



# On the Extension of the DIRECT Algorithm to Multiple Objectives

Alberto Lovison<sup>1</sup> · Kaisa Miettinen<sup>2</sup>

Received: 1 March 2019 / Accepted: 7 August 2020 / Published online: 5 September 2020  
© The Author(s) 2020

## Abstract

Deterministic global optimization algorithms like Piyavskii–Shubert, DIRECT, EGO and many more, have a recognized standing, for problems with many local optima. Although many single objective optimization algorithms have been extended to multiple objectives, completely deterministic algorithms for nonlinear problems with guarantees of convergence to global Pareto optimality are still missing. For instance, deterministic algorithms usually make use of some form of scalarization, which may lead to incomplete representations of the Pareto optimal set. Thus, all global Pareto optima may not be obtained, especially in nonconvex cases. On the other hand, algorithms attempting to produce representations of the globally Pareto optimal set are usually based on heuristics. We analyze the concept of global convergence for multiobjective optimization algorithms and propose a convergence criterion based on the Hausdorff distance in the decision space. Under this light, we consider the well-known global optimization algorithm DIRECT, analyze the available algorithms in the literature that extend DIRECT to multiple objectives and discuss possible alternatives. In particular, we propose a novel definition for the notion of potential Pareto optimality extending the notion of potential optimality defined in DIRECT. We also discuss its advantages and disadvantages when compared with algorithms existing in the literature.

**Keywords** Global convergence · Multiobjective optimization · Multiple criteria optimization · DIRECT algorithm · Deterministic optimization algorithms

**Mathematics Subject Classification** 90C29 · 90C26 · 90C30

---

This paper deepens and utilizes some ideas in the extended abstract *Exact extension of the DIRECT algorithm to multiple objectives* [26]. This research was partly funded by the Academy of Finland (Grant no. 287496).

---

✉ Alberto Lovison  
lovison@math.unipd.it

Kaisa Miettinen  
kaisa.miettinen@jyu.fi

<sup>1</sup> Department of Mathematics “Tullio Levi-Civita”, Università di Padova, Via Trieste 63, 35121 Padova, Italy

<sup>2</sup> University of Jyväskylä, Faculty of Information Technology, P.O. Box 35 (Agora), FI-40014 University of Jyväskylä, Jyväskylä, Finland

## 1 Introduction

We consider the class of exact (or deterministic) global optimization algorithms and their possible extensions to the multiobjective optimization case. Among their characteristics, these algorithms exhibit appreciable convergence speed when the dimension of the decision space is moderate, but their specific strength is *global convergence*, the proved ability of approximating the global optimum of a black-box function with an arbitrary accuracy as e.g., in [22,30,32,35,36,40,41].

Since many real-world problems have multiple conflicting objectives to be optimized simultaneously, we concentrate here on multiobjective optimization. Such problems have so-called Pareto optima, where none of the objective functions can be improved without impairing some of the others.

Some exact single objective optimization algorithms have inspired or have been used as components in multiobjective optimization algorithms like [9,16,21,24,25]. Nevertheless, the class of exact, global multiobjective optimization algorithms has not been developed to the same extent as its single objective counterpart. Indeed, in most cases,

1. a single objective exact global algorithm is employed after the multiobjective optimization problem has been scalarized (e.g., [3,4,45]) or
2. an exact global algorithm is hybridized at some level with one or more non-deterministic principles (e.g., [4,21,42]).

As a result, these algorithms can suffer from the following drawbacks:

1. If scalarization is used, a (discrete) subset of the Pareto optimal set will be accurately approximated and the quality of the approximation depends on the number of scalarized problems solved.
2. If hybridization with heuristics is used, a wide subset of the set of Pareto optima can be approximated, but heuristics cannot guarantee optimality. Moreover, in general, the convergence rate will be slower than in scalarization based algorithms.

In this paper, we are interested in globally convergent deterministic multiobjective optimization algorithms. By saying that a multiobjective optimization algorithm is globally convergent, we mean that it should be able to produce an arbitrarily accurate representation of all the connected components of a globally Pareto optimal set, given a sufficiently large number of function evaluations.

As mentioned above, although global convergence and deterministic characters are desirable features, it is not easy to attain them both in algorithm design. Furthermore, the question of global convergence has not been considered in many papers in the domain of multiobjective optimization [4,12,19,23–25,31,41].

The first contribution of this paper is clarifying the notion of global convergence in the case of multiple objectives on the basis of the Hausdorff distance between sets. Indeed, the use of set-based distances like Hausdorff is not relevant for single objective optimization algorithms, because the global optimum for single objective optimization is a unique value in the objective space, corresponding to a set of minimizers in the decision space that in a typical case consists of a unique point, but can consist of more than one point or even an infinite number of points in pathological cases. Therefore, the Euclidean distance between the exact minimizer and an approximated minimizer is the only necessary concept for establishing a convergence speed.

This is not the case in multiobjective optimization, where in the generic case, the set of Pareto optima is a  $(k - 1)$ -dimensional submanifold of the objective space, if  $k > 1$  is the

number of objectives. To establish the convergence speed of a multiobjective optimization algorithm, it is necessary to measure the distance between the exact Pareto optimal set and the approximated Pareto optimal set. This issue is not addressed univocally in the literature.

The second contribution is an analysis and a discussion of the well-known DIRECT algorithm [20] and its extensions to multiple objectives. We focus on DIRECT because it has been recognized as an efficient algorithm for medium sized problems in the decision space. Indeed, DIRECT does not suffer much of the curse of dimensionality unlike, e.g., the Piyavskii–Shubert algorithm [33,37]. Furthermore, DIRECT does not require any global information (e.g., a global Lipschitz constant) to be globally convergent and, finally, it has been a starting point for many efficient single objective optimization algorithms.

In the original DIRECT algorithm, the feasible set is the  $d$ -dimensional hyperinterval  $D = [0, 1]^d \subseteq \mathbb{R}^d$ .  $D$  is covered by small hyperintervals  $D_1, \dots, D_N$ , where  $D_\ell = [a_1^{(\ell)}, b_1^{(\ell)}] \times \dots \times [a_d^{(\ell)}, b_d^{(\ell)}]$ , with  $0 \leq a_i^{(\ell)} < b_i^{(\ell)} \leq 1$ ,  $1 \leq i \leq d$ , such that  $\cup_{\ell=1}^N D_\ell = D$ . DIRECT contains two fundamental steps: a subdivision step and a selection step. In the division step, one of the small hyperintervals  $D_\ell$  of the covering of  $D$  is partitioned into smaller hyperintervals. In the selection step, the hyperintervals that are likely to contain the minimum are selected for further subdivision. The selection rule is based on a notion of potential optimality. This notion cannot be extended to multiple objectives in a straightforward way. We present, therefore, a review and a discussion of the existing extensions of DIRECT. Next, we propose a novel formulation for the notion of potential Pareto optimality, which is as faithful as possible to the original single objective version. We also discuss the strengths and weaknesses in comparison with the existing alternatives.

As far as the rest of this paper is concerned, we provide background in Sect. 2 by introducing the main concepts and notations used in the rest of the paper. Furthermore, we give a literature survey of deterministic, global multiobjective optimization algorithms, in particular, those based on DIRECT. For the convenience of the reader, we also briefly resume the single objective DIRECT algorithm. In Sect. 3, we discuss the notion of global convergence in single objective optimization, as it has been interpreted in the literature. Then we extend the focus to how it should be formulated to guarantee that an algorithm will accurately represent every possible connected component of a Pareto optimal set in multiobjective optimization. In Sect. 4, we discuss how the principles of DIRECT have been extended to multiple objectives in the literature and analyze possible alternatives. These ideas are collected and presented in the proposed *multiDIRECT* algorithm. In Sect. 6, we discuss the behavior of *multiDIRECT*, the multiobjective extension of the Piyavskii–Shubert algorithm mPS [25] and another DIRECT extension developed recently [1] on a simple but paradigmatic example defined in [19]. Finally, we conclude in Sect. 7.

