



HAL
open science

Biobjective Hypervolume Based HV-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. Biobjective Hypervolume Based HV-ISOOMOO Algorithms Converge with At Least Sublinear Speed to the Entire Pareto Front. 2022. hal-03198414v3

HAL Id: hal-03198414

<https://hal.science/hal-03198414v3>

Preprint submitted on 22 Apr 2022

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 **BIOBJECTIVE HYPERVOLUME BASED HV-ISOOMOO ALGORITHMS**
2 **CONVERGE WITH AT LEAST SUBLINEAR SPEED TO THE ENTIRE**
3 **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 based on the hypervolume and
8 called HyperVolume based Incremental Single-Objective Optimization for MultiObjective Optimization
9 (HV-ISOOMOO) for approximating the Pareto front with an increasing number of points. The hypervolume
10 indicator is a set-quality indicator widely used for algorithms design and performance assessment.
11 The class of HV-ISOOMOO algorithms approximate the Pareto front by greedily maximizing this indicator.
12 At each meta-iteration of HV-ISOOMOO algorithms, a single-objective subproblem is solved.
13 We study the convergence to the entire Pareto front of HV-ISOOMOO under the assumption that these
14 subproblems are solved perfectly. The convergence is defined as the convergence of the hypervolume
15 of the sets of all meta-iterations incumbents towards the hypervolume of the Pareto front. We prove
16 tight lower bounds on the speed of convergence for convex and bilipschitz Pareto fronts in $O(1/n^c)$ with
17 $c = 1$ and $c \leq 1$, respectively. The index n denotes the number of meta-iterations of HV-ISOOMOO.
18 For convex Pareto fronts, the convergence speed is in $\Theta(1/n)$, namely the fastest convergence achievable
19 by a biobjective optimization algorithm. These are the first results on the speed of convergence of multi-
20 objective optimization algorithms towards the entire Pareto front. We also analyze theoretically the
21 asymptotic convergence behavior.

22 **Key words.** multiobjective optimization, convergence, hypervolume, Pareto front

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

24 **1. Introduction.** Real-world problems often involve the simultaneous optimization
25 of several conflicting objectives. The solution of such problems is the set of non-dominated
26 *decision vectors* (vectors of the search space), the *Pareto set*. It is defined as the set of
27 solutions that cannot be improved along one objective without degrading along another
28 one. Its image in the objective space is the Pareto front. A decision maker is then often
29 involved to choose, based on its preferences, a single best compromise. The shape of the
30 Pareto front informs him on the trade-off between objectives. Many algorithms such as
31 evolutionary algorithms approximate the Pareto front with a number of points fixed in
32 the beginning. But some algorithms, in particular stemming from direct search methods
33 [1, 7, 10, 11] approximate the Pareto set or the Pareto front with more and more points
34 during the run. Ideally, the quality of the Pareto front approximation (e.g. measured with
35 its size) increases gradually with time.

36 The speed of convergence towards a critical decision vector or a vector of the Pareto
37 front has been examined for many algorithms such as $(1 + 1)$ evolutionary multiobjective
38 algorithms [5] or Newton's method [18]. Convergence speeds are typically similar to
39 the ones obtained for single-objective optimization. Both aforementioned publications
40 describe convergence towards a single point and the first analysis is even reduced to the
41 study of the convergence of a single-objective optimization algorithm [5]. The convergence
42 of algorithms towards the whole Pareto set or front is of a different kind because iterates
43 are sets and not points. Such convergence has already been theoretically investigated
44 for some algorithms [10] and more abstract frameworks [26], but analysis of the speed
45 of convergence are generally missing. Empirical studies typically focus on determining
46 which algorithm is faster but more rarely provide information on the speed of convergence

47 such as order of convergence or complexity. Yet, while often overlooked, the study of the
48 speed of convergence, both theoretically and empirically, is important. In this context,
49 it has been proven that convergence of biobjective optimization algorithms towards the
50 whole Pareto front is always sublinear in the number of function evaluations, at least
51 when measuring convergence with the hypervolume indicator [22] or the multiplicative
52 ϵ -indicator [8], and thus much lower than typical speeds of convergence to a single point.
53 The hypervolume and the multiplicative ϵ -indicator are set-quality indicators widely used
54 in multiobjective optimization, both to guide algorithms and for performance assessment.
55 The hypervolume is at the core of all known strictly Pareto-compliant indicators [28].

56 In this paper, we introduce a new algorithm framework: HyperVolume based Incre-
57 mental Single-Objective Optimization for MultiObjective Optimization (HV-ISOOMOO).
58 Algorithms following this framework try to greedily maximize the hypervolume by adding
59 points approximating the largest hypervolume contribution achievable by a feasible vector.
60 Such points are obtained by running a single-objective solver. Relying on single-objective
61 optimization is a traditional approach in multiobjective optimization, the most significant
62 example being scalarization [16]. A greedy approach to finding a set of n points with a
63 large hypervolume is already used in the selection part of some multiobjective optimiza-
64 tion evolutionary algorithms such as SMS-EMOA [6]. The hypervolume of such discrete
65 greedy approximation is proven to be at least $(e - 1)/e$ times the one of a n -optimal
66 distribution [23]. To the best of our knowledge, we provide the first continuous equivalent
67 of this result.

68 The HV-ISOOMOO framework shares some similarities with a recent hyperboxing
69 algorithm [13]. At each meta-iteration, this hyperboxing algorithm finds a new point by
70 solving the Pascoletti-Serafini scalarization problem defined by the upper corner and the
71 diagonal of a box. The choice of the box relates to minimizing the ϵ -additive indicator.
72 In contrast, HV-ISOOMOO algorithms are built to minimize the hypervolume.

73 We analyze an ideal version of HV-ISOOMOO algorithms where the single-objective
74 solver returns a global optimum of the single-objective subproblems. This analysis is rel-
75 evant for practical HV-ISOOMOO algorithms (whose construction is left for future publi-
76 cations) when the single-objective solver returns good approximations¹ of global optima.
77 We investigate the speed of convergence of the ideal version of HV-ISOOMOO towards
78 the whole Pareto front when measuring the convergence with the hypervolume. For con-
79 vex and bilipschitz Pareto fronts, we prove that the convergence speed is in $O(1/n^c)$ with
80 $c = 1$ and $c \leq 1$, respectively, with n being the number of single-objective optimization
81 runs performed. For convex Pareto fronts, the convergence is exactly in $\Theta(1/n)$ as no
82 biobjective algorithm can converge faster to the Pareto front [22]. Additionally, we prove
83 that for simultaneously bilipschitz and smooth enough Pareto fronts doubling the number
84 of points in the approximation divides the optimality gap by a factor which converges
85 asymptotically to two. In the proof process, we obtain interesting intermediary results
86 such as bounds on the normalized maximum hypervolume and a geometric interpretation
87 of optimality conditions.

88 The paper is organized as follows. In [Section 2](#), we recall some classic concepts and
89 define our assumptions. In [Section 3](#), we introduce the HV-ISOOMOO framework to-
90 gether with connected mathematical abstractions and present numerical results on its

¹Good approximations are typically unavailable when the single-objective solver does not converge to a solution, stops too early or converges to a local optimum.

91 convergence rate. In [Section 4](#), we prove lower bounds on the so-called *normalized max-*
 92 *imum hypervolume*, later used to investigate convergence. In [Section 5](#), we derive lower
 93 bounds on the speed of convergence of HV-ISOOMOO algorithms under the assumption
 94 that every single-objective subproblem is solved perfectly and a theoretical result giving
 95 an insight on its asymptotic convergence behavior. For readability of the proof of this
 96 insight, we put some intermediary results and their proofs in the [Appendices A and B](#).

97 *Notations and conventions.* For $a, b \in \mathbb{N}$ with $a < b$, we note $\llbracket a; b \rrbracket$ the set $\{a, a +$
 98 $1, \dots, b - 1, b\}$. For a vector $u \in \mathbb{R}^2$, we note u_1 and u_2 respectively its first and its second
 99 coordinate. If the vector notation already contains an index, we separate the two indices
 100 with a comma. For simplicity sake, we often replace the set $\{u\}$ by u in the notations. We
 101 say that a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is decreasing (resp. strictly decreasing) when for all $x < y$,
 102 we have $f(x) \geq f(y)$ (resp. $f(x) > f(y)$). We only consider two-dimensional objective
 103 spaces and refer to the Lebesgue measure of a set as its area.

104 **2. Background and assumptions.** In this section, we recall some classic concepts
 105 of multiobjective optimization and define the assumptions used in the paper.

106 **2.1. Biobjective optimization problems, the Pareto front and the hyper-**
 107 **volume indicator.** We consider a biobjective minimization problem:

$$108 \quad (2.1) \quad \min_{X \in \Omega \subset \mathbb{R}^d} F(X)$$

110 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
 111 vectors in the objective space. We say that u weakly dominates v denoted by $u \preceq v$ if
 112 $u_1 \leq v_1$ and $u_2 \leq v_2$ and that u dominates v denoted by $u \prec v$ if $u \preceq v$ and $u \neq v$. A
 113 vector of the objective space \mathbb{R}^2 is said *feasible* when it belongs to $F(\Omega)$. Solving the
 114 optimization problem consists in finding a good approximation of the *Pareto front*, the
 115 set of non-dominated feasible vectors, $\{F(X) : X \in \Omega, \forall Y \in \Omega, F(Y) \not\prec F(X)\}$. We
 116 restrict ourselves to Pareto fronts with an explicit representation:

$$117 \quad (2.2) \quad \text{PF}_f = \{(x, f(x)) : x \in [x_{\min}, x_{\max}]\}$$

119 with $f : \mathbb{R} \mapsto \mathbb{R}$ decreasing. When both objective functions have and reach finite global
 120 minimum values in the search space, that we denote respectively v_1 and v_2 , we have
 121 $x_{\min} := v_1$ and $x_{\max} := \min_{X \in \Omega: F_2(X)=v_2} F_1(X)$.

122 We denote by $u_{\min} := (x_{\min}, f(x_{\min}))$ and $u_{\max} := (x_{\max}, f(x_{\max}))$ the extreme
 123 vectors of the Pareto front. Likewise, we denote by $\tilde{u}_{\min, r} := (\tilde{x}_{\min, r}, f(\tilde{x}_{\min, r}))$ and
 124 $\tilde{u}_{\max, r} := (\tilde{x}_{\max, r}, f(\tilde{x}_{\max, r}))$ the extremes vectors of the part of the Pareto front domi-
 125 nating a reference point r , with $\tilde{x}_{\min, r} := \max(x_{\min}, f^{-1}(r_2))$ and $\tilde{x}_{\max, r} := \min(x_{\max}, r_1)$.
 126 The vector $(x_{\max}, f(x_{\min}))$ is called the *nadir point*. All these notations are illustrated in
 127 [Figure 1](#).

128 The *hypervolume* with respect to a reference point r of a set S of objective vectors
 129 is the Lebesgue measure of the region of the objective space dominated by S and strictly
 130 dominating the reference point r . We denote it $\text{HV}_r(S)$. When no vector of the Pareto
 131 front dominates the reference point r , $\text{HV}_r(S) = 0$ for any set S of feasible points of the
 132 objective space. Since this particular case is not interesting, we only consider reference
 133 points dominated by at least one vector of the Pareto front from now on. We refer to such
 134 reference points as *valid*. When there are more than two points in the Pareto front, any

135 $(F_1(X_2), F_2(X_1))$ with $X_1 \in \arg \min_{X \in \Omega} F_1$ and $X_2 \in \arg \min_{X \in \Omega} F_2$ is a valid reference
 136 point.

137 The region of the objective space weakly dominated by S and dominating r (see the
 138 righthand plot of [Figure 1](#)) is denoted by \mathcal{D}_S^r and formally defined as:

$$139 \quad (2.3) \quad \mathcal{D}_S^r = \{w \in \mathbb{R}^2 : \exists u \in S : u \preceq w \prec r\} .$$

141 The hypervolume of a set S relative to the reference point r equals $\lambda(\mathcal{D}_S^r)$ with $\lambda(\cdot)$ being
 142 the Lebesgue measure. The set \mathcal{D}_S^r is the union of the \mathcal{D}_u^r for $u \in S$, \mathcal{D}_u^r being the
 143 rectangle $[u_1, r_1] \times [u_2, r_2]$ when u dominates r and \emptyset otherwise, see the righthand plot of
 144 [Figure 1](#). Note that the \mathcal{D}_u^r are not disjoint.

145 We use the hypervolume to characterize the convergence of a set S of objective vectors
 146 to the entire Pareto front. For a fixed valid reference point r , a set S is said to converge
 147 to the Pareto front when the hypervolume difference $\text{HV}_r(\text{PF}_f) - \text{HV}_r(S)$ converges
 148 to 0. We define the *optimality gap* of S with respect to a valid reference point r as
 149 $\text{HV}_r(\text{PF}_f) - \text{HV}_r(S)$. Another quantity of interest is how much adding a vector to a set
 150 affects its hypervolume. The *hypervolume improvement* with respect to r of the vector u
 151 to the set S is $\text{HVI}_r(u, S) = \text{HV}_r(S \cup \{u\}) - \text{HV}_r(S)$. We also use the term hypervolume
 152 improvement to refer to the hypervolume increase of an increasing² sequence of sets
 153 $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ at iteration n , $\text{HV}_r(\mathcal{S}_{n+1}) - \text{HV}_r(\mathcal{S}_n)$.

154 **2.2. Decomposition of the optimality gap using gap regions.** The optimality
 155 gap is the Lebesgue measure of the *total gap region* introduced below.

156 **DEFINITION 2.1.** *The total gap region of S with respect to a fixed valid reference*
 157 *point r , \mathcal{G}_S^r , is defined as the region of the objective space which dominates r and is*
 158 *weakly dominated by PF_f but not by S , namely $\mathcal{D}_{\text{PF}_f}^r \setminus \mathcal{D}_S^r$.*

159 We introduced \mathcal{D}_S^r in (2.3). Its Lebesgue measure is $\text{HV}_r(S)$.

160 The concept of *total gap region* is strongly connected to the concept of *search region*
 161 [20]. They can both be seen as the part of the objective space which may contain feasible
 162 nondominated points, but only the total gap region relies on the knowledge of the Pareto
 163 front. The total gap region is composed of the vectors of the search region which are
 164 weakly dominated by the Pareto front, a condition satisfied by all feasible objective-
 165 vectors.

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

170 **DEFINITION 2.2 (Gap regions and local nadir points).** *Let $S = \{v_1, \dots, v_n\}$ be a set*
 171 *of n distinct vectors of the Pareto front dominating a valid reference point r . Let σ be the*
 172 *permutation ordering the v_i by increasing F_1 -values: $v_{\sigma(1),1} < v_{\sigma(2),1} < \dots < v_{\sigma(n),1}$.*

- 173 • *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 $\mathcal{D}_{S'}^r$ with*
 174 *the associated local nadir point r_i^S being $r_1^S = (v_{\sigma(1),1}, r_2)$, $r_{n+1}^S = (r_1, v_{\sigma(n),2})$*
 175 *and $r_i^S = (v_{\sigma(i),1}, v_{\sigma(i-1),2})$ for all $i \in \llbracket 2, n \rrbracket$.*
- 176 • *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 ,*
 177 *respectively.*

²A sequence of set $\{A_n, n \geq 0\}$ is increasing if the following inclusions $A_0 \subset A_1, \dots \subset A_n \subset \dots$ hold.

