$
853 $r_2 - f(x^*)$ the lengths of the sides of the rectangle $\mathcal{D}_{u^*}^r$. For all $x, y \in [x_{\min}, x_{\max}]$, we have
854 $|f(x) - f(y)| \geq L_{\min} \times |x - y|$. Additionally, since $r_1 \leq x_{\max}$, the segment $[x^*, x^* + L_1]$ is
855 included in $[x_{\min}, x_{\max}]$. As a consequence, the section of the Pareto front on the right of
856 u^* dominates the segment between u^* and $u^* + L_1 \times (1, -L_{\min})$, see the lefthand plot of
857 Figure 6. Therefore, $\lambda(\mathcal{G}_{\text{right}}^{u^*})$ is larger than the area of the region of the objective space
858 dominated by this segment, not dominated by u^* and dominating r , that is $\frac{1}{2} \times L_{\min} \times L_1^2$.
859 For all $x, y \in [x_{\min}, x_{\max}]$, we also have $|f(x) - f(y)| \leq L_{\max} \times |x - y|$. Therefore, the
860 part of the Pareto front on the left of u^* is dominated by the segment between u^* and
861 $u^* + L_2 \times (-\frac{1}{L_{\min}}, 1)$, and $\lambda(\mathcal{G}_{\text{left}}^{u^*})$ is lower than $\frac{1}{2} \times \frac{1}{L_{\min}} \times L_2^2$. We have yet to prove a
862 lower bound on $\frac{L_1}{L_2}$. The vector u^* being different from u_{\min} , for $\epsilon > 0$ small enough, the
863 vector $u_\epsilon := (x^* - \epsilon, f(x^* - \epsilon))$ belongs to the Pareto front. As we can see in the middle
864 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
865 $\epsilon \times (L_2 - \epsilon \times L_{\max})$. Additionally, u^* being an hypervolume maximizer, $\text{HVI}_r(u^*, u_\epsilon)$ is
866 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
867 the limit of this inequality when $\epsilon \rightarrow 0$, we obtain that $L_1 \times L_{\max} \geq L_2$. Combining
868 the bounds on $\lambda(\mathcal{G}_{\text{left}}^{u^*})$ and $\lambda(\mathcal{G}_{\text{right}}^{u^*})$ with the lower-bound on $\frac{L_1}{L_2}$, we obtain the desired
869 lower bound on $\lambda(\mathcal{G}_{\text{right}}^{u^*})$. We can obtain the symmetric inequality when $r_2 \geq f(x_{\min})$ by
870 following the same approach. \square

871 In particular, when f is bilipschitz and r dominates the nadir point, both bounds hold.
872 We are now able to prove the desired bounds on the normalized area of the gap regions
873 $\mathcal{G}_{\text{left}}^{u^*}$ and $\mathcal{G}_{\text{right}}^{u^*}$.

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

878 *Proof.* Let note arbitrarily \mathcal{G}_1 and \mathcal{G}_2 the two gap regions of the set $S = \{u^*\}$.
879 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,
880 Propositions 3.10 and 3.11, the normalized maximum hypervolume $\max_{u \in \text{PF}_f} \text{HV}_r(u)$
881 over $\text{HV}_r(\text{PF}_f)$ 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
882 into bounds on $\text{HV}_r(\text{PF}_f) - \max_{u \in \text{PF}_f} \text{HV}_r(u)$, that is $\lambda(\mathcal{G}_1) + \lambda(\mathcal{G}_2)$. As a consequence,
883 $\lambda(\mathcal{G}_1)$ is between $(1 - \frac{1}{2} \times \frac{L_{\max}}{L_{\min}}) \times \text{HV}_r(\text{PF}_f) - \frac{L_{\max}^2}{L_{\min}^2} \times \lambda(\mathcal{G}_1)$ and $(1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}}) \times \text{HV}_r(\text{PF}_f) -$
884 $\frac{L_{\min}^2}{L_{\max}^2} \lambda(\mathcal{G}_1)$. Moving all the $\lambda(\mathcal{G}_1)$ terms on the same side and re-normalizing this side, we
885 obtain the desired bounds for \mathcal{G}_1 , which can be chosen to be either $\mathcal{G}_{\text{left}}^{u^*}$ or $\mathcal{G}_{\text{right}}^{u^*}$. \square

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

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

897 LEMMA B.1. We assume that the Pareto front is described by a function f which is
898 (L_{\min}, L_{\max}) -bilipschitz. For any valid reference point $r = (r_1, r_2)$ such that $r_1 > x_{\max}$ and
899 $f(x_{\max}) < r_2 < f(x_{\max}) + L_{\min} \times (r_1 - x_{\max})$, the right extreme of the Pareto front u_{\max} is
900 the only maximizer of $\text{HV}_r(\cdot)$. Additionally, for any valid reference point $r = (r_1, r_2)$ such
901 that $r_2 > f(x_{\min})$ and $x_{\min} < r_1 < x_{\min} + \frac{r_2 - f(x_{\min})}{L_{\max}}$, the vector $u_{\min} = (x_{\min}, f(x_{\min}))$
902 is the only maximizer of $\text{HV}_r(\cdot)$.