## 2 Background

### 2.1 Some concepts of single and multiobjective optimization

There is a fundamental aspect that distinguishes the global optimization of a single function and several objective functions. Let us first consider the optimization problem with a single objective continuous function

$$f : D \rightarrow \mathbb{R}, \quad D = [0, 1]^d \subseteq \mathbb{R}^d, \quad \text{minimize}_{x \in D} f(x), \quad (1)$$

where  $D$  is called a *feasible set* in the decision space  $\mathbb{R}^d$  and there are no other constraints. Because of compactness, the set of globally optimal objective function values for problem (1) is a single value that often corresponds to a unique optimal point in the decision space.

In some cases, even an infinite number of optimal points can correspond to the unique optimal value. However, this situation can be proved to be nongeneric, i.e., for a dense set in a suitable set of functions, there exists only one optimal value with a unique corresponding optimal point in the decision space. This is a consequence of Sard’s theorem and Morse’s lemma [17]). Therefore, the global convergence of single objective optimization algorithms is established by requiring that the globally optimal value is approximated as accurately as desired, given an infinite number of function evaluations. In the generic case, this determines a unique optimum also in the feasible set  $D$ .

In the case of multiple continuous objectives, i.e.,

$$f : D \rightarrow \mathbb{R}^k, \quad D = [0, 1]^d \subseteq \mathbb{R}^d, \quad \underset{x \in D}{\text{minimize}} f(x), \quad (\text{MOP})$$

where  $f(x) = (f_1(x), \dots, f_k(x))$  is called an objective vector in the objective space  $\mathbb{R}^k$ , in general, it is not possible to completely order multidimensional vectors. Indeed, one vector can be better than another one when considering their first components, but the converse may happen when comparing the second components. Therefore, none of the two vectors can be said to be better than the other one. Only in special cases, say, pathological ones, there could exist a point which is the global optimum with respect to all the objective functions at the same time. In such cases, multiobjective optimization methods are not needed.

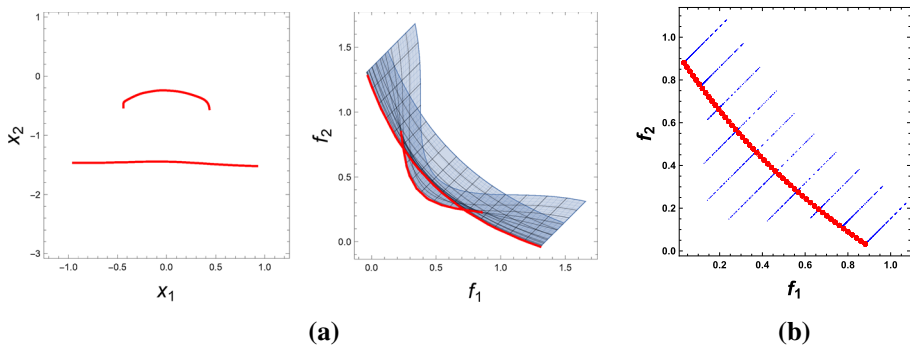
To deal more rigorously with optimality, it is necessary to consider the notion of *dominance* [18,29]. An objective vector  $z$  in  $\mathbb{R}^k$  *dominates* another objective vector  $w \in \mathbb{R}^k$ , and we write  $z < w$ , if  $z_i \leq w_i$  for all  $i$  and there exists  $j$  such that  $z_j < w_j$ .

We say that a point  $x$  is a (*global*) *Pareto optimum* if  $f(x)$  is nondominated by any other  $f(y)$ , for  $y$  in the feasible set  $D$ . The set of Pareto optima form a Pareto set  $P$ . The set of corresponding objective vectors, i.e.,  $\{f(x) : x \text{ is a Pareto optimum}\}$  is called a *Pareto front*. If a point  $x$  is nondominated by any other point in one of its neighborhoods, it is called a *local Pareto optimum*. If a Pareto optimum is not a minimizer for one of the objective functions considered individually, we call it a *tradeoff point*. If objective functions have been evaluated only for a finite number of points  $S = \{x_1, \dots, x_N\} \subseteq D$ , we can consider the subset of  $S$  consisting of the Pareto optimal points only with respect to  $S$  and not with respect to the entire  $D$ . These points cannot be guaranteed to be Pareto optimal, although they are used as an approximation of the Pareto set. We call this subset a *nondominated subset* of  $S$  and denote it by  $ND(S)$ . We recall also a widely used concept related to Pareto optimality [29, Definition 2.5.1]. A point  $x \in D$  and the corresponding objective vector  $f(x)$  are *weakly Pareto optimal* if there does not exist another point  $y \in D$  such that  $f_j(y) < f_j(x)$  for all  $j = 1, \dots, k$ .

## 2.2 Literature survey

In what follows, we discuss multiobjective optimization algorithms currently available in the literature that are either deterministic or related to the problem of finding the set of global Pareto optima. Special attention is dedicated to algorithms extending or involving DIRECT.

Normal Boundary Intersection (NBI) [10] is a deterministic algorithm capable of producing an evenly spaced discretization of the Pareto front by defining a family of scalar subproblems of the original multiobjective optimization problem. NBI is based on continuation, i.e., it finds connections among different Pareto optima. NBI starts from a Pareto optimal point, which is



**Fig. 1** Panel **a**: application of SICON [23] to  $L\&H_{2\times 2}$  [19]. The algorithm detects correctly all the different connected components of the Pareto set. Panel **b**: application of the NBI algorithm to the  $L\&H_{2\times 2}$  problem. The starting set is a regular grid. The algorithm correctly traces only one of the connected components of the Pareto optimal set, missing the second component. (Numerical experiment on MODEFRONTIER® courtesy by E. Rigoni at ESTECO s.r.l.).

the solution of one of the scalar subproblems, and identifies the local connected component of the Pareto set containing the starting point. In recent years NBI has been extended and generalized in other algorithms, see for instance [3]. As a drawback, in nonconvex cases, NBI possibly misses multiple isolated components, giving rise to an incomplete representation of the Pareto set.

This issue is observed in [10] and discussed in [23], where a continuation algorithm detecting correctly all the separated connected components, called SICON, is introduced. See Fig. 1a for an illustration of such a phenomenon for a nonlinear and nonconvex problem, called  $L\&H_{2\times 2}$  [19] (in both decision and objective spaces). In it, the Pareto set consists of two components, although their images are mapped in a single connected Pareto front.

Another deterministic continuation algorithm is proposed in [34] solving the first order Karush–Kuhn–Tucker (KKT) conditions for Pareto optimality. Points satisfying the KKT conditions are called *stationary*. Stationary points found are approximated by a collection of boxes obtained by an iterative subdivision of the feasible set of the problem. This algorithm has two main drawbacks: first it approximates only one connected component of the Pareto set and, second, the set of stationary points is a strict superset of the Pareto set.

In [12], the authors extend the ideas of [34] and propose an algorithm that, by using the properties of a suitably defined dynamical system, approximates a set of stationary points (as defined above). The output of the algorithm is a covering of a set composed by a collection of boxes. The authors discuss convergence towards a global Pareto set in a similar way as done in this paper by adopting the notion of a Hausdorff distance between sets. The main difference to our algorithm is that for the global convergence, the connectedness of the Pareto set is required and, therefore, multiple connected components of the Pareto set cannot be handled properly.