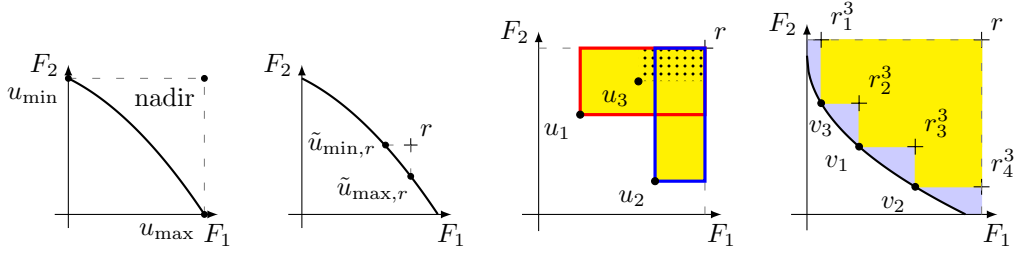


Fig. 1: Illustration of notations. The extreme vectors u_{\min} and u_{\max} and the nadir point (leftmost) ; the extreme vectors relative to the reference point r $\tilde{u}_{\min,r}$ and $\tilde{u}_{\max,r}$ (left) ; three vectors u_1 , u_2 and u_3 and the regions weakly dominated by them and dominating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

178 *Local nadir points* are also called *local upper bounds* [24, 20]. *Gap regions* are to search
 179 *zones* what total gap regions are to search regions. A gap region is composed of the
 180 vectors of a search zone which are weakly dominated by the Pareto front.

181 The left (resp. right) extreme gap region is empty when the left (resp. right) extreme
 182 vector of the Pareto front belongs to S . Non-extreme gap regions are never empty.

183 The total gap region of such a set S is the disjoint union of its gap regions: $\mathcal{G}_S^r =$
 184 $\dot{\cup}_{i=1}^{|S|+1} \mathcal{G}_{S,i}^r$. This decomposition of the total gap region, and thus of the optimality gap,
 185 is the cornerstone of our convergence analysis. Since the area of a gap region $\mathcal{G}_{S,i}^r$ is
 186 $\text{HV}_{r_i^S}(\text{PF}_f)$, we can write the optimality gap of a set S as the sum of $|S|+1$ hypervolumes
 187 of the Pareto front with respect to the local nadir points.

188 **LEMMA 2.3.** *Let S be a set of n distinct vectors of the Pareto front dominating a*
 189 *valid reference point r . The optimality gap of S with respect to a valid reference point r*
 190 *can be decomposed as*

$$191 \quad (2.4) \quad \text{HV}_r(\text{PF}_f) - \text{HV}_r(S) = \sum_{i=1}^{n+1} \text{HV}_{r_i^S}(\text{PF}_f) .$$

193 *Proof.* The optimality gap of S is the Lebesgue measure of the total gap region \mathcal{G}_S^r ,
 194 which is the disjoint union of the gap regions $\mathcal{G}_{S,i}^r = \mathcal{D}_{\text{PF}_f}^{r_i^S}$. Therefore, the optimality gap
 195 equals $\sum_{i=1}^{n+1} \lambda(\mathcal{D}_{\text{PF}_f}^{r_i^S}) = \sum_{i=1}^{n+1} \text{HV}_{r_i^S}(\text{PF}_f)$. \square

196 Additionally, we can express the hypervolume improvement of any vector to S as an
 197 hypervolume. It is a trivial assertion for vectors which do not dominate S . For other
 198 vectors, the reference point depends on the gap region to which the vector belongs.

199 **LEMMA 2.4.** *Let S be a set of n distinct vectors of the Pareto front dominating a*
 200 *valid reference point r . For any u belonging to the i -th gap region of S , $\mathcal{G}_{S,i}^r$, we have*

$$201 \quad (2.5) \quad \text{HVI}_r(u, S) = \text{HV}_{r_i^S}(u) .$$

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

205 **2.3. Assumptions on the Pareto front and the objective functions.** We
206 present and discuss here the assumptions on the function f describing the Pareto front
207 under which we derive convergence results. We typically assume that the function f is
208 bilipschitz, convex or simultaneously bilipschitz and with a Hölder continuous derivative.
209 Under any of these three assumptions, f is continuous. For the sake of conciseness, we
210 transfer the properties of f to the Pareto front. For example, we call *convex Pareto*
211 *front* a Pareto front described by a convex function. The ZDT test suite [27] contains
212 two biobjective optimization problems with a convex Pareto front: ZDT1 (see Figure 2)
213 and ZDT4. We recall that a function f is Hölder continuous with exponent α , namely
214 $\mathcal{C}^{1,\alpha}$, when there exists $H \geq 0$ such that $|f(x) - f(y)| \leq H \times |x - y|^\alpha$ for all x, y [17].
215 We note $[f]_\alpha$ the *minimum Hölder coefficient* with respect to the exponent α of a $\mathcal{C}^{1,\alpha}$
216 function f , that is $[f]_\alpha := \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}$. We recall that a function f is bilipschitz
217 if there exists L_{\min} and L_{\max} with $L_{\max} \geq L_{\min} > 0$ such that for all $x, y \in [x_{\min}, x_{\max}]$,
218 we have $L_{\min} \times |x - y| \leq |f(x) - f(y)| \leq L_{\max} \times |x - y|$. When needed, we detail
219 the bilipschitz constants and refer to f as (L_{\min}, L_{\max}) -bilipschitz. For example, the
220 problem of minimizing $F_1(X) = X^2$ and $F_2(X) = (X - 1)^2$ for $X \in [0.1, 0.9]$ has a
221 bilipschitz Pareto front. Its representation is the function $f : x \mapsto 1 - \sqrt{x}$ for $x \in$
222 $[F_1(0.1), F_1(0.9)] = [0.01, 0.81]$, which is (L_{\min}, L_{\max}) -bilipschitz with $L_{\min} := |f'(0.81)| =$
223 $0.555\dots$ and $L_{\max} := |f'(0.01)| = 5$. We also talk of *affine Pareto fronts* when $f(x) =$
224 $ax + b$ with $a < 0$ and $b \in \mathbb{R}$. As they form a line in the biobjective case, they are usually
225 referred to as *linear Pareto fronts*. They help to understand the results we prove on the
226 asymptotic convergence behavior. The biobjective optimization problem DTLZ1 of the
227 DTLZ test suite [15] has an affine Pareto front. Affine Pareto fronts are a special case of
228 (L_{\min}, L_{\max}) -bilipschitz Pareto fronts where $L_{\min} = L_{\max}$.

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

231 **PROPOSITION 2.5.** *Given a biobjective minimization problem as in (2.1) whose Pareto*
232 *front is described by a function f . If F_1 and F_2 are respectively $(L_{\min,1}, L_{\max,1})$ -bilipschitz*
233 *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. If the search space*
234 *Ω and the objective functions F_1 and F_2 are convex, then f is convex.*

235 The proofs of this proposition can be found for instance in [22]. The conditions on F_1 ,
236 F_2 and Ω are sufficient but non-necessary conditions. Indeed, adding small discontinuity in
237 the objective functions far from the Pareto set makes them non-convex and non-bilipschitz
238 without modifying the Pareto front.

239 Representing F_1 values on the absciss and F_2 values on the ordinate instead of the
240 converse is an arbitrary choice. When f is a bijection, if we chose to represent the F_2
241 values on the x -axis instead of on the y -axis, we would have another representation of
242 the Pareto front : $\{(y, f^{-1}(y)) : y \in [f(x_{\max}); f(x_{\min})]\}$. If so, the inverse function f^{-1}
243 would play the role of f . It is interesting to notice that the choice of the objective function
244 represented on the horizontal axis does not impact whether the function characterizing
245 the Pareto front is bilipschitz or convex. Indeed, f being bilipschitz is equivalent to both
246 f and f^{-1} being lipschitz. Additionally, given that the function f is decreasing, f being
247 convex is equivalent to its inverse f^{-1} being convex.

248 **3. The HV-ISOOMOO framework.** We introduce the HV-ISOOMOO frame-
249 work. We formalize its mathematical abstraction under the assumption that every single-

250 objective subproblem is solved perfectly, the *greedy set sequences*. Finally, we present
 251 numerical results on the rate of convergence of these greedy set sequences towards the
 252 Pareto front.

253 **3.1. Presentation of the framework.** The HV-ISOOMOO framework builds in-
 254 crementally an increasing sequence $(\mathcal{I}_n)_{n \in \mathbb{N}^*}$ of sets of vectors of the objective space. The
 255 pseudocode of HV-ISOOMOO is given in [Algorithm 3.1](#), where the current value of \mathcal{I}_n
 256 is denoted by \mathcal{I} . At each so-called meta-iteration, a single-objective maximization solver
 257 SOOPTIMIZER (line 2 in [Algorithm 3.1](#)) is run on the criterion $X \in \Omega \subset \mathbb{R}^d \mapsto J(\mathcal{I}, X)$
 258 and the resulting solution is added to \mathcal{I} (line 3 in [Algorithm 3.1](#)). We use the term meta-
 259 iteration to separate between the (meta-)iterations of HV-ISOOMOO and the iterations
 260 of SOOPTIMIZER. Since the set \mathcal{I} is composed of the final objective incumbents of previ-
 261 ous runs of SOOPTIMIZER and (ideally) provides an approximation of the Pareto front,
 262 we call it *final incumbents Pareto front approximation*.

263 The single-objective optimization procedure may vary between meta-iterations. More
 264 precisely, the run of SOOPTIMIZER depends on data about precedent runs stored in
 265 D (line 4 in [Algorithm 3.1](#)). This allows to alternate between various single-objective
 266 optimization solvers with different features, but also to adapt their initialization. This
 267 could be done by storing in D an iteration index and the final search space incumbents
 268 of SOOPTIMIZER runs.

Algorithm 3.1 HV-ISOOMOO Framework

```

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

```

269 The criterion J is chosen such that its maximization is *compliant* with the maxi-
 270 mization of the hypervolume improvement with respect to a reference point r as defined
 271 below.

272 *Assumption 3.1.* (Compliance to HVI_r maximization) The maximization of a crite-
 273 rion J as in HV-ISOOMOO is *compliant* with the maximization of HVI_r if for any set \mathcal{I}
 274 of objective vectors, we have

$$275 \quad (3.1) \quad \operatorname{argmax}_{X \in \mathbb{R}^d} J(\mathcal{I}, X) = \operatorname{argmax}_{X \in \mathbb{R}^d} \text{HVI}_r(F(X), \mathcal{I}) .$$

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

282 **DEFINITION 3.2.** We define the convergence of an HV-ISOOMOO algorithm as the
 283 convergence of $\text{HV}_r(\mathcal{I}_n)$ towards $\text{HV}_r(\text{PF}_f)$.

284 The performance of a specific HV-ISOOMOO algorithm depends crucially on the
 285 choice of the criterion J . In this respect, $\text{HVI}_r(\mathcal{I}, F(\cdot))$ itself is not a good candidate for

286 $J(\mathcal{I}, \cdot)$. Indeed, it is constant equal to zero in the region dominated by \mathcal{I} , which makes it
 287 difficult to optimize. A criterion whose maximization is compliant with the maximization
 288 of the hypervolume improvement and designed to be easier to optimize has already been
 289 introduced in [25] under the name *uncrowded hypervolume improvement* (UHVI). For
 290 $F(X)$ not dominated by \mathcal{I} , UHVI_r and HVI_r are equal. Otherwise, in the region where
 291 the hypervolume improvement is null, UHVI_r is negative and equals minus the distance
 292 to the empirical non-dominated front of the set \mathcal{I} relative to r . It is easy to see that
 293 UHVI_r satisfies (3.1).

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

297 *Assumption 3.3* (Perfect single-objective optimization). At every meta-iteration n ,
 298 for any final incumbents Pareto front approximation \mathcal{I}_n , the run of SOOPTIMIZER (line 2
 299 in Algorithm 3.1) returns $Y \in \arg\max_{X \in \Omega} J(\mathcal{I}_n, F(X))$.

300 The assumption of perfect single-objective optimization is reminiscent of the assump-
 301 tion of perfect line search which is common in the analysis of gradient based methods [14].
 302 Under this assumption, all choices of criteria verifying Assumption 3.1 are equivalent.

303 Convergence results under perfect conditions are useful to investigate the convergence
 304 (speed) we can expect with a given framework and in turn guide the construction of
 305 practical algorithms. A good theoretical convergence speed suggests that an approach is
 306 worth exploring while an unexplained gap between practical and ideal convergence speeds
 307 suggests that there is room for improvement. In our case, convergence rates under perfect
 308 conditions also give us an idea of what the convergence rates with respect to meta-
 309 iterations of practical HV-ISOOMOO algorithms will look like when SOOPTIMIZER
 310 returns good approximations of global optima.

311 **3.2. Greedy sets and greedy set sequences.** We introduce below mathematical
 312 abstractions of the HV-ISOOMOO framework under Assumption 3.3 of perfect single-
 313 objective optimization, *greedy sequences* and *greedy set sequences*.

314 DEFINITION 3.4 (Greedy sequence and greedy set sequence). *Given a valid reference*
 315 *point r , we define as greedy sequence relative to r , a sequence $(v_n)_{n \in \mathbb{N}^*}$ satisfying*

$$316 \quad (3.2) \quad v_1 \in \arg \max_{v \in F(\Omega)} HV_r(v) \text{ and}$$

$$317 \quad (3.3) \quad v_{n+1} \in \arg \max_{v \in F(\Omega)} HV_r(\{v_1, \dots, v_n, v\}) \text{ for all } n \geq 1 .$$

318
 319 *The greedy set sequence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ associated to the greedy sequence $(v_n)_{n \in \mathbb{N}^*}$ is composed*
 320 *of the greedy sets $\mathcal{S}_n := \{v_k, k \leq n\}$.*

321 When considering a greedy set \mathcal{S}_n , we denote the i -th gap region and the associated local
 322 nadir point defined in Definition 2.2 by \mathcal{G}_i^n and r_i^n , respectively.

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

326 The recurrence relation of the greedy sequence (3.3) is equivalent to v_{n+1} belonging
 327 to $\arg \max_{v \in F(\Omega)} \text{HVI}_r(v, \mathcal{S}_n)$ for all $n \geq 1$. It is immediate to see that under Assump-
 328 tion 3.3, the final incumbents generated by HV-ISOOMOO constitute a greedy sequence
 329 while the final incumbents Pareto front approximations form the associated greedy set

330 sequence $(\mathcal{I}_n)_{n \in \mathbb{N}^*}$. The indices n of both sequences correspond to HV-ISOOMOO meta-
 331 iterations. In this paper, we derive convergence results for greedy set sequences, which
 332 transfer to HV-ISOOMOO under [Assumption 3.3](#).

333 Since the hypervolume indicator associated to a valid reference point is strictly Pareto-
 334 compliant (see [\[21\]](#)), this sequence is composed of vectors of the Pareto front.

335 **PROPOSITION 3.5.** *If the Pareto front is described by a lower semi-continuous func-*
 336 *tion f , then any vector of a greedy sequence relative to a valid reference point r belongs to*
 337 *the Pareto front. Consequently, for such Pareto front and reference point and under [As-](#)*
 338 *sumption 3.3, all final incumbents Pareto front approximations \mathcal{I}_n of an HV-ISOOMOO*
 339 *algorithm relative to r are subsets of the Pareto front.*

340 *Proof.* Since for any valid reference point r , $HV_r(\cdot)$ is strictly Pareto-compliant [\[21\]](#),
 341 its maximum and the maximum of every function of the form $v \mapsto HV_r(S \cup \{v\})$ are
 342 non-dominated and belong to the Pareto front. Thus, in particular, a vector v verifying
 343 either [\(3.2\)](#) or [\(3.3\)](#) belongs to the Pareto front. \square

344 We can express the hypervolume improvement of a greedy set sequence at iteration
 345 $n + 1$, $HV_r(\mathcal{S}_{n+1}) - HV_r(\mathcal{S}_n)$, as the maximum of $n + 1$ hypervolume maximization
 346 problems.