903 *Proof.* This proof is illustrated in the righthand plot of Figure 6. Let r be a reference
904 point such that $r_2 > f(x_{\min})$ and $x_{\min} < r_1 < x_{\min} + \frac{r_2 - f(x_{\min})}{L_{\max}}$. Let $u = (x, f(x)) \neq u_{\min}$
905 be a vector of the Pareto front which dominates r . The hypervolume improvement of u_{\min}
906 to $\{u\}$ is $(r_2 - f(x_{\min})) \times (x - x_{\min})$. The hypervolume improvement of u to $\{u_{\min}\}$ is equal
907 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
908 dominates r and f is (L_{\min}, L_{\max}) -bilipschitz. Since we assume that $L_{\max} \times (r_1 - x_{\min}) <$
909 $r_2 - f(x_{\min})$, the upper bound on $\text{HVI}_r(u, u_{\min})$ is strictly lower than $\text{HVI}_r(u_{\min}, u)$. As
910 a consequence, the hypervolume of u_{\min} is strictly higher than the one of u . We conclude
911 that u_{\min} is the unique hypervolume maximizer. The symmetric result can be obtained
912 with the same approach. \square

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

916 PROPOSITION B.2. We assume that the the Pareto front is described by a bilipschitz
 917 function. Let $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ be a greedy set sequence associated to a valid reference point r .
 918 For n large enough, every reference point r_i^n corresponding to a non-empty gap region
 919 $\mathcal{G}_{\mathcal{S}_n, i}^r$ dominates the nadir point.

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

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

935

REFERENCES

- 936 [1] A. AUGER, J. BADER, AND D. BROCKHOFF, *Theoretically Investigating Optimal μ -Distributions*
 937 *for the Hypervolume Indicator: First Results for Three Objectives*, in Parallel Problem Solving
 938 from Nature, PPSN XI, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C.
 939 Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos,
 940 D. Tygar, M. Y. Vardi, G. Weikum, R. Schaefer, C. Cotta, J. Kolodziej, and G. Rudolph, eds.,
 941 vol. 6238, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 586–596, [https://doi.org/](https://doi.org/10.1007/978-3-642-15844-5_59)
 942 [10.1007/978-3-642-15844-5_59](https://doi.org/10.1007/978-3-642-15844-5_59).
 943 [2] A. AUGER, J. BADER, D. BROCKHOFF, AND E. ZITZLER, *Theory of the Hypervolume Indicator:*
 944 *Optimal μ -Distributions and the Choice of the Reference Point*, in Proceedings of the Tenth
 945 ACM SIGEVO Workshop on Foundations of Genetic Algorithms, FOGA '09, New York, NY,
 946 USA, Jan. 2009, Association for Computing Machinery, pp. 87–102, [https://doi.org/10.1145/](https://doi.org/10.1145/1527125.1527138)
 947 [1527125.1527138](https://doi.org/10.1145/1527125.1527138).
 948 [3] A. AUGER, J. BADER, D. BROCKHOFF, AND E. ZITZLER, *Hypervolume-based multiobjective opti-*
 949 *mization: Theoretical foundations and practical implications*, Theoretical Computer Science,
 950 425 (2012), pp. 75–103, <https://doi.org/10.1016/j.tcs.2011.03.012>.
 951 [4] N. BEUME, M. LAUMANN, AND G. RUDOLPH, *Convergence Rates of (1+1) Evolutionary Multiobjec-*
 952 *tive Optimization Algorithms*, in Parallel Problem Solving from Nature, PPSN XI, R. Schaefer,
 953 C. Cotta, J. Kolodziej, and G. Rudolph, eds., Lecture Notes in Computer Science, Berlin,
 954 Heidelberg, 2010, Springer, pp. 597–606, https://doi.org/10.1007/978-3-642-15844-5_60.
 955 [5] N. BEUME, B. NAUJOKS, AND M. EMMERICH, *SMS-EMOA: Multiobjective selection based on dom-*
 956 *inated hypervolume*, European Journal of Operational Research, 181 (2007), pp. 1653–1669,
 957 <https://doi.org/10.1016/j.ejor.2006.08.008>.
 958 [6] J. BIGEON, S. L. DIGABEL, AND L. SALOMON, *DMulti-MADS: Mesh adaptive direct multisearch for*
 959 *blackbox multiobjective optimization*, p. 30.
 960 [7] K. BRINGMANN AND T. FRIEDRICH, *The maximum hypervolume set yields near-optimal approxima-*
 961 *tion*, in Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation
 962 - GECCO '10, Portland, Oregon, USA, 2010, ACM Press, p. 511, [https://doi.org/10.1145/](https://doi.org/10.1145/1830483.1830576)
 963 [1830483.1830576](https://doi.org/10.1145/1830483.1830576).
 964 [8] D. BROCKHOFF, *Optimal μ -Distributions for the Hypervolume Indicator for Problems with Lin-*
 965 *ear Bi-objective Fronts: Exact and Exhaustive Results*, in Simulated Evolution and Learning,
 966 K. Deb, A. Bhattacharya, N. Chakraborti, P. Chakraborty, S. Das, J. Dutta, S. K. Gupta,

- 967 A. Jain, V. Aggarwal, J. Branke, S. J. Louis, and K. C. Tan, eds., vol. 6457, Springer Berlin
968 Heidelberg, Berlin, Heidelberg, 2010, pp. 24–34, https://doi.org/10.1007/978-3-642-17298-4_2.
- 969 [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>.
- 970
971
- 972 [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>.
- 973
974
- 975 [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>.
- 976
977
- 978 [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>.
- 979
980
- 981 [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>.
- 982
983
- 984 [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>.
- 985
986
- 987 [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.
- 988
989
- 990 [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>.
- 991
992
- 993 [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>.
- 994
995
- 996 [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.
- 997
998
- 999 [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.
- 1000
1001