Singular Continuation (SICON) proposed in [23] is another continuation algorithm based on both the first and second order optimality conditions proposed by Smale in [38]. It has some similarities with [12]. To be more specific, SICON aims at approximating the set of stationary points. In addition, SICON takes into consideration the second derivatives and, therefore, stationary points can be separated in the multiobjective analogue of minima, maxima and saddle points. When compared with the previous strategies, SICON can produce a piecewise linear approximation [2] of the Pareto set in the form of a triangulated surface,

with a provably Newton-like superlinear accuracy. Furthermore, the algorithm allows for determining all the connected components of the Pareto set, provided that the starting sampling is dense enough. Indeed, in Fig. 1b, two connected components are detected, but if the sample of points had been sparser, the algorithm could have missed the smaller component. The SICON algorithm relies on a Delaunay tessellation of the feasible set and can suffer from the curse of dimensionality. Because of how it is defined, it is not possible to add new points and refine the structure automatically, although extensions in this direction are possible.

In [24,25], the question of global convergence is discussed on the basis of the Hausdorff distance in the decision space. In [24], an algorithm globally converging towards the substationary set is proposed, while in [25] an extension of the Piyavskii–Shubert globally convergent algorithm is introduced and the convergence towards the global Pareto set is proven assuming that a global Lipschitz constant is known for each objective function. Lipschitz optimization for univariate problems is considered for the bi-objective case in [41, Chapter 7], in [46,47], and for the multivariate case in [31]. In [19], the authors discuss the problem of correctly identifying different components of the Pareto set that may superimpose when nonconvex objective functions are involved. The aim is to produce globally correct representations of the Pareto set by means of triangulated surfaces, extending a previous nonglobal algorithm [18]. A deterministic approach making use of the Hausdorff distance in the objective space is called the Non Uniform Space Covering Method [14,15]. It provides an  $\varepsilon$ -approximation of the Pareto front in the sense that the Hausdorff distance between the Pareto front and its approximation is guaranteed to be smaller than  $\varepsilon > 0$ , given some suitable assumptions on the estimation of lower bounds. MultiGLODS [9] is a hybrid local-global deterministic algorithm combining Direct Multisearch [7], a derivative-free multiobjective local optimization algorithm and GLODS [8], a space-filling optimization method that manages efficiently multiple sequences possibly converging towards the same Pareto optimal point. Nevertheless, the authors focused on local convergence features and proved local convergence towards at least one Pareto optimal point.

Let us finally survey some algorithms related to the DIRECT algorithm, although some of them make partial use of heuristics and are not, thus, fully in our scope of studying deterministic algorithms. In [1,45], the authors propose three algorithms called MO-DIRECT, derived from DIRECT by employing either the ranks of the nondominated sorting [11], the hypervolume indicator or the nondominance (ND) concept in the objective space augmented by an extra dimension consisting of the hyperinterval radius. Several numerical performance tests have been implemented and used for measuring the capabilities of these algorithms. For the purposes of comparison with the algorithm proposed here, we notice that only ND does not make use of scalarizations. A proof of global convergence in the sense that the algorithm will produce an everywhere dense sampling of the feasible set is provided. The code is available and we provide a theoretical and numerical comparison with MO-DIRECT-ND in the subsequent sections.

A series of papers [5,6] discusses applications of an extension of DIRECT to damage identification problems with a numerical comparison with MO-DIRECT [1,45]. The algorithm proposed uses a scalarization of the multiobjective optimization problem by assigning to a vector  $v$  in the objective space the rank  $R(v)$  given by the nondominated sorting. The papers are application-oriented and the problem of global convergence is not considered.

A multiobjective extension of DIRECT (called MODIR) is proposed in [4] exhibiting an engineering application. The algorithm uses a non-explicit scalarization of objectives because it selects as potentially optimal those hyperintervals that are potentially optimal according to at least one of the objectives. This definition excludes the selection of tradeoff points. Nevertheless, the authors provide a proof of global convergence in the sense of everywhere

dense sampling in an infinite time. As an alternative formulation, the authors also propose a selection rule based on the nondominated set in the augmented feasible set using the hyperinterval radius as in MO-DIRECT-ND. The authors also propose a hybridization with a stochastic local search strategy called DFMA.

An evolutionary hybrid algorithm called NSDIRECT-GA is proposed in [42]. There, DIRECT is combined with MOGA and compared with NSGA (both evolutionary algorithms, see [11]).

### 2.3 The DIRECT algorithm

DIRECT [20] is quite a natural generalization of the Piyavskii–Shubert algorithm [33,37] and outperforms its ancestor in several senses. First, DIRECT does not require the existence of a global Lipschitz constant, which usually is not known, and second, DIRECT exhibits a faster convergence.

In the following subsections, we outline the working principles of DIRECT. This will make more comprehensible the ideas behind the *multi*DIRECT algorithm to be proposed. First, we consider a problem with a Lipschitz continuous single objective function  $f$  for which the Lipschitz constant is unknown.

#### 2.3.1 Selection rule

In DIRECT, the feasible set  $D$  is the  $d$ -dimensional hyperinterval  $D = [0, 1]^d \subseteq \mathbb{R}^d$ , covered by a collection of hyperintervals  $D_1, D_2, \dots$ , where  $D_j = [a_1^{(j)}, b_1^{(j)}] \times \dots \times [a_d^{(j)}, b_d^{(j)}]$ , with  $0 \leq a_i^{(j)} < b_i^{(j)} \leq 1, 1 \leq i \leq d$ . The hyperintervals  $D_1, D_2 \dots$  can overlap only along the faces. The radius  $\sigma_j$  of the  $j$ th hyperinterval  $D_j$  is defined as  $\sigma_j = \max_{i=1, \dots, d} \frac{b_i^{(j)} - a_i^{(j)}}{2}$ . Some of these hyperintervals are selected for subdivision on the basis of a Lipschitz-like criterion. Assuming that the function  $f$  has been evaluated at the point in the center of the  $j$ th hyperinterval, it is possible to estimate a lower bound for  $f$  by fixing a *rate of change*  $K > 0$  (which has a role similar to a Lipschitz constant) by setting  $\ell_j = f(x_j) - K\sigma_j$ , where  $x_j$  is the point at the center of the  $j$ th hyperinterval and  $\sigma_j$  is the radius of the hyperinterval. By ordering the sequence of  $\ell_j$  we obtain a ranking of the hyperintervals that are most likely to contain the global optimum of the function  $f$ . This ranking can vary when different rates of change  $K$  are chosen. If  $K$  is small, hyperintervals with low values of  $f(x_j)$  will be highly ranked, even if they have small  $\sigma_j$ . On the other hand, when  $K$  is large, hyperintervals with a large radius  $\sigma_j$  will be highly ranked.