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

$$349 \quad (3.4) \quad HV_r(\mathcal{S}_{n+1}) - HV_r(\mathcal{S}_n) = \max_{i \in \llbracket 1, n+1 \rrbracket} \max_{u \in PF_f} HV_{r_i^n}(u) .$$

351 *Proof.* The hypervolume improvement $HV_r(\mathcal{S}_{n+1}) - HV_r(\mathcal{S}_n)$ is the hypervolume
 352 improvement of v_{n+1} to \mathcal{S}_n , namely the highest hypervolume improvement of a vector
 353 $u \in PF_f$ to \mathcal{S}_n by [\(3.3\)](#) and [Proposition 3.5](#). We can reformulate the hypervolume
 354 improvement of any vector u to \mathcal{S}_n as $\max_{i \in \llbracket 1, n+1 \rrbracket} HV_{r_i^n}(u)$ by [Lemma 2.4](#) since gap
 355 regions are disjoint and $HV_{r_i^n}(\cdot)$ is null outside the i -th gap region of \mathcal{S}_n . \square

356 Similarly, the problem of maximizing the hypervolume improvement to a greedy set
 357 \mathcal{S}_n can be rewritten as the maximum of a finite number of hypervolume maximization
 358 problems.

359 **LEMMA 3.7.** *At any iteration n , the recurrence relation satisfied by v_{n+1} , i.e. [\(3.3\)](#),*
 360 *can be reformulated as*

$$361 \quad (3.5) \quad v_{n+1} \in \arg \max_{i \in \llbracket 1, n+1 \rrbracket} \max_{u \in PF_f} HV_{r_i^n}(u) .$$

363 *Proof.* It is a direct consequence of the fact that the hypervolume improvement of
 364 any vector u to \mathcal{S}_n is $\max_{i \in \llbracket 1, n+1 \rrbracket} HV_{r_i^n}(u)$, as stated in the proof of [Lemma 3.6](#), and
 365 that $v_{n+1} \in PF_f$ by [Proposition 3.5](#). \square

366 As a consequence, we can infer from [\[4, Theorem 1\]](#) that as soon as the Pareto front
 367 is lower semi-continuous, there exists a greedy sequence, and thus a greedy set sequence.

368 **PROPOSITION 3.8.** *If the Pareto front is described by a lower semi-continuous func-*
 369 *tion f , then there exists a greedy sequence $(v_n)_{n \in \mathbb{N}^*}$ relative to any valid reference point.*

370 *Proof.* If f is lower semi-continuous, then for any valid reference point r , the maxi-
 371 mum of $HV_r(\cdot)$ exists, see [\[4, Theorem 1\]](#). Therefore, there exists a vector verifying [\(3.2\)](#)

372 and the problem of maximizing the maximum of a finite number of hypervolume functions
 373 defined in (3.5) admits a solution. Since (3.3) and (3.5) are equivalent by Lemma 3.7, we
 374 can build a sequence $(v_n)_{n \in \mathbb{N}^*}$ verifying (3.2) and (3.3), namely a greedy sequence. \square

375 Yet, in general, there exists more than one greedy sequence, and thus greedy set sequence.
 376 For example, there are infinitely many greedy sequences associated to any affine Pareto
 377 front with a reference point dominating the nadir point. This statement relies on the fact
 378 that the unique maximizer of the hypervolume relative to a reference point r dominating
 379 the nadir point is the middle of the section of the Pareto front dominating r , see [3,
 380 Theorem 5]. As a consequence, the middle of the section of the Pareto front dominating r
 381 is the only candidate for v_1 but v_2 can be either at $1/4$ or at $3/4$ of this section. Similarly,
 382 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
 383 section of the Pareto front dominating r . For any n , we can find an iteration m such
 384 that v_m can be placed at 2^n different points, whatever the $m - 1$ first terms of the greedy
 385 sequence are.

386 **3.3. Numerical results.** In this section, we investigate empirically the convergence
 387 rate of HV-ISOOMOO algorithms with respect to meta-iterations on seven benchmark
 388 Pareto fronts (see Figure 2). We try to be as close as possible to the Assumption 3.3 of
 389 perfect single-objective optimization. We iteratively find a vector of the Pareto front that
 390 we estimate close to a global optimum of the single-objective optimization subproblem of
 391 maximizing the hypervolume improvement to the vectors found so far. More precisely,
 392 we estimate that this vector differs from a global optimum by less than 10^{-12} (measured
 393 in terms of objective function values). We assimilate these vectors to greedy vectors in
 394 the following.

395 We consider the six concrete Pareto fronts with an explicit representation: $\{(x, f(x)) :$
 396 $x \in [0, 1]\}$ considered in [22], plus an affine Pareto front (see Figure 2). The Pareto fronts
 397 **zdt1**, **zdt2** and **dtlz2** belong to the ZDT and DTLZ test suites [27, 15]. Four of the
 398 Pareto fronts examined are convex (**affine**, **convex-bil**, **doublesphere** and **zdt1**), while
 399 three are bilipschitz (**affine**, **convex-bil** and **concave-bil**). Two are neither convex
 400 nor bilipschitz (**dtlz2** and **zdt2**), and thus do not belong to the class of Pareto fronts
 401 investigated theoretically in this paper. We take the nadir point $(1, 1)$ as reference point.

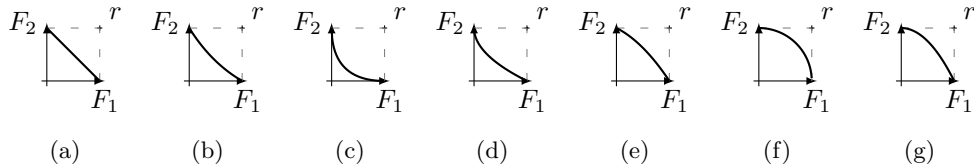


Fig. 2: The benchmark Pareto fronts and their representations (a): **affine** with $f : x \mapsto 1 - x$, (b): **convex-bil** with $f : x \mapsto \frac{e}{e-1} \times e^{-x} + 1 - \frac{e}{e-1}$, (c): **doublesphere** with $f : x \mapsto 1 + x - 2 \times \sqrt{x}$, (d): **zdt1** with $f : x \mapsto 1 - \sqrt{x}$, (e): **concave-bil** with $f : x \mapsto 1 - 0.5x - 0.5x^2$, (f): **dtlz2** with $f : x \mapsto \sqrt{1 - x^2}$ and (g): **zdt2** with $f : x \mapsto 1 - x^2$ for $x \in [0, 1]$. The reference point is $r = (1, 1)$.

402

403

Greedy vectors are defined as the true solutions of single-objective problems involving

404 the objective functions F_1 and F_2 in the search-space $\Omega \subset \mathbb{R}^d$. In order to compute an
 405 approximation of the greedy vectors, we exploit the explicit representation f of the Pareto
 406 front. Any greedy vector v_n belongs to the Pareto front (see [Proposition 3.5](#)), and thus is
 407 of the form $(v_{n,1}, f(v_{n,1}))$. We compute (an approximation of) $v_{n,1}$ by solving numerically
 408 the one-dimensional optimization problems defined in (3.7) and (3.8).

409 For all n , let note σ_n the permutation of $\llbracket 1, n \rrbracket$ which orders the vectors of \mathcal{S}_n by
 410 increasing F_1 -values and the so-called *ordered greedy set F_1 -values*:

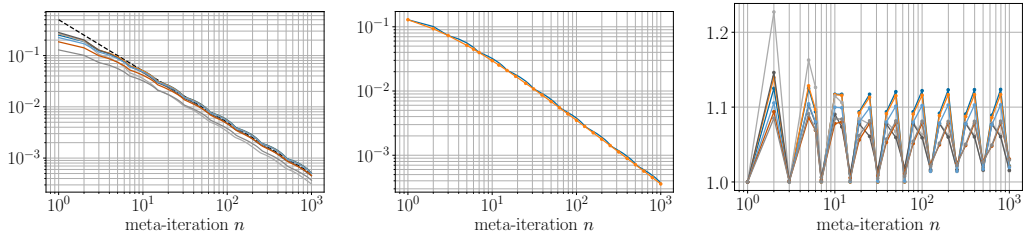
$$411 \quad (3.6) \quad w_{i,r}^n := v_{\sigma_n(i),1} \text{ for } i \in \llbracket 1, n \rrbracket, w_{0,r}^n := \tilde{x}_{\min,r} \text{ and } w_{n+1,r}^n := \tilde{x}_{\max,r} .$$

413 The following single-objective optimization problems

$$414 \quad (3.7) \quad v_{1,1} \in \arg \max_{x \in [0,1]} \text{HV}_r((x, f(x)))$$

$$415 \quad (3.8) \quad v_{n+1,1} \in \arg \max_{i \in \llbracket 1, n+1 \rrbracket} \max_{x \in [w_{1,r}^n, w_{n+1,r}^n]} \text{HV}_{w_{n+1,r}^n, f(w_{1,r}^n)}((x, f(x)))$$

417 are solved using the SLSQP version implemented in the python module `scipy.optimize`
 418 with the stopping criterions being set to `ftol:1e-13` and `maxiter:1000`. For each
 419 problem, we run the solver SLSQP three times starting it uniformly at random in the
 420 search interval. We ensured that the objective functions optimized did not differ by
 421 more than 10^{-12} between the runs. The source code is available at <https://github.com/enguiemarescaux/hypervolume-greedy-sequences>.
 422



(a) The optimality gap of \mathcal{S}_n and $1/2n$ (the dashed line) for n up to 1000.
 (b) The optimality gap of \mathcal{S}_n — and of a n -optimal distribution — for n up to 1000 for the **affine** Pareto front.
 (c) The optimality gap of \mathcal{S}_n normalized by the optimality gap of a n -optimal distribution for n up to 1000.

Fig. 3: Numerical speed of convergence of the greedy set sequence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ towards the Pareto front and comparison with the one of a n -optimal distribution, for $r = (1, 1)$. In (a) and (b), all benchmark Pareto fronts are examined: **affine** —, **convex-bil** —, **doublesphere** —, **zdt1** —, **concave-bil** —, **zdt2** — and **dtlz2** —.

423 In [Figure 3a](#), we display the optimality gap of \mathcal{S}_n with respect to the meta-iteration
 424 n for all benchmark Pareto fronts. We rely on the analytical expression of the Pareto
 425 front PF_f to compute its hypervolume and the optimality gap. We observe very similar
 426 convergence rates for all benchmark Pareto fronts. They are all close to the $(1/(2n))_{n \in \mathbb{N}}$
 427 line in the log-log scale. It is compliant with theory for the **affine** Pareto front. Indeed,
 428 let define the sequence of indices $(n_k)_{k \in \mathbb{N}}$ such that $n_0 = 1$ and $n_{k+1} = 2n_k + 1$. For

429 an affine Pareto front and the nadir chosen as reference point, there is a unique greedy
 430 set \mathcal{S}_{n_k} , which consists of objective-vectors regularly distributed on the Pareto front and
 431 is equal to the n_k -th optimal distribution. A direct consequence is that for any such
 432 n_k , the optimality gaps of \mathcal{S}_{n_k} and of the n_k -th optimal distribution are equal. For the
 433 **affine** Pareto front and $r = (1, 1)$, the optimality gap of the n -th optimal distribution is
 434 $1/(2n + 2)$ [3, Theorem 5].

435 Our underlying assumption of perfect single-objective optimization is theoretical and
 436 cannot be verified by a real solver. For HV-ISOOMOO algorithms coupled with a real
 437 single-objective optimization solver SOOPTIMIZER, we can still display convergence
 438 graphs with respect to meta-iterations as in Figure 3a. We can also display convergence
 439 graphs with respect to cumulated SOOPTIMIZER iterations.

440 In Figures 3b and 3c, we compare the optimality gap of \mathcal{S}_n with the smallest opti-
 441 mality gaps achievable by a set of n points. These are the optimality gaps of *n-optimal*
 442 *distributions* [3], the sets of n objective vectors with the highest hypervolume achievable
 443 by a set of this cardinal. We reuse the optimality gaps of *n-optimal* distributions which
 444 were computed for [22], for $n = 1, 2, 3, 5, 6, 7, 10, 12, 15, 19, 25, 31, 39, 50, 63, 79,$
 445 $100, 125, 158, 199, 251, 316, 398, 501, 630, 794$ and 1000. The details of the computation
 446 method can be found in [22, Section 5.2].

447 In Figure 3b, we display the optimality gap of \mathcal{S}_n with respect to the meta-iteration n
 448 for the **affine** Pareto front only, along with the optimality gap of *n-optimal* distributions.
 449 The curve of the optimality gap of the greedy set sequence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ follows the one of
 450 the *n-optimal* distribution, moving away and getting closer periodically. This is what we
 451 would expect theoretically, as detailed above.

452 In Figure 3c, we display the relation between the optimality gap of \mathcal{S}_n and of *n-*
 453 *optimal* distributions for all benchmark Pareto fronts. We see similar fluctuations as
 454 for the **affine** Pareto front, with the same periodicity. At the bottom of the curve, the
 455 optimality gap of \mathcal{S}_n is only a few percent larger than the one of a *n-optimal* distribution.
 456 In the worst case, that is for **doublesphere** Pareto front and for $n = 2$, the optimality
 457 gap of \mathcal{S}_n is only 23% larger than the one of a *n-optimal* distribution. For $n \geq 10$, what
 458 is lost in proportion by taking \mathcal{S}_n instead of a *n-optimal* distribution is always smaller
 459 than for the **affine** Pareto front for the displayed value of n . We conjecture that it is true
 460 for all $10 \leq n \leq 1000$. The **affine** curve stops reaching regularly the value 1, in contrast
 461 to what is known theoretically. It is explained by the discretization in n .

462 **4. Lower bounds on the normalized maximum hypervolume.** In this section,
 463 we provide bounds on the maximum hypervolume achievable by a single feasible vector
 464 normalized by the maximum hypervolume of a feasible set: $\frac{\max_{u \in \text{PF}_f} \text{HV}_r(u)}{\text{HV}_r(\text{PF}_f)}$. We refer
 465 to this ratio as the *normalized maximum hypervolume* with respect to r . Bounds on the
 466 normalized maximum hypervolume are exploited in Section 5 to provide bounds on the
 467 speed of convergence of greedy set sequences towards the Pareto front.

468 **4.1. Lower bound on the normalized maximum hypervolume for convex**
 469 **Pareto fronts.** The hypervolume relative to a reference point r of a vector $u = (x, f(x))$
 470 of the Pareto front is $\text{HV}_r(u) = (r_1 - x) \times (r_2 - f(x))$. From this simple formula, we
 471 derive in the next proposition necessary conditions for a vector of the Pareto front to be
 472 an hypervolume maximizer when f has at least left and right derivatives in x^* .

473 PROPOSITION 4.1. *Let $x^* \in]x_{\min}, x_{\max}[$ such that $u^* := (x^*, f(x^*))$ maximizes the*

474 *hypervolume with respect to a valid reference point r . If the function f describing the*
 475 *Pareto front admits left and right derivatives in x^* , respectively $f'_-(x^*)$ and $f'_+(x^*)$, then*

476 (4.1)
$$-f'_+(x^*) \leq \frac{r_2 - f(x^*)}{r_1 - x^*} \leq -f'_-(x^*) .$$

 477

478 *Proof.* We define the function $\text{HV}_{x,r}(\cdot)$ as $x \mapsto \text{HV}_r((x, f(x)))$. If x^* maximizes
 479 $\text{HV}_{x,r}(\cdot)$, then the left and the right derivatives of $\text{HV}_{x,r}(\cdot)$ are positive and negative,
 480 respectively. By replacing the left and right derivatives of $\text{HV}_{x,r}(\cdot)$ by their explicit
 481 formulas and reorganizing the terms we obtain (4.1). \square

482 Equation (4.1) states that the slope of the diagonal of the rectangle $\mathcal{D}_{u^*}^r$ is between the
 483 absolute values of the slopes of the right and the left derivatives of f at x^* (see the middle
 484 plot of Figure 4). To the best of our knowledge, this geometric interpretation is new. It
 485 becomes simpler when f is differentiable. Then, the absolute value of the slope of the
 486 tangent of the front at a non-extreme vector u^* is equal to the slope of the diagonal of
 487 the rectangle $\mathcal{D}_{u^*}^r$ (see the lefthand plot of Figure 4).