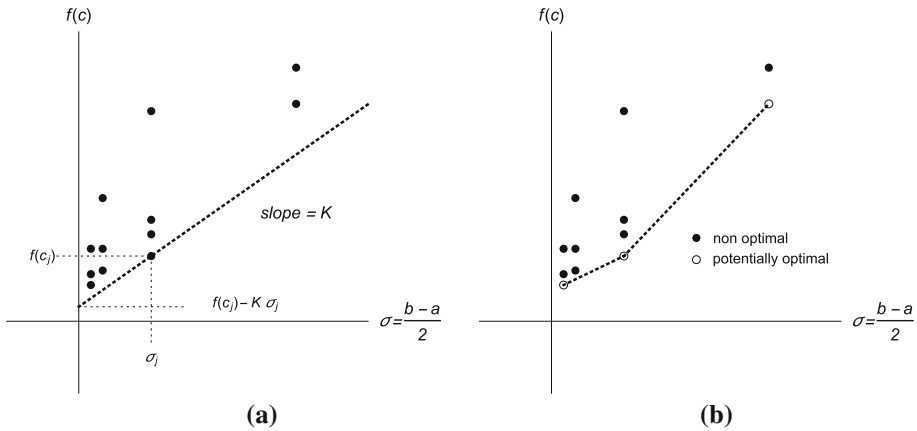
We call *potentially optimal* a hyperinterval  $j$  whose lower bound is optimal for a certain fixed rate of change  $K$ , i.e.,

$$f(x_j) - K\sigma_j \leq f(x_i) - K\sigma_i \quad \text{for all } i. \tag{2}$$

We can show that all potentially optimal hyperintervals correspond to points in the lower face of the convex hull of the following set of points as the rate of change  $K$  varies from 0 to  $\infty$ , as depicted in Fig. 2:

$$\left\{ (\sigma_j, f(x_j)) \mid j = 1, 2, \dots \right\} \subseteq \mathbb{R}^2. \tag{3}$$

We call the plane  $\{(\sigma, f) : \sigma, f \in \mathbb{R}\}$  as an *augmented plane of values*. At each iteration, DIRECT picks these potentially optimal hyperintervals on the convex hull and promotes them for further subdivision in smaller hyperintervals. It is typical that several hyperintervals are



**Fig. 2** DIRECT selection rule. **a** Function values evaluated at the center of a hyperinterval versus hyperinterval radius  $\sigma = \max_{i=1,\dots,d}(b_i - a_i)/2$ . The lower bound corresponding to a fixed rate of change  $K$  is the intersection of the line  $y = f(c_j) + K(\sigma - \sigma_j)$  with the  $y$  axis. **b** Hyperintervals corresponding to points lying on the lower, right-hand side of the convex hull are potentially optimal

to be subdivided, because as  $K$  varies, the lower bounds of each hyperinterval vary and, therefore, the ranking changes.

This selection rule indeed selects a subset of the Pareto set of a biobjective optimization problem, where the objective functions are the hyperinterval radius, to be maximized, and the function value, to be minimized. The subset in fact corresponds to the convex hull of the Pareto front for this problem.

For practical reasons, a further rule of selection is adopted: the lower bound computed should improve the current lowest value computed by at least a fixed quantity  $\epsilon$ , i.e.,

$$f(x_j) - K\sigma_j \leq f_{\min} - \epsilon |f_{\min}|. \tag{4}$$

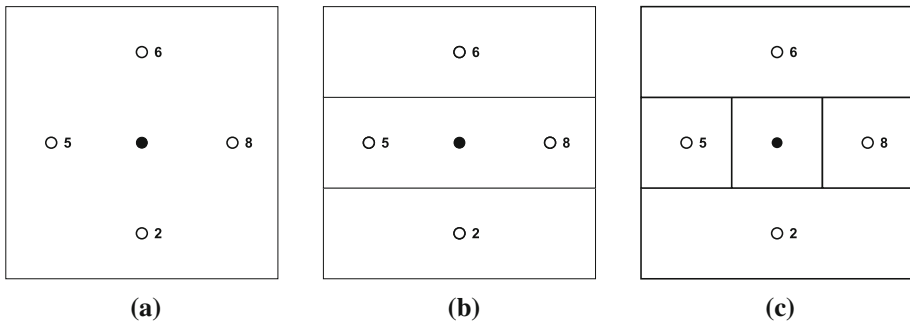
This extra rule prevents oversubdivision of the smallest hyperintervals which could easily lead to the depletion of the available computational resources without significantly improving the algorithm performance.

### 2.3.2 Division rule

A crucial aspect of the algorithm is the way hyperintervals are divided in smaller ones. The original algorithm proposes first to compute two extra test points along each coordinate direction of the feasible set, and then on the basis of the values computed, the hyperinterval is divided in thirds of equal size by first considering the directions where function values are more promising. So, if we call  $x_{i1}$  and  $x_{i2}$  the new points considered along the  $i$ th direction and  $f_{i1}$  and  $f_{i2}$  the corresponding function values, directions are ranked from the lowest to the highest  $\min\{f_{i1}, f_{i2}\}$ . Then, the most promising directions according to this criterion are divided in thirds at first, such that the best values are contained in the subintervals with the largest diameter. This feature improves the convergence speed, because best rated directions are more likely to be subdivided earlier which increases the density in the neighborhoods of test points with highest ranking. The division rule is exemplified in Fig. 3.

A further detail is that actually not all the directions are divided but only the ones corresponding to sides with the largest length. We notice that because of the division in thirds at





**Fig. 3** DIRECT division rule. **a** Two new points are computed along each direction. **b-c** Directions are ranked according to the values computed at the test points. More promising directions are divided first such that the resulting hyperintervals remain larger

every iteration, the possible radius sizes form a discrete set, although in an infinite time they will accumulate around zero. This mechanism reduces the computational complexity of the algorithm and at the same time guarantees that the hyperintervals will not become too small or skinny.

The ancestor hyperinterval is removed from the list of the radii and function values, while the new hyperintervals are incorporated. Therefore, a new diagram with radii  $\sigma_i$  and corresponding function values  $f_i$  as in Fig. 2a can be produced and the subsequent steps performed.

From these rules it can be proved that the DIRECT algorithm is convergent in a global sense. In other words, in an infinite time, the sequence of points becomes infinitely dense everywhere.

### 3 Point-wise and set-wise convergence

In a general nonconvex case, the Pareto front is an infinite set and may be composed of a number of separate connected components corresponding to different local Pareto optima (see Fig. 1a). We also refer to this kind of a problem as *multimodal*. The smoothness of the shapes usually observed for Pareto sets and fronts is not casual, but it can be proved that under reasonable regularity conditions, the set of Pareto optima in the feasible set is composed of a number of separate connected components. It can be proved that the connected components of the Pareto set are  $(k - 1)$ -dimensional submanifolds with a boundary and corners, called also *stratified sets* [27,28]. These connected components can superimpose and intersect each other when they are mapped on the Pareto front as in the example in Fig. 1 (for the explicit definition of the  $L\&H_{2 \times 2}$  function, see [19]).

We also observe that the intersection of the mapped components cannot be eliminated by a small perturbation of the functions. Therefore, unlike in the single objective case, there are nonpathological cases when more than one Pareto optimal point is mapped on the same Pareto optimal objective value. In other words, the correspondence between optimal points and optimal objective function values is not a bijection. Only in very special situations, i.e., pathological cases, the set of Pareto optima is a unique point in the feasible set, or a finite number of points [18,19,23,27].

Because the sets of global optima for single objective and multiobjective optimization problems are very different even from the point of view of geometry and topology, the corresponding concepts and criteria for global convergence of algorithms must take these differences into account. In the literature, the global convergence of multiobjective optimization algorithms is sometimes discussed in the same terms as single objective cases, i.e., by showing that an algorithm converges towards a (single) global Pareto optimum.

**Definition 1** A *deterministic algorithm*  $\mathcal{A}$  is a sequence of instructions that given a function  $f : D \rightarrow \mathbb{R}^k$ , produces a sequence of points  $x_1, \dots, x_n, \dots$ .

**Definition 2** If an algorithm  $\mathcal{A}$ , applied to a function  $f$  produces a sequence  $\{x_1, x_2, \dots\}$ , and  $x^* \in \text{cl}\{x_1, x_2, \dots\}$ , where  $\text{cl}(S)$  is the closure of the set  $S$  and  $x^*$  is a global Pareto optimum for  $f$ , we say that  $\mathcal{A}$  is *point-wise convergent*.

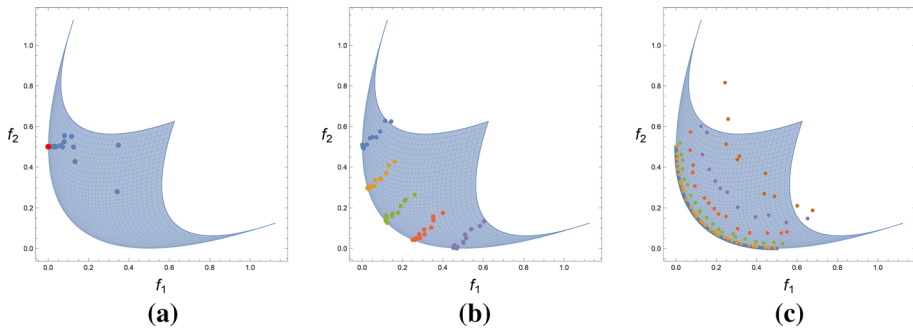
We refer to this property as *point-wise convergence*. Clearly, this property completely misses the global features of the set of Pareto optima discussed above. Indeed, we can give the following example.

**Example 1** Let  $f : D \rightarrow \mathbb{R}^k$ ,  $f(x) = (f_1(x), \dots, f_k(x))$ , and let  $\mathcal{A}$  be a single objective globally convergent optimization algorithm. If we apply  $\mathcal{A}$  to, e.g.,  $f_1$ , we obtain a sequence  $x_1, \dots, x_n, \dots$  having as a limit point one of the global optima  $x^*$  of the *scalar-valued* function  $f_1$ . Trivially,  $x^*$  is Pareto optimum, and the algorithm is also globally convergent in a point-wise sense. This example is illustrated in Fig. 4a, where the red dot is the limit point, which is also the global minimum for the first objective function, corresponding to the  $x$  axis.

The algorithm in Example 1 confuses the optimum of a single objective function with the set of Pareto optima of a multiobjective optimization problem. What is worse, this approach, which is completely unsatisfactory from the point of view of multiobjective optimization, is satisfactory in terms of point-wise convergence. Let us consider another example.

**Example 2** With the same hypotheses of Example 1, we consider a finite family of scalarizations  $g_\ell : D \rightarrow \mathbb{R}$ ,  $\ell = 1, \dots, m$ , defined by suitable linear convex combinations  $g_\ell(x) := \lambda_1^\ell f_1(x) + \dots + \lambda_k^\ell f_k(x)$ , with  $\lambda_1^\ell \geq 0, \dots, \lambda_k^\ell \geq 0$  and  $\lambda_1^\ell + \dots + \lambda_k^\ell = 1$ . If we apply  $\mathcal{A}$  to each  $g_\ell$ , we obtain a sequence  $x_1, \dots, x_n, \dots$  having as a limit point one of the global optima  $x_\ell^*$  of the *scalar* function  $g_\ell$ , which is a global Pareto optimum for the vector-valued function  $f$ . Because all of the  $x_\ell^*$  are global Pareto optima, this approach is also globally convergent in a point-wise sense. This example is illustrated in Fig. 4b, where the optimization of five different scalar subproblems is represented by dots of five different colors. An almost equally spaced sequence of five points on the Pareto front is obtained.

A well-known issue of the scalarization in Example 2 is that optima located in concave parts of the Pareto front cannot be approximated, but there are scalarizations like [10,44], that can overcome this problem. In particular, with the method in [10], the Pareto front can be covered as finely as desired by setting accordingly the number of subproblems. Nevertheless, as the number of scalar subproblems is fixed at the beginning, the density of points on the Pareto front does not change. Because the Pareto set is not a discrete and finite set, it is desirable that, at least in infinite time, the whole Pareto set becomes densely sampled. This feature is exemplified in Fig. 4c. An optimization algorithm should produce at every new iteration (each iteration is marked with a different colour) a new set of points approximating



**Fig. 4** Globally failing point-wise convergence: **a** optimization is carried on considering only one function, converging towards a (possibly global) optimum of the selected function. **b** A priori subdivision of the original problem in a fixed number of subproblems. Standard optimization carried on separate scalar subproblems leads to point-wise convergence towards a discrete subset of the Pareto set. Global set-wise convergence: **c** The approximated Pareto set becomes progressively denser and closer to the exact Pareto set with respect to the Hausdorff distance between sets. Notice that the Hausdorff distance is measured in the feasible set while the objective space is shown here

the Pareto set. The points of this set should be closer to the Pareto set at every iteration, but also the number of points should increase and become more dense.

Based on the observations discussed, it seems more appropriate to consider a *set-wise approach*, i.e., the set of Pareto optima should not be approximated in a-point-at-a-time fashion, as in multi-start algorithms, but it should be approximated as a whole by means of a sequence of approximating sets of points. Moreover, because it can happen that the same value in the Pareto front is the image of two or more different points in the feasible set (see the example in Fig. 1b), it is important to study the approximating sequences of sets in the feasible set. This is preferable also because the Pareto set in the feasible set usually has a regular manifold structure that is not guaranteed to be maintained in the objective space. A suitable concept of distance between sets is needed, and a natural choice is the Hausdorff distance.

A *set-wise global convergence* is realized if a multiobjective optimization algorithm produces at every iteration  $n$  a candidate set  $S_n$  approximating the whole set of Pareto optima. In fact, the algorithm can produce one or a few points at every iteration. Nevertheless, a candidate set can be defined or updated at every iteration by selecting the nondominated points among the whole sequence of new and old points evaluated since the algorithm started. Consistently, a definition for global convergence can be given as follows:

**Definition 3** Let  $P \subseteq D$  be the set of Pareto optima of problem (MOP). Consider an optimization algorithm producing the sequence of approximating sets  $S_1, S_2, \dots$ . We say that the algorithm *converges globally* for the function  $f$  if

$$\lim_{n \rightarrow \infty} d_{\mathcal{H}}(S_n, P) = 0, \tag{5}$$

where  $d_{\mathcal{H}}$  stands for the Hausdorff distance.

### 4 multiDIRECT: an extension of DIRECT to multiple objectives

Earlier, a multiobjective extension of the Piyavskii–Shubert algorithm [33,37] called *multiPS* has been proposed in [25]. Its global convergence in the sense discussed in Sect. 3 can be

proven. However, as mentioned in the introduction, the Piyavskii–Shubert algorithm has shortcomings that the DIRECT algorithm does not share.

In this section, we address the problem of extending the strengths of DIRECT to multiple objectives, focusing on the selection rule (2), (4) and on the division rule defined in Sect. 2.3.2. We propose the extension to multiple objectives in an as straightforward manner as possible for both selection and division rules. Before we can introduce the actual new algorithm, we must present some theoretical considerations.