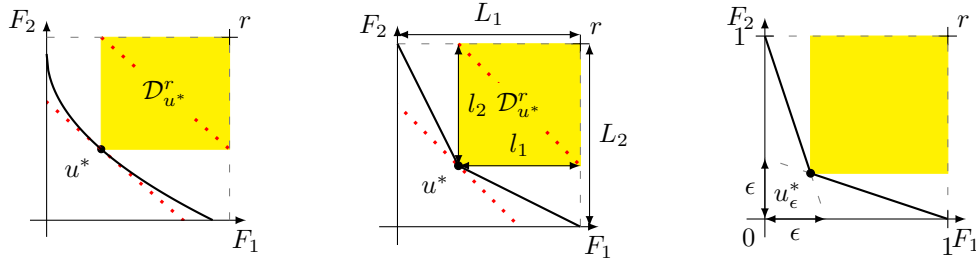


Fig. 4: Left and middle: Two convex Pareto fronts and their respective hypervolume maximizers u^* , one differentiable (left) and one non-differentiable (middle). The slopes of the two dotted lines, namely PF_g and the diagonal of $\mathcal{D}_{u^*}^r$, are equal. Right: The Pareto front PF_ϵ and the hypervolume maximizer u_ϵ^* for $\epsilon = 1/3$ and $r = (1, 1)$.

488 COROLLARY 4.2. *Let $x^* \in]x_{\min}, x_{\max}[$ be such that $u^* := (x^*, f(x^*))$ maximizes the*
 489 *hypervolume with respect to a valid reference point r . If the Pareto front is described by*
 490 *a differentiable function f in x^* , then $f'(x^*)$ satisfies*

491 (4.2)
$$-f'(x^*) = \frac{r_2 - f(x^*)}{r_1 - x^*} .$$

 492

493 *Proof.* It is a direct consequence of Proposition 4.1 \square

494 A convex function may not be differentiable, but it always has left and right de-
 495 rivatives. It is also above its left and right tangent lines respectively on the left and
 496 on the right of x^* . Therefore, Proposition 4.1 implies that the affine function $g : x \mapsto$
 497 $f(x^*) - \frac{r_2 - f(x^*)}{r_1 - x^*} \times (x - x^*)$ is a minorant of f . This is the key idea of the proof of the
 498 following lower bound on the normalized maximum hypervolume.

499 PROPOSITION 4.3. *If the Pareto front is described by a convex function f , then the*
 500 *following lower bound on the normalized maximum hypervolume with respect to any valid*

501 *reference point r holds:*

$$502 \quad (4.3) \quad \frac{\max_{u \in PF_f} HV_r(u)}{HV_r(PF_f)} \geq \frac{1}{2}$$

503

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

506 *Proof.* As explained in the above paragraph, the convexity of f implies that the
 507 affine function $g : x \mapsto f(x^*) - \frac{r_2 - f(x^*)}{r_1 - x^*} \times (x - x^*)$ is a minorant of f . Therefore,
 508 $PF_g := \{g(x) : x \in [x_{\min}, x_{\max}]\}$ dominates PF_f , and thus has a higher hypervolume.
 509 We denote $L_1 := r_1 - \tilde{x}_{\min, r}$ and $L_2 := r_2 - f(\tilde{x}_{\max, r})$ the lengths of the rectangle
 510 $\mathcal{R} := [\tilde{x}_{\min, r}, r_1] \times [f(\tilde{x}_{\max, r}), r_2]$. We denote $l_1 := r_1 - x^*$ and $l_2 := r_2 - f(x^*)$ the lengths
 511 of the rectangle $\mathcal{D}_{u^*}^r$. The region of \mathcal{R} which dominates PF_g is a right-angled triangle.
 512 Additionally, by definition, the slope of its hypotenuse is l_2/l_1 , and thus the lengths of
 513 the other sides are $L_1 - l_1 + (L_2 - l_2) \times \frac{l_1}{l_2}$ and $L_2 - l_2 + (L_1 - l_1) \times \frac{l_2}{l_1}$ (see the middle
 514 plot of [Figure 4](#)). Therefore, we have

$$515 \quad HV_r(PF_g) = \lambda(\mathcal{R}) - \lambda(\{u \in \mathbb{R}^2 : u \in \mathcal{R}, u \preceq PF_g\})$$

$$516 \quad = L_1 L_2 - \frac{1}{2} \times (L_1 - l_1 + (L_2 - l_2) \times \frac{l_1}{l_2}) \times (L_2 - l_2 + (L_1 - l_1) \times \frac{l_2}{l_1})$$

$$517 \quad = l_1 l_2 \times \left[-2 + 2 \times \frac{L_2}{l_2} - \frac{1}{2} \times \left(\frac{L_2}{l_2}\right)^2 + 2 \times \frac{L_1}{l_1} - \frac{1}{2} \times \left(\frac{L_1}{l_1}\right)^2 \right].$$

518

519 For all x , we have $(x - 2)^2 \geq 0$ and thus $2x - \frac{1}{2}x^2 \leq 2$. Therefore, we can conclude that
 520 $HV_r(PF_g)$, and thus $HV_r(PF_f)$ is smaller than $2 \times l_1 l_2$, that is $2 \times HV_r(u^*)$. If either
 521 $L_1/l_1 \neq 2$ or $L_2/l_2 \neq 2$, the inequality is strict. Thus, when the inequality is an equality,
 522 the center of \mathcal{R} belongs to the Pareto front. Since f is convex, it happens only when f is
 523 affine and the reference point r dominates the nadir point. Conversely, if both conditions
 524 are met, we know that the optimum is in the middle of the Pareto front and that we have
 525 the equality (see [[3](#), Theorem 5]). \square

526 We just proved that one half is a tight lower bound on the normalized maximum
 527 hypervolume for convex Pareto fronts. However, except for the trivial upper bound 1,
 528 there is no upper bound valid for every convex Pareto front, even when r dominates the
 529 nadir point. Let consider the Pareto fronts $PF_\epsilon := \{\max(1 - \frac{x}{\epsilon}, \epsilon - \epsilon \times x) : x \in [0, 1]\}$
 530 (represented in the righthand plot of [Figure 4](#) for $\epsilon = \frac{1}{3}$). The normalized maximum
 531 hypervolume of PF_ϵ for the reference point $r = (1, 1)$ converges to 1 when ϵ goes to 0.³

532 **4.2. Lower and upper bounds on the normalized maximum hypervolume**
 533 **for bilipschitz Pareto fronts.** In this section, we examine lower and upper bounds on
 534 the normalized maximum hypervolume in the case of bilipschitz Pareto fronts.

535 We consider two affine fronts with the same left extreme vector as PF_f and slopes
 536 $-L_{\min}$ and $-L_{\max}$, see the lefthand plot of [Figure 5](#). We call them PF_{\min} and PF_{\max} ,
 537 respectively. Formally:

$$538 \quad (4.4) \quad PF_{\max} := \{(x, f_{\max}(x)) : x \in [x_{\min}, x_{\max}]\} \text{ and}$$

$$539 \quad (4.5) \quad PF_{\min} := \{(x, f_{\min}(x)) : x \in [x_{\min}, x_{\max}]\}$$

³The normalized hypervolume equals $\frac{(1-\epsilon+\epsilon^2)^2}{1-\epsilon \times (1-\epsilon)^2 + (\epsilon-\epsilon^2)^2}$ which converges to 0 when ϵ goes to 0.

541 with $f_{\min}(x) = f(x_{\min}) - (x - x_{\min}) \times L_{\min}$ and $f_{\max}(x) = f(x_{\min}) - (x - x_{\min}) \times$
 542 L_{\max} . For a (L_{\min}, L_{\max}) -bilipschitz function f , $f_{\max}(x) \leq f(x) \leq f_{\min}(x)$ for all $x \in$
 543 $[x_{\min}, x_{\max}]$, and thus the Pareto front is dominated by PF_{\max} and dominates PF_{\min} .
 These two affine fronts provide bounds on both the hypervolume of the Pareto front

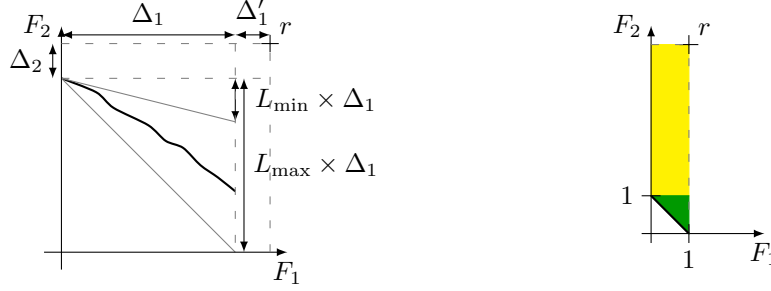


Fig. 5: Left : The Pareto front PF_f surrounded by PF_{\max} (below) and PF_{\min} (above).
 Right : An illustration that $\text{HV}_r(u_{\min}) - \text{HV}_r(\text{PF}_f)$ (■) becomes negligible compared to
 $\text{HV}_r(u_{\min})$ (■) for $r_1 = 1$ and $r_2 \rightarrow \infty$.

544 and the largest hypervolume of a vector of the Pareto front. They are key to prove
 545 the following lower bound on the normalized maximum hypervolume of a (L_{\min}, L_{\max}) -
 546 bilipschitz Pareto front.
 547

548 PROPOSITION 4.4. *If the Pareto front is described by a (L_{\min}, L_{\max}) -bilipschitz func-*
 549 *tion f , then for any valid reference point r , we have*

$$550 \quad (4.6) \quad \frac{\max_{u \in \text{PF}_f} \text{HV}_r(u)}{\text{HV}_r(\text{PF}_f)} \geq \frac{1}{2} \times \frac{L_{\min}}{L_{\max}} .$$

552 *Proof.* The fronts PF_{\max} and PF_{\min} are defined respectively in (4.4) and (4.5). We
 553 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 $V := \Delta_2 \times (r_1 -$
 554 $\tilde{x}_{\min, r})$, see the lefthand plot of Figure 5. Since the front PF_{\max} dominates the Pareto
 555 front, the hypervolume of PF_f is smaller than the hypervolume of PF_{\max} , $V + L_{\max} \times$
 556 $\Delta_1 \times \Delta'_1 + \frac{1}{2} \times L_{\max} \times \Delta_1^2$. Additionally, since each vector of PF_{\min} is dominated by a
 557 vector of PF_f , the maximum hypervolume of a vector of PF_f is larger than the maximum
 558 hypervolume of a vector of PF_{\min} . The front PF_{\min} being an affine and therefore convex
 559 front, we know by Proposition 4.3 that the maximum hypervolume of a vector of PF_{\min} is
 560 larger than half of $\text{HV}_r(\text{PF}_{\min})$, which is equal to $\frac{1}{2} \times (V + L_{\min} \times \Delta_1 \times \Delta'_1 + \frac{1}{2} \times L_{\min} \times \Delta_1^2)$.
 561 To summarize, the maximum hypervolume of a vector of PF_f is larger than $\frac{1}{2} \times (V + L_{\min} \times$
 562 $\Delta_1 \times \Delta'_1 + \frac{1}{2} \times L_{\min} \times \Delta_1^2)$. Combining the upper bound on the hypervolume of PF_f and the
 563 lower bound on the maximum hypervolume of a vector of PF_f , the normalized maximum
 564 hypervolume is larger 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
 565 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
 566 conclude that the normalized maximum hypervolume is larger than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}}$. \square

567 We cannot guarantee any upper bound strictly smaller than 1 on the normalized maximum
 568 hypervolume without adding an assumption on the reference point. Indeed, for a given

569 bounded Pareto front, it is easy to show that the normalized maximum hypervolume goes
570 to 1 for $r_1 = x_{\max}$ and $r_2 \rightarrow \infty$ (see the righthand plot of [Figure 5](#)). However, if f is
571 (L_{\min}, L_{\max}) -bilipschitz and r dominates the nadir point, we can prove that the normalized
572 maximum hypervolume is larger than $\frac{1}{2} \times \frac{L_{\max}}{L_{\min}}$. The proof relies on the fact that if the
573 reference point r dominates the nadir point, the vector of an affine front with the largest
574 hypervolume relative to r is its middle (see [[3](#), Theorem 5]), whose hypervolume is half
575 of the hypervolume of the entire Pareto front.

576 **PROPOSITION 4.5.** *If the Pareto front is described by a (L_{\min}, L_{\max}) -bilipschitz func-*
577 *tion f and the reference point r is valid and dominates the nadir point, the following*
578 *upper-bound on the normalized maximum hypervolume with respect to r holds*

$$579 \quad (4.7) \quad \frac{\max_{u \in PF_f} HV_r(u)}{HV_r(PF_f)} \leq \frac{1}{2} \times \frac{L_{\max}}{L_{\min}} .$$

581 *Proof.* We use the same notations as in the proof of [Proposition 4.4](#). Since r dominates
582 the nadir point, both Δ'_1 , Δ_2 and V equal 0, and thus the hypervolumes of PF_{\max}
583 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
584 PF_{\min} by PF_f implies that the hypervolume of the Pareto front is below $\frac{1}{2} \times L_{\min} \times$
585 Δ_1^2 . Since PF_{\max} is an affine front whose extremes dominate r , its middle is the unique
586 hypervolume maximizer (see [[2](#), Theorem 5]) with an hypervolume equal to $\frac{1}{4} \times L_{\max} \times$
587 Δ_1^2 . The domination of PF_f by PF_{\max} implies that the maximum hypervolume of a vector
588 of PF_f is larger than $\frac{1}{4} \times L_{\max} \times \Delta_1^2$. Gathering the lower bound on $HV_r(PF_f)$ and the
589 upper bound on the maximum hypervolume of a vector of PF_f , we retrieve (4.7). \square

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

593 **5. Convergence of HV-ISOOMOO coupled with perfect single-objective**
594 **optimization.** We prove in this section various convergence results for HV-ISOOMOO
595 algorithms coupled with perfect single-objective optimization. We first prove that when
596 the Pareto front is either convex or bilipschitz, these algorithms converge to the entire
597 Pareto front. We transform the bounds on the normalized maximum hypervolume proven
598 in [Section 4](#) into lower bounds on the speed of convergence. Then, we analyze the asymp-
599 totic convergence behavior when the Pareto front is bilipschitz with a Hölder continuous
600 derivative.

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

608 **5.1. Convergence of HV-ISOOMOO with guaranteed speed of conver-**
609 **gence.** We prove some upper bounds on the relation between the optimality gap at
610 iteration $2n + 1$ and at iteration n . These bounds translate into lower bounds on the
611 speed of convergence of HV-ISOOMOO under [Assumption 3.3](#) of perfect single-objective

612 optimization. The proof relies on inequalities of the form

613 (5.1)
$$\max_{u \in \text{PF}_f} \text{HV}_{r'}(u) \geq C \times \text{HV}_{r'}(\text{PF}_f)$$

614

615 stated in [Propositions 4.3](#) and [4.4](#) and equations regarding optimality gaps, areas of gap
616 regions and hypervolume improvement presented in [Subsection 2.2](#). A consequence of
617 (5.1) being true for any reference point r' is that the optimality gap at iteration $2n + 1$
618 is at most $(1 - C)$ times the optimality gap at iteration n .