### 4.1 Potential Pareto optimality

In what follows, we consider a feasible set  $D$  covered by a set of non-overlapping hyperintervals  $H_1, \dots, H_N$ . To each hyperinterval  $H_i = [a_1^i, b_1^i] \times \dots \times [a_d^i, b_d^i]$ , we associate the coordinates of its center  $c := (c_1, \dots, c_d)$ , the vector of radii  $\left(\sigma_1 = \frac{b_1^i - a_1^i}{2}, \dots, \sigma_d = \frac{b_d^i - a_d^i}{2}\right)$ , and a vector of objective functions values  $f(c) := (f_1(c), \dots, f_k(c))$ , computed at the center point  $c$  of the hyperinterval. In other words, we refer to a hyperinterval by implicitly considering the triplet  $(c, f(c), \sigma)$ . When referring to the radius  $\sigma$  of a hyperinterval, we refer to  $\max\{\sigma_1, \dots, \sigma_d\}$ . More details will be given in Sect. 4.5 related to the division rule. By recalling the definition of potential optimality (2), we notice that we first need an extension of the notion of lower bound:

**Definition 4** Let  $H$  be a hyperinterval with  $(c, f(c), \sigma)$ . We denote  $F = f(c)$ . Given a sequence  $A = (\alpha_1, \dots, \alpha_k)$  of rates of change, one for each objective function, the corresponding *vector lower bound* is given by

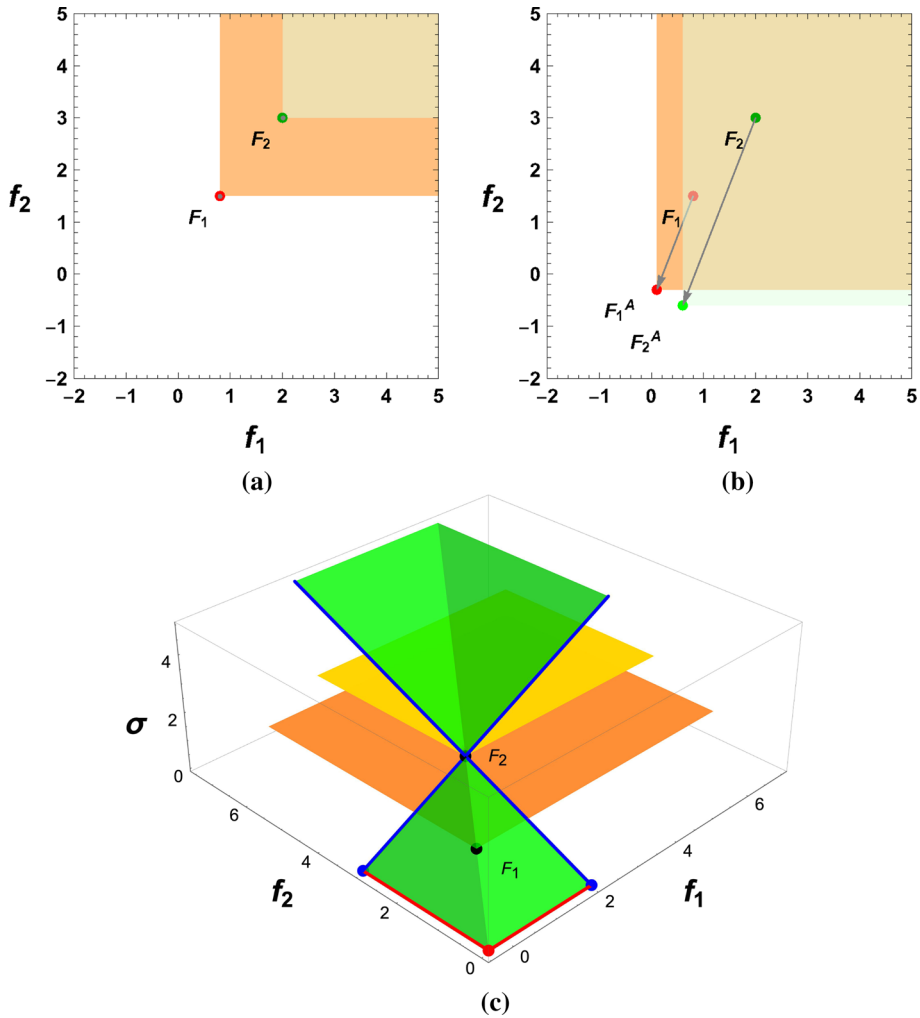
$$F^A := F - A\sigma = (f_1(c) - \alpha_1\sigma, \dots, f_k(c) - \alpha_k\sigma). \tag{6}$$

Therefore, a multiobjective version of the notion of potential optimality can be obtained from (2) by replacing “lower bound” with “vector lower bound” and “optimality” with “Pareto optimality”. In analogy with the single objective DIRECT, we augment the space of the objectives  $\{(f_1, \dots, f_k)\}$  with an extra dimension representing the radius  $\sigma$  of a hyperinterval  $H$ . We refer to a vector in this *augmented objective space* as an *augmented objective vector*  $(F, \sigma) = (f_1, \dots, f_k, \sigma) \in \mathbb{R}^{k+1}$  of a hyperinterval  $H = (c, F, \sigma)$ .

**Definition 5** A hyperinterval  $H_i = (c_i, F_i, \sigma_i)$  is *potentially Pareto optimal* if there exists a sequence  $A = (\alpha_1, \dots, \alpha_k)$  of rates of change such that the vector lower bound  $F_i^A := F_i - A\sigma_i = (f_1(c_i) - \alpha_1\sigma_i, \dots, f_k(c_i) - \alpha_k\sigma_i)$  is nondominated in the set of vector lower bounds corresponding to the same sequence of rates of change:

$$\left\{ F_j^A = F_j - A\sigma_j \mid j = 1, \dots, N, i \neq j \right\}.$$

**Example 3** Consider only two hyperintervals  $H_1, H_2$ , with objective vectors  $F_1, F_2 \in \mathbb{R}^2$ , with  $F_1$  dominating  $F_2$ , and different radii  $\sigma_1 < \sigma_2$  as in Fig. 5a. A set of rates of changes  $A := (\alpha_1, \alpha_2), \alpha_{1,2} \geq 0$  produces vector lower bounds  $F_1^A = F_1 - A\sigma_1$  and  $F_2^A = F_2 - A\sigma_2$ . It is clear that if  $\alpha_1, \alpha_2$  are sufficiently small, or even zero,  $F_1^A$  still dominates  $F_2^A$ , while for large values of  $\alpha_1$  and (or)  $\alpha_2$  with  $\sigma_2 > \sigma_1$ ,  $F_2^A$  dominates (or at least will not be dominated by)  $F_1^A$  (see Fig. 5b). Therefore, both  $H_1$  and  $H_2$  are potentially Pareto optimal.



**Fig. 5** Potential Pareto optimality for two hyperintervals  $H_1$  dominating  $H_2$ , but with  $\sigma_2 > \sigma_1$ . **a** For small values of  $A = (\alpha_1, \alpha_2)$  we have  $F_2^A \prec F_1^A$ . **b** On the other hand, for sufficiently large  $A$ , the lower bound  $F_2^A$  is no more dominated by  $F_1^A$ . Therefore, both the hyperintervals  $H_1$  and  $H_2$  are potentially Pareto optimal. **c** A corresponding situation in the augmented objective space. Orange and yellow planes are the  $\sigma$ -levels of  $H_1$  and  $H_2$ , respectively. The red line is the virtual Pareto front, while the green cone is the dominance cone. The blue lines are the dominance rays. Dominance rays trivially do not intersect higher  $\sigma$ -levels, because there are no higher  $\sigma$ -levels at all

**Remark 1** Let us extend Example 3 to an arbitrary set of hyperintervals  $\{H_j \mid j = 1, \dots, N\}$ . We partition the hyperintervals in classes corresponding to the same radius  $\sigma$ , called  $\sigma$ -levels. Obviously, all nondominated objective vectors correspond to potentially Pareto optimal hyperintervals. Indeed, they remain nondominated for small values of the rates of change  $(\alpha_1, \dots, \alpha_k)$ . At the same time, let us consider hyperintervals in the  $\sigma_M$ -level, with  $\sigma_M$  being the maximal radius of the hyperintervals. Then, also the nondominated hyperintervals in the  $\sigma_M$ -level are potentially Pareto optimal, because for sufficiently large values of  $(\alpha_1, \dots, \alpha_k)$  they surpass all the vector lower bounds of hyperintervals with a smaller radius.

**Example 4** Consider three hyperintervals  $H_1, H_2, H_3$  with objective vectors  $F_1, F_2, F_3 \in \mathbb{R}^2$ , with  $F_1 \prec F_2$ , and  $F_2 \prec F_3$  and different diameters  $\sigma_1 < \sigma_2 < \sigma_3$  as in Fig. 6. Because of the above observations, it is clear that  $H_1$  and  $H_3$  are potentially Pareto optimal, corresponding to very small and very large values of the rates of change  $\alpha_1$  and  $\alpha_3$ , respectively. Let us then consider  $H_2$  to determine if some  $(\alpha_1, \alpha_2)$  exist such that the vector lower bound is not dominated by lower bounds corresponding to other points.

As can be observed in Fig. 6a, for values of  $\alpha_1, \alpha_2$  being large enough, the vector lower bound of  $H_2$  surpasses the vector lower bound of  $H_1$ . The problem is that at the same time, the corresponding vector lower bound of  $H_3$  can surpass  $F_2^A$ , vanishing its attempt to become nondominated. This happens for  $A = (\alpha_1, \alpha_2)$  close to the line connecting  $F_1$  to  $F_2$ . Nevertheless, we observe in Fig. 6b, that there are directions, close to the vertical and horizontal lines, where there is still a possibility for  $F_2^A$  to become nondominated.

All of these possibilities are more understandable by examining the animation available at <http://www.mit.jyu.fi/optgroup/extramaterial.html>.

The definition of potential Pareto optimality requires to check for infinite values of the rates of change  $(\alpha_1, \dots, \alpha_k)$ . In the standard DIRECT, it is possible to characterize geometrically the condition of potential optimality, as illustrated in Fig. 2, with the convex hull of the set  $\{(\sigma_i, f_i)\}$  in the augmented plane and reduce the problem to a finite number of checks.

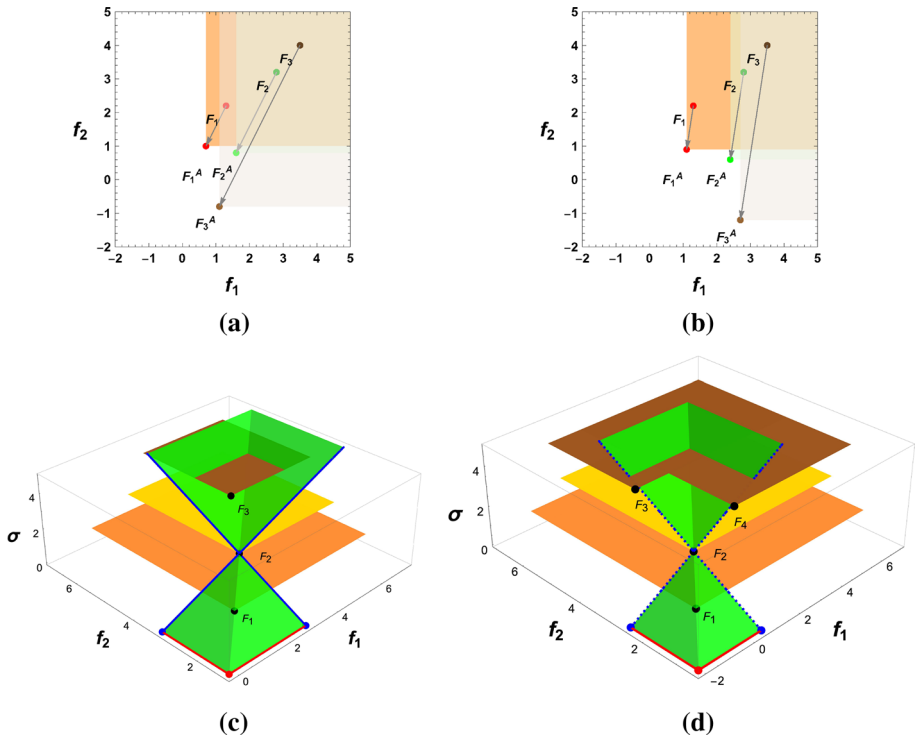
**Question 1** *Is it possible to characterize geometrically the set of potentially Pareto optimal hyperintervals?*

In the literature, there have been some attempts to extend the notion of potential optimality. In [4], a hyperinterval is called potentially optimal if it is potentially optimal for at least one of the objectives. In [1], the set of potentially Pareto optimal hyperintervals is approximated by a set of nondominated vectors in the augmented objective space. In a previous paper [26], we have considered the convex hull of the set of nondominated vectors in the augmented objective space. As we will see in the following, the notions of [4] and [26] are too restrictive, while the one in [1] is too conservative.

### 4.2 Selection rule

Considering Example 3, we visualize hyperintervals in the augmented objective space by adding  $\sigma$  as extra dimension. In this framework, the set of vector lower bounds  $F_i^A := F^A(c_i) = (f_1(c_i) - \alpha_1\sigma_i, f_2(c_i) - \alpha_2\sigma_i), i = 1, 2$ , corresponds to the intersections of the lines  $\sigma \mapsto (f_1(c_i) - \alpha_1(\sigma_i - \sigma), f_2(c_i) - \alpha_2(\sigma_i - \sigma), \sigma) \in \mathbb{R}^3$  with the hyperplane  $\{\sigma = 0\} \subset \mathbb{R}^3$ . If the values  $(\alpha_1, \alpha_2)$  are small, the lines are steep and the intersections are close to the original values, and  $F_1^A$  dominates  $F_2^A$ . On the other hand, if the values  $(\alpha_1, \alpha_2)$  are sufficiently large, i.e., the lines are closer to the horizontal plane and  $F_2^A$  will be no more dominated by  $F_1^A$ .

**Definition 6** Let  $\widehat{H} = (\widehat{c}, \widehat{F}, \widehat{\sigma})$  and  $A = (\alpha_1, \dots, \alpha_k)$ . We denote by  $\widetilde{F}(A, \sigma) \in \mathbb{R}^k$  the line given by  $\sigma \mapsto \widetilde{F}(A, \sigma) := \widehat{F} + A(\sigma - \widehat{\sigma})$ . With this notation, the vector lower bound  $\widehat{F}^A$  is given by  $\sigma = 0$ , i.e.,  $\widetilde{F}(A, 0) = \widehat{F} - A\widehat{\sigma}$ . If  $\widehat{F}$  is dominated by an  $F_i$  with  $\sigma_i < \widehat{\sigma}$ , we call a *dominance ray* the line in  $\mathbb{R}^{k+1}$  connecting  $(\widehat{F}, \widehat{\sigma})$  to  $(F_i, \sigma_i)$ . With the notation above, the dominance ray is defined by setting  $A = \frac{F_i - \widehat{F}}{\sigma_i - \widehat{\sigma}}$ .



**Fig. 6** Potential Pareto optimality for three hyperintervals on different  $\sigma$ -levels. **a** Vector lower bounds for  $H_1$  and  $H_3$  both dominate the vector lower bound for  $H_2$  for rates of changes  $\alpha_1$  and  $\alpha_2$  with the same magnitude. **b** If  $\alpha_1 \ll \alpha_2$ , the vector lower bound for  $H_2$  is no more dominated. **c** Correspondingly, in the augmented space we see that some of the dominance rays intersect the higher  $\sigma$ -levels in points that are nondominated in the  $\sigma$ -level. **d** If there are two objective vectors  $F_3$  and  $F_4$  in the upper level dominating the intersection of the dominance cone of  $H_2$  with their  $\sigma$ -level, then  $H_2$  is not potentially Pareto optimal. Notice that the most important rays (in blue) are at the extrema of the virtual Pareto front (i.e., the red line). Colors are available in the online version of the paper

$$\sigma \mapsto (\tilde{F}(A, \sigma), \sigma) = (\tilde{F}_i(\sigma), \sigma) := \left( \hat{F} + \frac{F_i - \hat{F}}{\sigma_i - \hat{\sigma}}(\sigma - \hat{\sigma}), \sigma \right). \tag{7}$$

Note that  $\tilde{F}_i(\hat{\sigma}) = \hat{F}$  and  $\tilde{F}_i(\sigma_i) = F_i$ , as claimed. Moreover  $\tilde{F}_i(0)$  is the vector lower bound both for  $\hat{F}$  and  $F_i$  when  $A$  as above is used.