619 We sketch the proof idea in the simple case where each of the v_k ($k \in \llbracket n + 1, 2n + 1 \rrbracket$)
620 is inserted in a distinct gap region of \mathcal{S}_n , see the lefthand plot of [Figure 6](#). Inserting v_k in
621 a gap region leads to an hypervolume improvement larger than C times the area of this
622 gap region by (5.1). Thus, the hypervolume improvement from iteration n to $2n + 1$
623 is larger than C times the area of the union of all gap regions of \mathcal{S}_n , namely the optimality
gap at iteration n . A detailed proof is presented after the theorem statement.

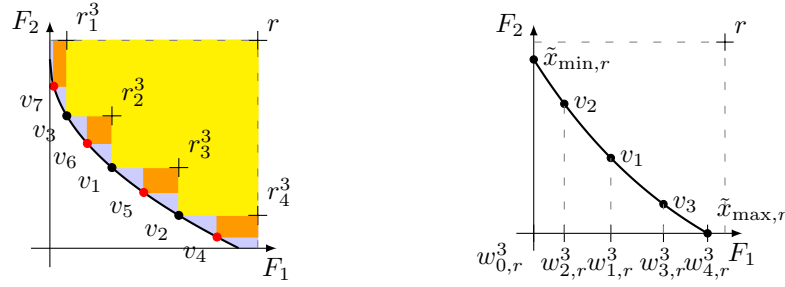


Fig. 6: Left: A Pareto front where each of the gap regions of \mathcal{S}_3 is filled by one of the greedy vectors v_k for $k \in \llbracket 4, 7 \rrbracket$. The front is described by $f(x) = 1 - \sqrt{x}$ for $x \in [0, 1]$. We represent the region $\mathcal{D}_{\mathcal{S}_3}^r$ (yellow), the gap regions of \mathcal{S}_3 (blue) and the regions corresponding to $\text{HVI}_r(v_k, \mathcal{S}_{k-1})$ for $k \in \llbracket 4, 7 \rrbracket$ (orange). Right: The ordered greedy set 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 $x \in [0, 1]$.

624

625 **PROPOSITION 5.1.** *Consider a biobjective optimization problem with a Pareto front*
626 *described by a function f . Any greedy set sequence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ relative to a valid reference*
627 *point r satisfies for all n*

628 (5.2)
$$\frac{\text{HV}_r(\text{PF}_f) - \text{HV}_r(\mathcal{S}_{2n+1})}{\text{HV}_r(\text{PF}_f) - \text{HV}_r(\mathcal{S}_n)} \leq 1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}}$$
 if f is (L_{\min}, L_{\max}) -bilipschitz and

629 (5.3)
$$\frac{\text{HV}_r(\text{PF}_f) - \text{HV}_r(\mathcal{S}_{2n+1})}{\text{HV}_r(\text{PF}_f) - \text{HV}_r(\mathcal{S}_n)} \leq \frac{1}{2}$$
 if f is convex.

630

631 *Proof.* Fix $n \geq 1$. We note σ a permutation of $\llbracket 1, n + 1 \rrbracket$ such that $n + \sigma(i)$ is the
632 index of the first greedy vector v_k inserted in $\mathcal{G}_{\mathcal{S}_n, i}^r$ when possible. With this choice of σ ,
633 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
634 improvement to $\mathcal{S}_{n+\sigma(i)-1}$ of any vector u belonging to the i -th gap region of \mathcal{S}_n , $\mathcal{G}_{\mathcal{S}_n, i}^r$,
635 is equal to $\text{HV}_{r_i^n}(u)$ by [Lemma 2.4](#). The hypervolume improvement of the greedy vector

636 $v_{n+\sigma(i)}$ to $\mathcal{S}_{n+\sigma(i)-1}$ being maximal, it is in particular larger than the one of any vector
637 of $\mathcal{G}_{\mathcal{S}_n, i}^r$ and thus than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times \text{HV}_{r_i^n}(\text{PF}_f)$ by [Proposition 4.4](#). In other words, the
638 hypervolume improvement at any iteration $n + \sigma(i)$ is larger than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times \text{HV}_{r_i^n}(\text{PF}_f)$.
639 By adding these inequations for all $i \in \llbracket 1, n + 1 \rrbracket$, we deduce that the hypervolume
640 improvement from iteration n to $2n + 1$ is larger than $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}} \times \sum_{i=1}^{n+1} \text{HV}_{r_i^n}(\text{PF}_f)$.
641 Since the sum of the $\text{HV}_{r_i^n}(\text{PF}_f)$ is the optimality gap at iteration n , we have [\(5.2\)](#). If
642 f is convex instead of bilipschitz, we use [Proposition 4.3](#) instead of [Proposition 4.4](#) and
643 obtain [\(5.3\)](#). \square

644 Since the optimality gaps form a decreasing sequence, such lower bounds on the relation
645 between the optimality gaps at iteration $2n + 1$ and at iteration n imply that the optimal-
646 ity gap associated to a greedy set sequence converges asymptotically to 0. Equivalently,
647 sets constructed by HV-ISOOMOO algorithms coupled with perfect single-objective opti-
648 mization converge to the entire Pareto front as stated formally below.

649 **THEOREM 5.2.** *Consider a biobjective optimization problem with a Pareto front de-*
650 *scribed by a bilipschitz or convex function f .*

651 *The hypervolume of a greedy set sequence relative to a valid reference point r converges*
652 *to the hypervolume of the entire Pareto front, i.e. $\text{HV}_r(\mathcal{S}_n) \xrightarrow[n \rightarrow \infty]{} \text{HV}_r(\text{PF}_f)$.*

653 *Equivalently, for such Pareto fronts and under [Assumption 3.3](#) of perfect single-*
654 *objective optimization, HV-ISOOMOO algorithms relative to a valid reference point r*
655 *converge in the sense of [Definition 3.2](#).*

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

659 **COROLLARY 5.3.** *Consider a biobjective optimization problem with a Pareto front*
660 *described by a (L_{\min}, L_{\max}) -bilipschitz function. A greedy set sequence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ relative*
661 *to a valid reference point r satisfies for all n*

$$662 \quad (5.4) \quad \frac{\text{HV}_r(\text{PF}_f) - \text{HV}_r(\mathcal{S}_n)}{\text{HV}_r(\text{PF}_f)} \leq \left(1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}}\right)^{\lfloor \log_2(n+1) \rfloor} \leq (2n+2)^{\log_2(1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}})} .$$

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

$$666 \quad (5.5) \quad \frac{\text{HV}_r(\text{PF}_f) - \text{HV}_r(\mathcal{S}_n)}{\text{HV}_r(\text{PF}_f)} \leq \left(\frac{1}{2}\right)^{\lfloor \log_2(n+1) \rfloor} \leq \frac{1}{2n+2} .$$

668 *Hence, for such reference points and under [Assumption 3.3](#) of perfect single-objective*
669 *optimization, HV-ISOOMOO algorithms relative to r satisfy [\(5.4\)](#) if f is (L_{\min}, L_{\max}) -*
670 *bilipschitz and [\(5.5\)](#) if f is convex where \mathcal{S}_n is replaced by \mathcal{I}_n , the final incumbents Pareto*
671 *front approximation at iteration n .*

672 *Proof.* The k -th term of the sequence defined by $u_0 = 1$ and $u_{n+1} = 2 \times u_n + 1$ for
673 all $n \geq 1$ is $2^k - 1$. Thus, [\(5.2\)](#) and [\(5.3\)](#) imply that when f is (L_{\min}, L_{\max}) -bilipschitz or
674 convex, the normalized optimality gap at iteration $2^k - 1$ is smaller than $(1 - C)^k$ with C
675 equal to $\frac{1}{2} \times \frac{L_{\min}}{L_{\max}}$ and $\frac{1}{2}$, respectively. Since the hypervolume of the greedy set increases
676 with n , and thus the optimality gap decreases with n , we deduce the first inequalities in
677 [\(5.4\)](#) and [\(5.5\)](#) via the change of variable $k = \lfloor \log_2(n + 1) \rfloor$.

678 Additionally, for every n , $\lfloor \log_2(n+1) \rfloor$ is smaller than $\log_2(n+1)+1$, that is $\log_2(2n+$
679 $2)$. For every C , $\log_2(2n+2)$ equals $\log_C(2n+2) \times \log_2(C)$, and thus $C^{\log_2(2n+2)}$ equals
680 $(2n+2)^{\log_2(C)}$. Therefore, we can infer that $(2n+2)^{\log_2(C)}$ is an upper bound of the
681 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}) -
682 bilipschitz and convex, respectively. \square

683 We focused on the relation between the optimality gap at iteration n and at iteration
684 $2n+1$. We can similarly examine the relation between the optimality gap at iteration n
685 and at any later iteration.

686 **LEMMA 5.4.** *If f is (L_{\min}, L_{\max}) -bilipschitz (resp. convex), then for all n , for all*
687 $k \leq n+1$, $\frac{HV_r(\text{PF}_f) - HV_r(\mathcal{S}_{n+k})}{HV_r(\text{PF}_f) - HV_r(\mathcal{S}_n)}$ *is smaller than* $1 - 1/2 \times L_{\min}/L_{\max} \times k/(n+1)$ *(resp.*
688 $1 - 1/2 \times k/(n+1)$ *).*

689 *Proof.* Consider the k gap regions of \mathcal{S}_n with the largest areas. If f is (L_{\min}, L_{\max}) -
690 bilipschitz (resp. convex), the hypervolume improvement from iteration n to $n+k$ is at
691 least $1/2 \times L_{\min}/L_{\max}$ (resp. $1/2$) times the area of the union of these gap regions, which
692 is at least $\frac{k}{n+1}$ times the optimality gap at iteration n . \square

693 Yet, the previous lemma leads to looser lower bounds on the convergence rate. To illus-
694 trate this, we detail the case $k=1$ in the following lemma.

695 **LEMMA 5.5.** *From the relation between optimality gaps at one iteration and the next*
696 *one given in Lemma 5.4, we deduce lower bounds on the relation between the optimality*
697 *gaps at iteration n and at iteration $2n+1$ of $1/\sqrt{e} \approx 0.61$ and $e^{-1/2 \times L_{\min}/L_{\max}}$ for convex*
698 *and (L_{\min}, L_{\max}) -bilipschitz Pareto fronts, respectively.*

699 *Proof.* By Lemma 5.4, the relation between the optimality gap at iteration n and
700 at iteration $n+1$ is smaller than $1 - C/(n+1)$ for convex and (L_{\min}, L_{\max}) -bilipschitz
701 Pareto fronts with $C = 1/2$ and $C = 1/2 \times L_{\min}/L_{\max}$, respectively. By recurrence, this
702 implies that the relation between the optimality gap at iteration n and at iteration $2n+1$
703 is smaller than $(1 - C \times \frac{1}{n+1})^{n+1}$. The sequence of lower bounds $((1 - C \times \frac{1}{n+1})^{n+1})_{n \in \mathbb{N}}$
704 is increasing and converges towards e^{-C} . This is a direct consequence of classic results
705 on the sequence $((1 - 1/n)^n)_{n \in \mathbb{N}}$. \square

706 These lower bounds are smaller than the lower bounds of Proposition 5.1: $1/\sqrt{e} \geq 0.5$
707 while $e^{-1/2 \times L_{\min}/L_{\max}} \geq 1 - 1/2 \times L_{\min}/L_{\max}$.

708 5.2. Asymptotical behavior of the convergence of $HV_r(\mathcal{S}_n)$ to $HV_r(\text{PF}_f)$.

709 In this section, we analyze the asymptotic convergence behavior for a Pareto front de-
710 scribed by a bilipschitz function with a Hölder continuous derivative. We prove that, in
711 this case, doubling the number of vectors in the greedy set divides the optimality gap by a
712 factor which converges asymptotically to two as stated in Theorem 5.12. This asymptotic
713 limit corresponds to the case of affine Pareto fronts with a reference point dominating
714 the nadir point. For such Pareto fronts and reference points, the optimality gap is always
715 halved when the number of vectors in the greedy set goes from n to $2n+1$, see Figure 7.

716 First, we study the properties of the part of the Pareto front corresponding to a
717 specific gap region of \mathcal{S}_n . We introduced the *ordered greedy set F_1 -values* in (3.6).
718 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 $i \in$
719 $\llbracket 1, n+1 \rrbracket$ form a partition of $[\hat{x}_{\min,r}, \hat{x}_{\max,r}]$, see the righthand plot of Figure 6. The
720 interval $[w_{i-1,r}^n, w_{i,r}^n]$ corresponds to the part of the Pareto front belonging to the i -th gap
721 region of \mathcal{S}_n . When the Pareto front is bilipschitz, the lengths of these intervals converge

722 asymptotically to 0 as stated in the next lemma. This result is a direct consequence of
 723 the convergence of $\text{HV}_r(\mathcal{S}_n)$ to $\text{HV}_r(\text{PF}_f)$ stated in [Theorem 5.2](#).

724 **LEMMA 5.6.** *If the Pareto front is described by a bilipschitz function f and the greedy*
 725 *set sequence is associated to a valid reference point r , then the ordered greedy set F_1 -values*
 726 *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 (3.6).*

727 *Proof.* Let L_{\min} and L_{\max} be constants such that f is (L_{\min}, L_{\max}) -bilipschitz. The
 728 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
 729 $\int_{w_{i-1,r}^n}^{w_{i,r}^n} L_{\min} \times (w_{i,r}^n - x) dx$, which equals $\frac{1}{2} \times L_{\min} \times (w_{i,r}^n - w_{i-1,r}^n)^2$. Since the area
 730 of any gap region of \mathcal{S}_n is smaller than the optimality gap at iteration n , this implies that
 731 the difference $w_{i,r}^n - w_{i-1,r}^n$ is smaller than $\sqrt{2 \times (\text{HV}_r(\text{PF}_f) - \text{HV}_r(\mathcal{S}_n))}$ for all n , for all
 732 $i \in \llbracket 1, n+1 \rrbracket$. Therefore, the convergence of $\text{HV}_r(\mathcal{S}_n)$ to $\text{HV}_r(\text{PF}_f)$ stated in [Theorem 5.2](#)
 733 implies that the maximum over i of $w_{i,r}^n - w_{i-1,r}^n$ converges to 0. \square

734 We prove in the next lemma that if the Pareto front is described by a bilipschitz
 735 function f with a Hölder continuous derivative, then the the part of the Pareto front
 736 belonging to a specific gap region of \mathcal{S}_n is bilipschitz for some constants whose ratio con-
 737 verges asymptotically to 1. Affine functions being the only functions to be (L_{\min}, L_{\max}) -
 738 bilipschitz with $L_{\min}/L_{\max} = 1$, it supports the interpretation that the convergence of
 739 a greedy set sequence for such Pareto fronts and for affine Pareto fronts share some
 740 asymptotic similarities.

741 When f is bilipschitz, its restriction to the part of the Pareto front dominating r_i^n ,
 742 that is $[w_{i-1,r}^n, w_{i,r}^n]$, is $(L_{\min}^{i,n}, L_{\max}^{i,n})$ -bilipschitz with

$$743 \quad (5.6) \quad \begin{aligned} L_{\min}^{i,n} &:= \inf \left\{ \left| \frac{f(x) - f(y)}{x - y} \right|, x, y \in [w_{i-1,r}^n, w_{i,r}^n], x \neq y \right\} \text{ and} \\ L_{\max}^{i,n} &:= \sup \left\{ \left| \frac{f(x) - f(y)}{x - y} \right|, x, y \in [w_{i-1,r}^n, w_{i,r}^n], x \neq y \right\}. \end{aligned}$$

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

$$746 \quad (5.7) \quad q_n := \max \left\{ \frac{L_{\max}^{i,n}}{L_{\min}^{i,n}}, i \in \llbracket 1, n+1 \rrbracket : [w_{i-1,r}^n, w_{i,r}^n] \neq \emptyset \right\}.$$

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

751 **LEMMA 5.7.** *We consider a greedy set sequence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ relative to a valid reference*
 752 *point r . If the Pareto front is described by a bilipschitz function with a Hölder continuous*
 753 *derivative, then q_n defined in (5.7) converges asymptotically to 1.*

754 *Proof.* We take α such that f' is Hölder continuous with exponent α , i.e f is $\mathcal{C}^{1,\alpha}$,
 755 and $L_{\min}, L_{\max} > 0$ such that the function f describing the Pareto front is (L_{\min}, L_{\max}) -
 756 bilipschitz. We recall that f is decreasing, and thus for all $x < y$, we have $f(x) - f(y) \geq 0$.
 757 Since f is $\mathcal{C}^{1,\alpha}$ and therefore \mathcal{C}^1 , the Taylor formula with Lagrange remainder states that
 758 for all $x < y$, there exists $\xi \in [x, y]$ such that $f(y) = f(x) + (y - x) \times f'(\xi)$. Since f is $\mathcal{C}^{1,\alpha}$,

759 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,
760 $\frac{f(y)-f(x)}{x-y}$ is smaller than $-f'(x) + [f']_{\mathcal{C}^\alpha} \times (y - x)^\alpha$. We now restrict ourselves to x and
761 y belonging to the non-empty interval $[w_{i-1,r}^n, w_{i,r}^n]$. Our goal is to find an upper bound
762 depending on i but not on either x or y . Since f is $\mathcal{C}^{1,\alpha}$, the difference between $-f'(x)$
763 and $-f'(w_{i-1,r}^n)$ is smaller than $[f']_{\mathcal{C}^\alpha} \times (x - w_{i-1,r}^n)^\alpha$, and thus $[f']_{\mathcal{C}^\alpha} \times (w_{i,r}^n - w_{i-1,r}^n)^\alpha$
764 . Additionally, the difference between x and y is smaller than $w_{i,r}^n - w_{i-1,r}^n$. We conclude
765 that for $x, y \in [w_{i-1,r}^n, w_{i,r}^n]$, $\frac{f(y)-f(x)}{x-y}$ is smaller than $-f'(w_{i-1,r}^n) + 2[f']_{\mathcal{C}^\alpha} \times (w_{i,r}^n -$
766 $w_{i-1,r}^n)^\alpha$, and thus so is $L_{\max}^{i,n}$ defined in (5.6).
767 Following the same approach, we can also infer that $L_{\max}^{i,n}$ defined in (5.6) is greater
768 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
769 $-f'(w_{i-1,r}^n)$ is greater than L_{\min} and $(w_{i,r}^n - w_{i-1,r}^n)^\alpha$ is smaller than $\max_{i \in [1, n+1]} (w_{i,r}^n -$
770 $w_{i-1,r}^n)^\alpha$. As a consequence, q_n is smaller than $\frac{L_{\min} + 2[f']_{\mathcal{C}^\alpha} \times \max_{i \in [1, n+1]} (w_{i,r}^n - w_{i-1,r}^n)^\alpha}{L_{\min} - 2[f']_{\mathcal{C}^\alpha} \times \max_{i \in [1, n+1]} (w_{i,r}^n - w_{i-1,r}^n)^\alpha}$. By
771 Lemma 5.6, $\max_{i \in [1, n+1]} w_{i,r}^n - w_{i-1,r}^n$ converges to 0 and thus, this upper bound on q_n
772 converges to 1. Since q_n is always larger than 1, it converges to 1. \square

773 A consequence of the previous lemma is that the bounds on the hypervolume improvement
774 of v_{n+1} to \mathcal{S}_n normalized by the area of the gap region filled by v_{n+1} that we can infer from
775 Propositions 4.4 and 4.5 converge asymptotically to 1/2, see (5.8). Similarly, the bounds
776 on the normalized area of the child of a gap region that we can infer from Lemma A.2
777 converge to 1/4, see (5.9). These asymptotic values correspond to the case of an affine
Pareto front with a reference point dominating the nadir point, see Figure 7.

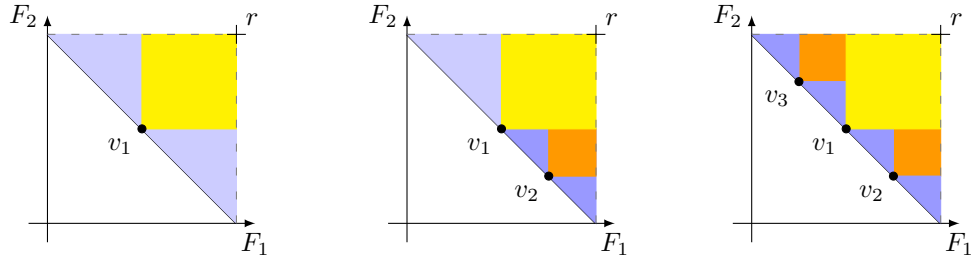


Fig. 7: The three greedy sets \mathcal{S}_1 (left), \mathcal{S}_2 (middle) and \mathcal{S}_3 (right) and their gap regions for an affine Pareto front with a reference point r dominating the nadir point. The area of any of the gap regions of \mathcal{S}_1 are half of $\text{HV}_r(\text{PF}_f)$ (left). The area of any 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 .

778

779 LEMMA 5.8. We consider a greedy set sequence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ relative to a valid reference
780 point r . If the Pareto front is described by a bilipschitz function f with a Hölder continuous
781 derivative, then for all $\epsilon > 0$, for n large enough, for every non-empty gap region $\mathcal{G}_{\mathcal{S}_n, i}^r$

782 and every child $\mathcal{G}_{\mathcal{S}_m, j}^r$ of $\mathcal{G}_{\mathcal{S}_n, i}^r$, we have

783 (5.8)
$$\frac{1}{2} \times (1 - \epsilon) \leq \frac{\max_{u \in \mathcal{G}_{\mathcal{S}_n, i}^r} \text{HVI}_r(u, \mathcal{S}_n)}{\lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)} \leq \frac{1}{2} \times (1 + \epsilon) \text{ and}$$

784 (5.9)
$$\frac{1}{4 \times (1 + \epsilon)} \leq \frac{\lambda(\mathcal{G}_{\mathcal{S}_m, j}^r)}{\lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)} \leq \frac{1}{4 \times (1 - \epsilon)} .$$

785

786 *Proof.* The interval $[w_{i-1, r}^n, w_{i, r}^n]$ is the set of the first coordinates of the vectors of
787 the Pareto front which dominate r_i^n . The restriction to $[w_{i-1, r}^n, w_{i, r}^n]$ of f is (L_{\min}, L_{\max}) -
788 bilipschitz for some L_{\min} and L_{\max} such that $\frac{L_{\max}}{L_{\min}} = q_n$ with q_n defined in (5.7). Ad-
789 ditionally, as stated in Proposition B.2, for n large enough, all the r_i^n corresponding to
790 non-empty gap regions dominate the nadir point. As a consequence, the conditions to
791 apply Lemma A.2 and Proposition 4.5 are met by non-extremes gap regions.

792 By Propositions 4.4 and 4.5, $\max_{u \in \text{PF}_f} \text{HV}_{r_i^n}(u) / \text{HV}_{r_i^n}(\text{PF}_f)$ is between $1/2 \times 1/q_n$ and
793 $1/2 \times q_n$. Additionally, by Lemma A.2, $\lambda(\mathcal{G}_{\mathcal{S}_m, j}^r) / \text{HV}_{r_i^n}(\text{PF}_f)$ is between $(1 - 1/2 \times q_n) / (1 +$
794 $q_n^2)$ and $(1 - 1/2 \times 1/q_n) / (1 + 1/q_n^2)$. The maximum over the vectors u belonging to the
795 Pareto front of $\text{HV}_{r_i^n}(u)$ is equal to the maximum over u belonging to the i -th gap region
796 of \mathcal{S}_n of $\text{HVI}_r(u, \mathcal{S}_n)$. Indeed, $\text{HV}_{r_i^n}(\cdot)$ is null for vectors outside the i -th gap region of \mathcal{S}_n
797 while it is nonnegative, equal to $\text{HVI}_r(\cdot, \mathcal{S}_n)$, otherwise. Additionally, $\text{HV}_{r_i^n}(\text{PF}_f)$ equals
798 $\lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)$. The convergence of q_n to 1 stated in Lemma 5.7 imply that the bounds proven
799 so far converge to a half and a quarter, respectively. Thus, we have (5.8) and (5.9) for n
800 large enough. \square

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

805 LEMMA 5.9. *We consider a greedy set sequence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ relative to a valid reference*
806 *point r . If the Pareto front is described by a bilipschitz function with a Hölder continuous*
807 *derivative, then for all $\epsilon > 0$, for n large enough and for any non-empty gap regions of*
808 *\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*

809 (5.10)
$$\frac{\lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)}{\lambda(\mathcal{G}_{\mathcal{S}_n, j}^r)} \leq 4 \times \frac{(1 + \epsilon)^2}{1 - \epsilon} .$$

810

811 *Proof.* Fix $\epsilon > 0$. By Lemma 5.8, there exists $N_1 \in \mathbb{N}^*$ such that for all n greater
812 than N_1 , (5.8) and (5.9) are verified for any non-empty gap region of \mathcal{S}_n and its children.
813 Since $\max_{i \in \llbracket 1, n + 1 \rrbracket} w_{i, r}^n - w_{i-1, r}^n$ converges to 0 by Lemma 5.6, every non-empty gap region
814 is filled at some point. Take N_2 such that all the non-empty gap regions of \mathcal{S}_{N_1} are filled
815 at iteration N_2 . For all n greater than N_2 , (5.8) and (5.9) are true for any non-empty
816 gap region of \mathcal{S}_n and its children, but also for its parents.

817 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
818 of \mathcal{S}_n , and \mathcal{P}_1 and \mathcal{P}_2 their respective parents. When two sets correspond to gap regions
819 relative to the same greedy set \mathcal{S}_m , we say that they cohabit at iteration m . Since only
820 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
821 and \mathcal{P}_2 or \mathcal{G}_2 and \mathcal{P}_1 cohabit at some earlier iteration. In the first case, there necessarily
822 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

823 to \mathcal{P}_2 , otherwise, \mathcal{G}_1 and \mathcal{G}_2 would not cohabit. By (5.8), the maximum hypervolume
824 improvement to \mathcal{S}_m of a vector of \mathcal{G}_1 and of a vector of \mathcal{P}_2 are at least $\frac{1}{2} \times (1 - \epsilon) \times \lambda(\mathcal{G}_1)$
825 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
826 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
827 $\lambda(\mathcal{P}_2)$ is smaller than $4 \times (1 + \epsilon)$ times the area of its child $\lambda(\mathcal{G}_2)$ by (5.9), this inequality
828 implies (5.10). In the second case, \mathcal{P}_2 is filled before \mathcal{P}_1 . Thus, there exists $m \geq N_2$ such
829 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
830 larger than the one of its child \mathcal{G}_1 , the hypervolume improvement of v_{m+1} to \mathcal{S}_m is still
831 larger than $\frac{1}{2} \times (1 - \epsilon) \times \lambda(\mathcal{G}_1)$. The rest of the argumentation remains valid. \square

832 We now have all the results needed to analyze the asymptotic impact of doubling the
833 number of points in the greedy set. To prove the following asymptotic upper bound,
834 we rely on similar arguments as for its nonasymptotic counterpart, Proposition 5.1. The
835 previous lemma guarantees that the impact of doubling the number of points in the greedy
836 set is asymptotically similar to the impact of passing from n points to $2n + 1$.

837 **PROPOSITION 5.10.** *Let $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ be a greedy set sequence relative to valid reference*
838 *point r . If the Pareto front is described by a bilipschitz function f with a Hölder continuous*
839 *derivative, then for all $\epsilon > 0$, we have for n large enough*

$$840 \quad (5.11) \quad \frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \leq \frac{1}{2} + o(\epsilon) .$$

842 *Proof.* Fix $\epsilon > 0$. Fix n large enough to verify (5.8) and (5.10) for this particular ϵ .

843 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
844 $v_{n+\sigma(i)}$ when it is filled before iteration $2n + 1$. With this choice of permutation, $\mathcal{G}_{\mathcal{S}_n, i}^r$
845 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
846 maximum hypervolume improvement of a vector of $\mathcal{G}_{\mathcal{S}_n, i}^r$ to $\mathcal{S}_{n+\sigma(i)-1}$, which is superior
847 to $\frac{1}{2} \times (1 - \epsilon) \times \lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)$ by (5.8). It is equivalent to say that the hypervolume improvement
848 at iteration $n + \sigma(i)$ is larger than $\frac{1}{2} \times (1 - \epsilon) \times \lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)$. Summing over $i \in \llbracket 1, n + 1 \rrbracket$, we
849 obtain that the hypervolume improvement between iteration n and $2n + 1$ is larger than
850 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 at
851 iteration n .

852 Now, we need to bound the hypervolume improvement at iteration $2n + 1$, that is
853 $HVI_r(v_{2n+1}, \mathcal{S}_{2n})$. It is smaller 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 (3.5) and
854 (5.8). Since the area of a gap region is smaller than the one of its parent, the maximum
855 area of a gap region is lower at iteration $2n$ than at iteration n . The maximum area of
856 one of the more than $n - 1$ gap regions of \mathcal{S}_n is itself smaller than $\frac{1}{n-1} \times \frac{4 \times (1 + \epsilon)^2}{1 - \epsilon}$ times
857 the optimality gap at iteration n by (5.10).

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

860 We roughly follow the same approach to obtain the following asymptotic lower bound on
861 the impact of doubling the number of points in the greedy set. Lemmas 5.8 and 5.9 are
862 key to prove an upper bound on the hypervolume improvement at iteration k . They allow
863 to prove that filling a gap region of \mathcal{S}_n more than once gives, up to a factor $1 + o(\epsilon)$, a
864 lower hypervolume improvement than filling a gap region which was not filled. Indeed,
865 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
866 parent by Lemma 5.8, which is itself at most $4 + o(\epsilon)$ times the area of any other gap
867 region of \mathcal{S}_n by Lemma 5.9.

868 PROPOSITION 5.11. Let $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ be a greedy set sequence relative to a valid reference
869 point r . If the Pareto front is described by a bilipschitz function f with a Hölder continuous
870 derivative, then for all $\epsilon > 0$, we have for n large enough

$$871 \quad (5.12) \quad \frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \geq \frac{1}{2} + o(\epsilon) .$$

873 *Proof.* Fix $\epsilon > 0$. Fix n large enough to verify (5.8), (5.9) and (5.10) for this particular
874 ϵ . Let $\delta \in \{-1, 0, 1\}$ be such that \mathcal{S}_n has $n + \delta$ non-empty gap regions. Let $i_0 := 1$ when
875 the left extreme gap region is empty and $i_0 := 0$ otherwise.

876 Let σ be a permutation of $\llbracket 1, n + \delta \rrbracket$ such that the i -th non-empty gap region of
877 \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
878 distinguish two cases. In the first case, $v_{n+\sigma(i)}$ is the child of the i -th non-empty gap
879 region of \mathcal{S}_n , and consequently its hypervolume improvement to $\mathcal{S}_{n+\sigma(i)-1}$ is at most
880 $\frac{1}{2} \times (1 + \epsilon) \times \lambda(\mathcal{G}_{\mathcal{S}_n, i_0+i}^r)$ by (5.8). In the second case, $v_{n+\sigma(i)}$ belongs to $\mathcal{G}_{\mathcal{S}_n, i_0+j}^r$, the
881 j -th non-empty gap region of \mathcal{S}_n , with $j \neq i$ and, by definition of σ , fills a descendant of
882 this gap region not $\mathcal{G}_{\mathcal{S}_n, i_0+j}^r$ itself. By (5.8), the hypervolume improvement of $v_{n+\sigma(i)}$ to
883 $\mathcal{S}_{n+\sigma(i)-1}$ is still at most $\frac{1}{2} \times (1 + \epsilon)$ times the area of the gap region it fills. By (5.9),
884 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.
885 By (5.10), we also know that the area of the i -th non-empty gap region of \mathcal{S}_n is at most
886 $4 \times \frac{(1+\epsilon)^2}{1-\epsilon}$ times the area of any other gap region of \mathcal{S}_n , in particular its i -th non-empty
887 gap region. We conclude that the hypervolume improvement of $v_{n+\sigma(i)}$ to $\mathcal{S}_{n+\sigma(i)-1}$ is
888 smaller 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 smaller than $\frac{(1+\epsilon)^3}{(1-\epsilon)^2}$, the
889 hypervolume improvement at any iteration $n + \sigma(i)$ is smaller than $\frac{1}{2} \times \frac{(1+\epsilon)^3}{(1-\epsilon)^2} \times \lambda(\mathcal{G}_{\mathcal{S}_n, i}^r)$.
890 Summing over $i \in \llbracket 1, n + \delta \rrbracket$, the hypervolume improvement from iteration n to $2n + \delta$
891 is smaller 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
892 iteration n .

893 Now, it is left to prove an upper bound on $HV_r(\mathcal{S}_{2n}) - HV_r(\mathcal{S}_{2n+\delta})$. This quantity
894 is maximal for $\delta = -1$, where it is simply the hypervolume improvement at iteration $2n$.
895 As in the previous proof, it is smaller than $\frac{1+\epsilon}{2 \times (n-1)}$ times the optimality gap at iteration
896 n . Therefore, the relation between the optimality gap at iteration $2n$ and at iteration n
897 is larger than $1 - \frac{1}{2} \times \frac{(1+\epsilon)^3}{(1-\epsilon)^2} - \frac{1+\epsilon}{2 \times (n-1)}$. \square

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

899 THEOREM 5.12. Consider a biobjective optimization problem and a greedy set se-
900 quence $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ relative to a valid reference point r . If the Pareto front is described by a
901 bilipschitz function f with a Hölder continuous derivative, we have

$$902 \quad (5.13) \quad \frac{HV_r(PF_f) - HV_r(\mathcal{S}_{2n})}{HV_r(PF_f) - HV_r(\mathcal{S}_n)} \xrightarrow{n \rightarrow \infty} \frac{1}{2} .$$

904 Consequently, for such Pareto front and reference point and under Assumption 3.3 of
905 perfect single-objective optimization, HV-ISOMOO algorithms relative to r satisfy (5.13)
906 where \mathcal{S}_n is replaced by \mathcal{I}_n , the final incumbents Pareto front approximation at iteration
907 n .

908 **6. Conclusion.** We prove that HV-ISOMOO algorithms coupled with a perfect
909 single-objective solver have a $O(1/n)$ convergence rate on convex Pareto fronts and a

910 $O(1/n^c)$ convergence rate on bilipschitz Pareto fronts with $c \leq 1$ depending on the bilips-
 911 chitz constants where n is the number of meta-iterations. Each meta-iteration corresponds
 912 to a single-objective optimization run. Both bounds are tight over the class of Pareto
 913 fronts and reference points considered. They are reached for affine Pareto fronts and
 914 reference points dominating the nadir point. On convex Pareto fronts, the convergence
 915 rate is exactly $\Theta(1/n)$, the fastest convergence rate achievable by biobjective optimization
 916 algorithms [22].

917 We also investigate numerically the non-asymptotic speed of convergence of HV-
 918 ISOOMOO algorithms coupled with a perfect single-objective solver on some simple con-
 919 vex and concave Pareto fronts. The optimality gap of the final incumbents Pareto front
 920 approximation at meta-iteration n is close to the optimality gap of n -optimal distributions,
 921 that is the lowest optimality gap achievable by a Pareto front representation of cardinal
 922 n . The ratio between these two optimality gaps fluctuates periodically with respect to
 923 n . At the lowest, the optimality gap of the final incumbents Pareto front approximation
 924 is only a few percent larger than the optimality gap of n -optimal distributions, while at
 925 the largest, it is 23% larger in the worst case. Both of these numerical and theoretical
 926 results show that greedily adding points maximizing the hypervolume contribution as in
 927 HV-ISOOMOO algorithms is an effective way to quickly increase the hypervolume.

928 Finally, we prove that for bilipschitz Pareto fronts with a Hölder continuous deriva-
 929 tive, doubling the number of meta-iterations divides the optimality gap by a factor which
 930 converges asymptotically to two. This asymptotic behavior resembles what we would ob-
 931 serve with an affine Pareto front and a reference point dominating the nadir point. Yet, it
 932 does not guarantee a $\Theta(1/n)$ convergence rate. Both $\left(\frac{\log(n)}{n}\right)_{n \in \mathbb{N}^*}$ and $\left(\frac{1}{n \times \log(n)}\right)_{n \in \mathbb{N}^*}$
 933 are examples of sequences verifying this property which do not have a $\Theta(1/n)$ convergence
 934 rate. The convergence rate on nonconvex Pareto fronts could theoretically be worse than
 935 $\Theta(1/n)$, but not better [22].

936 Convergence rates with respect to meta-iterations similar to those achieved under
 937 the assumption of perfect single-objective optimization may be observed in practice. We
 938 expect it to be the case for efficient implementations of HV-ISOOMOO and easy multi-
 939 objective problems, where the single-objective solver should return good approximations
 940 of global optima. Additionally, lower bounds on the speed of convergence may be directly
 941 derived for practical HV-ISOOMOO algorithms if a non-asymptotic lower bound on the
 942 speed of convergence towards a global optimum is known for the single-objective solver
 943 SOOPTIMIZER and if SOOPTIMIZER stops late enough. To do this, one could rely on
 944 the approach described in Lemma 5.4 with $k = 1$ and consider the vectors v_n^* towards
 945 which the single-objective solver converges instead of the greedy vectors v_n . The v_n^* are
 946 global optima of the true subproblems solved during meta-iterations.

947 We expect that the approach we use to prove a lower bound on the speed of conver-
 948 gence of HV-ISOOMOO coupled with a perfect single-objective solver generalizes to any
 949 number of objectives. Indeed, the gap regions can still be defined using local nadir points
 950 for any number of objectives $m \geq 3$ but they are no longer disjoint. This implies that
 951 we need to consider the hypervolume improvement from iteration n to $n + 1$ instead of
 952 $2n + 1$. Lemma 5.4 details how to do so. On top of that, the proof of Proposition 5.1
 953 only requires an upper bound on the number of gap regions and a lower bound on the
 954 normalized maximum hypervolume for some categories of Pareto fronts. It is known that
 955 for $m = 3$ and $m > 3$, there are respectively less than $2n + 1$ [12] and $O(n^{\lfloor \frac{m}{2} \rfloor})$ [19] gap re-
 956 gions associated to a set of n points. We conjecture that lower bounds on the normalized

957 maximum hypervolume can be proven for $m \geq 3$ for convex Pareto fronts and for Pareto
958 fronts with an explicit representation which is (L_{\min}, L_{\max}) -bilipschitz w.r.t. all variables.
959 A possible generalization of [Theorem 5.12](#) asymptotic insight is more open.

960 **Acknowledgments.** We would like to thank Dimo Brockhoff and Nikolaus Hansen
961 for several helpful discussions, notably about the importance of decomposing the opti-
962 mality gap and for the conjecture of [\(5.13\)](#).

963

REFERENCES

- 964 [1] C. AUDET, G. SAVARD, AND W. ZGHAL, *Multiobjective Optimization Through a Series of Single-*
965 *Objective Formulations*, SIAM Journal on Optimization, 19 (2008), pp. 188–210, <https://doi.org/10.1137/060677513>.
966 [2] A. AUGER, J. BADER, AND D. BROCKHOFF, *Theoretically investigating optimal μ -distributions for*
967 *the hypervolume indicator: first results for three objectives*, in Proceedings of the 11th interna-
968 tional conference on Parallel problem solving from nature: Part I, PPSN'10, Berlin, Heidelberg,
969 Sept. 2010, Springer-Verlag, pp. 586–596.
970 [3] A. AUGER, J. BADER, D. BROCKHOFF, AND E. ZITZLER, *Theory of the Hypervolume Indicator:*
971 *Optimal μ -Distributions and the Choice of the Reference Point*, in Proceedings of the Tenth
972 ACM SIGEVO Workshop on Foundations of Genetic Algorithms, FOGA '09, New York, NY,
973 USA, Jan. 2009, Association for Computing Machinery, pp. 87–102, [https://doi.org/10.1145/](https://doi.org/10.1145/1527125.1527138)
974 [1527125.1527138](https://doi.org/10.1145/1527125.1527138).
975 [4] A. AUGER, J. BADER, D. BROCKHOFF, AND E. ZITZLER, *Hypervolume-based multiobjective opti-*
976 *mization: Theoretical foundations and practical implications*, Theoretical Computer Science,
977 425 (2012), pp. 75–103, <https://doi.org/10.1016/j.tcs.2011.03.012>.
978 [5] N. BEUME, M. LAUMANN, AND G. RUDOLPH, *Convergence Rates of (1+1) Evolutionary Multiobjec-*
979 *tive Optimization Algorithms*, in Parallel Problem Solving from Nature, PPSN XI, R. Schaefer,
980 C. Cotta, J. Kołodziej, and G. Rudolph, eds., Lecture Notes in Computer Science, Berlin,
981 Heidelberg, 2010, Springer, pp. 597–606, https://doi.org/10.1007/978-3-642-15844-5_60.
982 [6] N. BEUME, B. NAUJOKS, AND M. EMMERICH, *SMS-EMOA: Multiobjective selection based on dom-*
983 *inated hypervolume*, European Journal of Operational Research, 181 (2007), pp. 1653–1669,
984 <https://doi.org/10.1016/j.ejor.2006.08.008>.
985 [7] J. BIGEON, S. LE DIGABEL, AND L. SALOMON, *DMulti-MADS: Mesh adaptive direct multisearch*
986 *for bound-constrained blackbox multiobjective optimization*, Computational Optimization and
987 Applications, 79 (2021), pp. 301–338, <https://doi.org/10.1007/s10589-021-00272-9>.
988 [8] K. BRINGMANN AND T. FRIEDRICH, *The maximum hypervolume set yields near-optimal approxima-*
989 *tion*, in Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation
990 - GECCO '10, Portland, Oregon, USA, 2010, ACM Press, p. 511, [https://doi.org/10.1145/](https://doi.org/10.1145/1830483.1830576)
991 [1830483.1830576](https://doi.org/10.1145/1830483.1830576).
992 [9] D. BROCKHOFF, *Optimal μ -distributions for the hypervolume indicator for problems with linear bi-*
993 *objective fronts: exact and exhaustive results*, in Proceedings of the 8th international conference
994 on Simulated evolution and learning, SEAL'10, Berlin, Heidelberg, Dec. 2010, Springer-Verlag,
995 pp. 24–34.
996 [10] A. L. CUSTÓDIO AND J. F. A. MADEIRA, *MultiGLODS: Global and local multiobjective optimization*
997 *using direct search*, Journal of Global Optimization, 72 (2018), pp. 323–345, [https://doi.org/](https://doi.org/10.1007/s10898-018-0618-1)
998 [10.1007/s10898-018-0618-1](https://doi.org/10.1007/s10898-018-0618-1).
999 [11] A. L. CUSTÓDIO, J. F. A. MADEIRA, A. I. F. VAZ, AND L. N. VICENTE, *Direct Multisearch for*
1000 *Multiobjective Optimization*, SIAM Journal on Optimization, 21 (2011), pp. 1109–1140, [https://doi.org/](https://doi.org/10.1137/10079731X)
1001 [10.1137/10079731X](https://doi.org/10.1137/10079731X).
1002 [12] K. DÄCHERT AND K. KLAMROTH, *A linear bound on the number of scalarizations needed to solve*
1003 *discrete tricriteria optimization problems*, Journal of Global Optimization, 61 (2015), pp. 643–
1004 676, <https://doi.org/10.1007/s10898-014-0205-z>.
1005 [13] K. DÄCHERT AND K. TEICHERT, *An improved hyperboxing algorithm for calculating a Pareto front*
1006 *representation*, arXiv:2003.14249, (2020), <https://arxiv.org/abs/2003.14249>.
1007 [14] E. DE KLERK, F. GLINEUR, AND A. B. TAYLOR, *On the worst-case complexity of the gradient method*
1008 *with exact line search for smooth strongly convex functions*, Optimization Letters, 11 (2017),
1009 pp. 1185–1199, <https://doi.org/10.1007/s11590-016-1087-4>.
1010

- 1011 [15] K. DEB, L. THIELE, M. LAUMANN, AND E. ZITZLER, *Scalable multi-objective optimization test*
1012 *problems*, in Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat.
1013 No.02TH8600), vol. 1, Honolulu, HI, USA, USA, May 2002, IEEE, pp. 825–830 vol.1, <https://doi.org/10.1109/CEC.2002.1007032>.
1014
- 1015 [16] M. EHRGOTT, *Multicriteria Optimization*, Springer-Verlag, Berlin Heidelberg, second ed., 2005,
1016 <https://doi.org/10.1007/3-540-27659-9>.
- 1017 [17] R. FIORENZA, *Hölder and locally Hölder Continuous Functions, and Open Sets of Class C^k , $C^{k,\lambda}$* ,
1018 *Frontiers in Mathematics*, Birkhäuser Basel, 2016, <https://doi.org/10.1007/978-3-319-47940-8>.
- 1019 [18] E. H. FUKUDA AND L. M. G. DRUMMOND, *On the convergence of the projected gradient method*
1020 *for vector optimization*, *Optimization*, 60 (2011), pp. 1009–1021, <https://doi.org/10.1080/02331934.2010.522710>.
1021
- 1022 [19] H. KAPLAN, N. RUBIN, M. SHARIR, AND E. VERBIN, *Efficient Colored Orthogonal Range Counting*,
1023 *SIAM Journal on Computing*, 38 (2008), pp. 982–1011, <https://doi.org/10.1137/070684483>.
- 1024 [20] K. KLAMROTH, R. LACOUR, AND D. VANDERPOOTEN, *On the representation of the search region in*
1025 *multi-objective optimization*, *European Journal of Operational Research*, 245 (2015), pp. 767–
1026 778, <https://doi.org/10.1016/j.ejor.2015.03.031>.
- 1027 [21] J. KNOWLES AND D. CORNE, *On metrics for comparing nondominated sets*, in Proceedings of the
1028 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600), vol. 1, May 2002,
1029 pp. 711–716 vol.1, <https://doi.org/10.1109/CEC.2002.1007013>.
- 1030 [22] E. MARESCAUX AND N. HANSEN, *Hypervolume in biobjective optimization cannot converge faster*
1031 *than $\Omega(1/p)$* , in Proceedings of the Genetic and Evolutionary Computation Conference, GECCO
1032 '21, New York, NY, USA, June 2021, Association for Computing Machinery, pp. 430–438,
1033 <https://doi.org/10.1145/3449639.3459371>.
- 1034 [23] G. L. NEMHAUSER, L. A. WOLSEY, AND M. L. FISHER, *An analysis of approximations for max-*
1035 *imizing submodular set functions—I*, *Mathematical Programming*, 14 (1978), pp. 265–294,
1036 <https://doi.org/10.1007/BF01588971>.
- 1037 [24] S. N. PARRAGH AND F. TRICOIRE, *Branch-and-Bound for Bi-objective Integer Programming*, *IN-*
1038 *FORMS Journal on Computing*, 31 (2019), pp. 805–822, [https://doi.org/10.1287/ijoc.2018.](https://doi.org/10.1287/ijoc.2018.0856)
1039 0856.
- 1040 [25] C. TOURÉ, N. HANSEN, A. AUGER, AND D. BROCKHOFF, *Uncrowded hypervolume improvement:*
1041 *COMO-CMA-ES and the softmax framework*, in Proceedings of the Genetic and Evolutionary
1042 Computation Conference, GECCO '19, New York, NY, USA, July 2019, Association for
1043 Computing Machinery, pp. 638–646, <https://doi.org/10.1145/3321707.3321852>.
- 1044 [26] D. A. V. VELDUIZEN AND G. B. LAMONT, *Evolutionary Computation and Convergence to a Pareto*
1045 *Front*, in Proceedings of the Third Annual Conference on Genetic Programming, San Francisco,
1046 1998, Stanford University Bookstore, pp. 221–228.
- 1047 [27] E. ZITZLER, K. DEB, AND L. THIELE, *Comparison of Multiobjective Evolutionary Algorithms: Em-*
1048 *pirical Results*, *Evolutionary Computation*, 8 (2000), pp. 173–195, <https://doi.org/10.1162/106365600568202>.
1049
- 1050 [28] E. ZITZLER, J. KNOWLES, AND L. THIELE, *Quality Assessment of Pareto Set Approximations*, in
1051 *Multiobjective Optimization: Interactive and Evolutionary Approaches*, J. Branke, K. Deb,
1052 K. Miettinen, and R. Słowiński, eds., *Lecture Notes in Computer Science*, Springer, Berlin,
1053 Heidelberg, 2008, pp. 373–404, https://doi.org/10.1007/978-3-540-88908-3_14.

1054 **Appendix A. Normalized areas of the gap regions relative to an hyper-**
1055 **volume maximizer.** The goal of this section is to prove bounds on the normalized
1056 areas of the gap regions $\mathcal{G}_{\text{left}}^{u^*}$ and $\mathcal{G}_{\text{right}}^{u^*}$ relative to an hypervolume maximizer u^* (see the
1057 lefthand plot of [Figure 8](#)) in the case of a bilipschitz Pareto front and of a reference point
1058 r dominating the nadir point. These bounds are stated in [Lemma A.2](#). The proof relies
1059 on the bounds on the normalized maximum hypervolume proven in [Subsection 4.2](#) and
1060 the following lower and upper bounds on the relation between $\lambda(\mathcal{G}_{\text{left}}^{u^*})$ and $\lambda(\mathcal{G}_{\text{right}}^{u^*})$.

1061 **PROPOSITION A.1.** *We assume that the Pareto front is described by a (L_{\min}, L_{\max}) -*
1062 *bilipschitz function f . Let u^* be a non-extreme vector of the Pareto front which maximizes*
1063 *the hypervolume with respect to a valid reference point r . If $r_1 \leq x_{\max}$, we have $\lambda(\mathcal{G}_{\text{right}}^{u^*}) \geq$
1064 $\frac{L_{\min}^2}{L_{\max}^2} \times \lambda(\mathcal{G}_{\text{left}}^{u^*})$. *If $r_2 \leq f(x_{\min})$, we have $\lambda(\mathcal{G}_{\text{left}}^{u^*}) \geq \frac{L_{\min}^2}{L_{\max}^2} \times \lambda(\mathcal{G}_{\text{right}}^{u^*})$.**

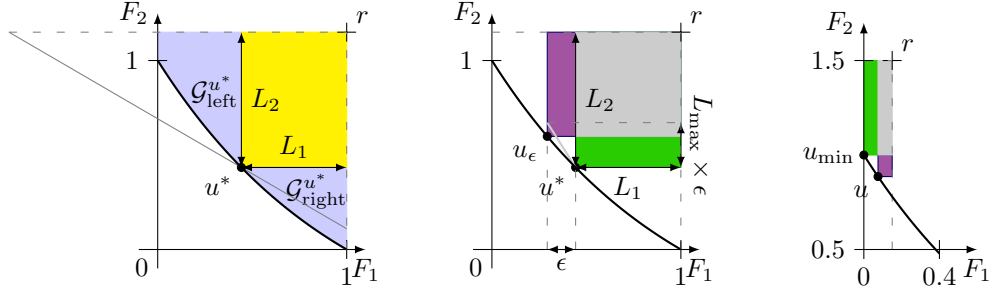


Fig. 8: Illustration of elements of the proofs of Proposition A.1 in the case $r_1 \leq x_{\max}$ (left and middle) and of Lemma B.1 (right). The Pareto front is described by $f(x) = \frac{e}{e-1} \times e^{-x} + 1 - \frac{e}{e-1}$ for $x \in [0, 1]$. Left: the gap regions $\mathcal{G}_{\text{left}}^{u^*}$ 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 .

1065 *Proof.* We consider the case where $r_1 \leq x_{\max}$. Let x^* be the first coordinate of u^* .
1066 We denote $L_1 := r_1 - x^*$ and $L_2 := r_2 - f(x^*)$ the lengths of the sides of the rectangle
1067 $\mathcal{D}_{u^*}^r$. For all $x, y \in [x_{\min}, x_{\max}]$, we have $|f(x) - f(y)| \geq L_{\min} \times |x - y|$. Additionally,
1068 since $r_1 \leq x_{\max}$, the segment $[x^*, x^* + L_1]$ is included in $[x_{\min}, x_{\max}]$. As a consequence,
1069 the section of the Pareto front on the right of u^* dominates the segment between u^* and
1070 $u^* + L_1 \times (1, -L_{\min})$, see the lefthand plot of Figure 8. Therefore, $\lambda(\mathcal{G}_{\text{right}}^{u^*})$ is larger than
1071 the area of the region of the objective space dominated by this segment, not dominated
1072 by u^* and dominating r , that is $\frac{1}{2} \times L_{\min} \times L_1^2$. For all $x, y \in [x_{\min}, x_{\max}]$, we also have
1073 $|f(x) - f(y)| \leq L_{\max} \times |x - y|$. Therefore, the part of the Pareto front on the left of u^* is
1074 dominated by the segment between u^* and $u^* + L_2 \times (-\frac{1}{L_{\min}}, 1)$, and $\lambda(\mathcal{G}_{\text{left}}^{u^*})$ is smaller
1075 than $\frac{1}{2} \times \frac{1}{L_{\min}} \times L_2^2$. We have yet to prove a lower bound on $\frac{L_1}{L_2}$. The vector u^* being
1076 different from u_{\min} , for $\epsilon > 0$ small enough, the vector $u_\epsilon := (x^* - \epsilon, f(x^* - \epsilon))$ belongs
1077 to the Pareto front. As we can see in the middle plot of Figure 8, $\text{HVI}_r(u^*, u_\epsilon)$ is smaller
1078 than $L_1 \times L_{\max} \times \epsilon$ and $\text{HVI}_r(u_\epsilon, u^*)$ is larger than $\epsilon \times (L_2 - \epsilon \times L_{\max})$. Additionally,
1079 u^* being an hypervolume maximizer, $\text{HVI}_r(u^*, u_\epsilon)$ is larger than $\text{HVI}_r(u_\epsilon, u^*)$, and thus
1080 $L_1 \times L_{\max} \geq L_2 - \epsilon \times L_{\max}$ for all $\epsilon > 0$. Taking the limit of this inequality when $\epsilon \rightarrow 0$,
1081 we obtain that $L_1 \times L_{\max} \geq L_2$. Combining the bounds on $\lambda(\mathcal{G}_{\text{left}}^{u^*})$ and $\lambda(\mathcal{G}_{\text{right}}^{u^*})$ with the
1082 lower-bound on $\frac{L_1}{L_2}$, we obtain the desired lower bound on $\lambda(\mathcal{G}_{\text{right}}^{u^*})$. We can obtain the
1083 symmetric inequality when $r_2 \geq f(x_{\min})$ by following the same approach. \square

1084 In particular, when f is bilipschitz and r dominates the nadir point, both bounds hold.
1085 We now prove the desired bounds on the normalized area of the gap regions $\mathcal{G}_{\text{left}}^{u^*}$ and
1086 $\mathcal{G}_{\text{right}}^{u^*}$.

1087 LEMMA A.2. Let u^* be a vector which maximizes the hypervolume with respect to
1088 a valid reference point r . If the Pareto front is described by a (L_{\min}, L_{\max}) -bilipschitz
1089 function f and the reference point r dominates the nadir point, both $\lambda(\mathcal{G}_{\text{left}}^{u^*})$ and $\lambda(\mathcal{G}_{\text{right}}^{u^*})$
1090 are between $(1 - \frac{1}{2} \times \frac{L_{\max}}{L_{\min}}) / (1 + \frac{L_{\max}^2}{L_{\min}^2})$ and $(1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}}) / (1 + \frac{L_{\min}^2}{L_{\max}^2})$.

1091 *Proof.* Let note arbitrarily \mathcal{G}_1 and \mathcal{G}_2 the two gap regions of the set $S = \{u^*\}$. By
1092 **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, by
1093 **Propositions 4.4** and **4.5**, the normalized maximum hypervolume $\max_{u \in \text{PF}_f} \text{HV}_r(u)$ over
1094 $\text{HV}_r(\text{PF}_f)$ 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 into
1095 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, $\lambda(\mathcal{G}_1)$
1096 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 $(1 - \frac{1}{2} \times \frac{L_{\min}}{L_{\max}}) \times \text{HV}_r(\text{PF}_f) -$
1097 $\frac{L_{\min}^2}{L_{\max}^2} \lambda(\mathcal{G}_1)$. Moving all the $\lambda(\mathcal{G}_1)$ terms on the same side and re-normalizing this side, we
1098 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

1099 **Appendix B. The nadir point is dominated by all the r_i^n corresponding to**
1100 **non-empty gap regions for n large.** We show in this section that for bilipschitz
1101 Pareto fronts, the nadir point is dominated by all the local nadir points r_i^n corresponding
1102 to non-empty gap regions, for n large enough. This result is stated in **Proposition B.2** and
1103 used in **Subsection 5.2**. It is equivalent to prove that the extreme vectors which dominate
1104 the reference point belong to the greedy set for n large enough.

1105 First, we prove in the next proposition that if $r_1 > x_{\max}$ (resp. $r_2 > f(x_{\min})$), then
1106 for r_2 (resp. r_1) close enough to $f(x_{\max})$ (resp. x_{\min}) the extreme vector u_{\max} (resp. u_{\min})
1107 is the only hypervolume maximizer, see the righthand plot of **Figure 8**. There are similar
1108 statements in [9] for the set of μ points maximizing the hypervolume, but they only apply
1109 to $\mu \geq 2$.

1110 **LEMMA B.1.** *We assume that the Pareto front is described by a function f which is*
1111 *(L_{\min}, L_{\max}) -bilipschitz and that the reference point r is valid. If $r_1 > x_{\max}$ and $f(x_{\max}) <$*
1112 *$r_2 < f(x_{\max}) + L_{\min} \times (r_1 - x_{\max})$, the right extreme of the Pareto front u_{\max} is the only*
1113 *maximizer of $\text{HV}_r(\cdot)$. Additionally, if $r_2 > f(x_{\min})$ and $x_{\min} < r_1 < x_{\min} + \frac{r_2 - f(x_{\min})}{L_{\max}}$,*
1114 *the vector $u_{\min} = (x_{\min}, f(x_{\min}))$ is the only maximizer of $\text{HV}_r(\cdot)$.*

1115 *Proof.* This proof is illustrated in the righthand plot of **Figure 8**. Let r be a reference
1116 point such that $r_2 > f(x_{\min})$ and $x_{\min} < r_1 < x_{\min} + \frac{r_2 - f(x_{\min})}{L_{\max}}$. Let $u = (x, f(x)) \neq u_{\min}$
1117 be a vector of the Pareto front which dominates r . The hypervolume improvement of u_{\min}
1118 to $\{u\}$ is $(r_2 - f(x_{\min})) \times (x - x_{\min})$. The hypervolume improvement of u to $\{u_{\min}\}$ is
1119 $(f(x_{\min}) - f(x)) \times (r_1 - x)$, which is smaller than $L_{\max} \times (x - x_{\min}) \times (r_1 - x_{\min})$ since u
1120 dominates r and f is (L_{\min}, L_{\max}) -bilipschitz. Since we assume that $L_{\max} \times (r_1 - x_{\min}) <$
1121 $r_2 - f(x_{\min})$, the upper bound on $\text{HVI}_r(u, u_{\min})$ is strictly smaller than $\text{HVI}_r(u_{\min}, u)$.
1122 Therefore, the hypervolume of u_{\min} is strictly larger than the one of u . We conclude that
1123 u_{\min} is the unique hypervolume maximizer. The symmetric result can be obtained with
1124 the same approach. \square

1125 It is left to prove that when $r_1 > x_{\max}$ (resp. $r_2 > f(x_{\min})$), the second coordinate of
1126 r_{n+1}^n (resp. the first coordinate of r_0^n) indeed converge to $f(x_{\max})$ (resp. x_{\min}). It is a
1127 straightforward consequence of **Lemma 5.6**. Therefore, we are able to conclude.

1128 **PROPOSITION B.2.** *We assume that the Pareto front is described by a bilipschitz func-*
1129 *tion f . Let $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ be a greedy set sequence relative to a valid reference point r . For*
1130 *n large enough, every local nadir point r_i^n corresponding to a non-empty gap region $\mathcal{G}_{\mathcal{S}_n, i}^r$*
1131 *dominates the nadir point.*

1132 *Proof.* By **Lemma 5.6**, $w_{n,r}^n$ converges to x_{\max} , and thus the right extreme local nadir
1133 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

1134 is strictly larger than x_{\max} , then there exists N such that for all $n \geq N$, r_{n+1}^n verifies
 1135 the assumptions on the reference point of [Lemma B.1](#) which guarantee that u_{\max} is the
 1136 unique maximizer of $\text{HV}_r(\cdot)$ over the right extreme gap region $\mathcal{G}_{\mathcal{S}_n, n+1}^r$. Let assume that
 1137 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
 1138 left extreme gap region $\mathcal{G}_{\mathcal{S}_n, i}^r$ is necessarily filled at some later iteration. When the right
 1139 extreme gap region is filled, u_{\max} , the unique minimizer of $\text{HV}_r(\cdot)$ over this gap region,
 1140 is added to the greedy set. To summarize, if $r_1 > x_{\max}$, then for n large enough \mathcal{S}_n
 1141 contains u_{\max} , and thus the right extreme gap region is empty. We can prove with the
 1142 same approach that for $r_2 > f(x_{\min})$, \mathcal{S}_n contains u_{\min} for n large enough.

1143 At any iteration, the non-extreme local nadir points dominate the nadir point. Ad-
 1144 ditionally, we proved that either $r_1 < x_{\max}$ (resp. $r_2 < f(x_{\min})$), and thus the left (resp.
 1145 right) extreme local nadir point dominates the nadir point or for n large enough, the left
 1146 (resp. right) extreme gap region is empty. \square