**Proposition 1** A hyperinterval  $\hat{H}$  is potentially Pareto optimal if and only if there exists a set of rates of change  $A = (\alpha_1, \dots, \alpha_k)$  such that the intersection of the ray  $\sigma \mapsto (F^A(\sigma), \sigma) := (\hat{F} + A(\sigma - \hat{\sigma}), \sigma)$  with every  $\sigma$ -level is nondominated by the objective vectors  $F_i$  corresponding to that  $\sigma$ -level, i.e., such that  $\sigma_i = \sigma$ .

**Proof** We observe that  $\hat{H}$  is potentially Pareto optimal if and only if there exists  $A$  such that  $\hat{F}^A$  is nondominated in the set of lower bounds associated with  $A$ , i.e.,  $\{F_i^A = F_i - A\sigma_i \mid i = 1, 2, \dots\}$ . Moreover,

$$\hat{F}^A \prec F_i^A \iff \hat{F} - A\hat{\sigma} \prec F_i - A\sigma_i \iff \hat{F} + A(\sigma_i - \hat{\sigma}) \prec F_i. \tag{8}$$

But  $\widehat{F} + A(\sigma_i - \widehat{\sigma})$  is exactly the intersection with the  $\sigma_i$ -level of the line  $\sigma \mapsto (\widetilde{F}(A, \sigma), \sigma)$  passing through  $(\widehat{F}, \widehat{\sigma})$ . The claim follows immediately.  $\square$

We introduce a subset of the objective space very important for the geometric characterization of potential Pareto optimality.

**Definition 7** Let  $E = \left\{ v \in \mathbb{R}^k \mid \widetilde{F}_j(0) \leq v \leq \widehat{F} \text{ for some } j \in \{1, \dots, t\} \right\}$ . We refer to the following set as a *virtual Pareto front*:

$$VP[\widehat{F}] := \left\{ (v, 0) \in \mathbb{R}^{k+1} \mid v \in E, v \text{ is weakly Pareto optimal in } E \right\}. \tag{9}$$

The cone of lines connecting  $(\widehat{F}, \widehat{\sigma})$  with the virtual Pareto front is called *dominance cone* of  $\widehat{H}$ .

In Figs. 5c, 6c, d, and 7b, the virtual Pareto front is highlighted in red, while the dominance cone is the green surface (colors available in the online version of the paper).

Proposition 1 requires that a high-dimensional infinite set of possible rates of change  $A = (\alpha_1, \dots, \alpha_k)$ ,  $\alpha_i \geq 0$ , is checked in order to determine whether  $\widehat{H}$  is potentially Pareto optimal. Nevertheless, it suffices to check only for  $A$  in a lower-dimensional subset.

**Proposition 2** A hyperinterval  $\widehat{H}$  is potentially Pareto optimal if and only if there exists  $A = (\alpha_1, \dots, \alpha_k)$  such that  $(A, \widehat{\sigma})$  is in the dominance cone of  $\widehat{H}$ , and the intersection of the ray  $\widetilde{F}^A(\sigma) := \widehat{F} + A(\sigma - \widehat{\sigma})$  with every  $\sigma$ -level is nondominated in the  $\sigma$ -level, for every level  $\sigma > \widehat{\sigma}$ .

**Proof** If such a ray  $A \in \mathbb{R}^k$  existed, it would be possible to find one  $\widetilde{A}$  in the dominance cone by picking the ray passing by the intersection of the virtual Pareto front and the line connecting  $\widehat{F}^A$  and  $\widehat{F}$ . Indeed, consider  $\lambda A$  in place of  $A$ , with  $0 < \lambda < 1$ . By continuity, there exists a neighborhood of 1 such that  $\widetilde{F}^{\lambda A}$  is still nondominated, as well as for the intersections of the dominance ray with the  $\sigma$ -levels with  $\sigma > \widehat{\sigma}$ . The infimum of such values of  $\lambda$ , by definition, corresponds to a point in the virtual Pareto front.  $\square$

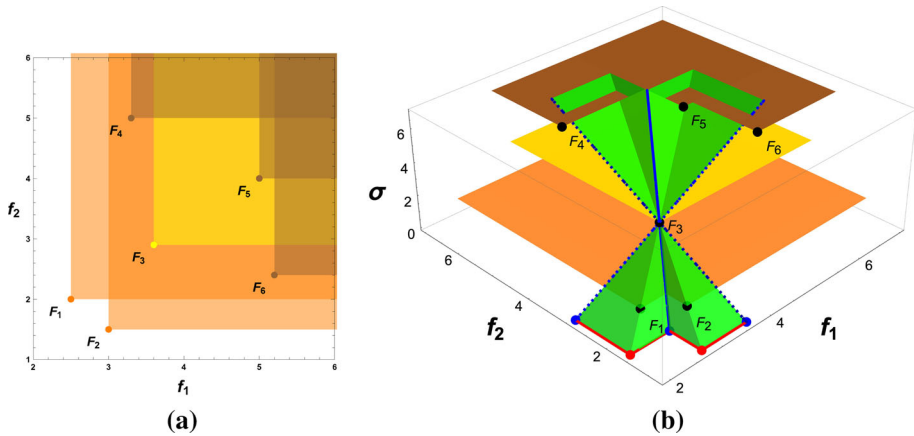
Proposition 1 is the multiobjective extension of the geometric criterion in (2). As equation (2) requires an infinite number of checks, i.e., all values  $\alpha$ ,  $0 < \alpha < \infty$ , also Propositions 1 and 2 would require to test an infinite number of  $A$ 's, at least to say that  $\widehat{H}$  is not potentially Pareto optimal. In the single objective case, it is possible to limit oneself to the slopes of the lines connecting points in the convex hull of the augmented function values, therefore to reduce to a finite number of tests. This is possible also in the multiobjective case.

**Definition 8** *Tipping vectors*  $Tip$  are elements in the virtual Pareto front which are non-dominated if the dominance relation is reversed by using  $>$  in place of  $<$ , i.e.,  $Tip = -ND(-VP[\widehat{F}])$ .

**Remark 2** In the biobjective case, tipping vectors are the concave vertices of the virtual Pareto front, plus the two extrema. In Figs. 5, 6, 7, the tipping vectors and the corresponding dominance rays are highlighted in blue (colors are available in the online version of this paper). Solid blue lines correspond to intersections that are nondominated in all the higher  $\sigma$ -levels, while dashed lines correspond to dominated intersections in some of the higher  $\sigma$ -levels.

**Theorem 1** An hyperinterval  $\widehat{H}$  is potentially Pareto optimal if and only if there exists a tipping vector  $T$  such that the ray connecting  $(\widehat{F}, \widehat{\sigma})$  with  $T$  is not intersecting the regions dominated by the remaining vectors  $F_i$  for each of the  $\sigma$ -levels with  $\sigma > \widehat{\sigma}$ .





**Fig. 7** Geometric characterization of potential Pareto optimality for six hyperintervals. Here,  $F_1$  and  $F_2$  belong to the same  $\sigma$ -level, as well as  $F_4, F_5$  and  $F_6$ , i.e.,  $\sigma_1 = \sigma_2 < \sigma_3 < \sigma_4 = \sigma_5 = \sigma_6$ . Since  $F_1$  and  $F_2$  are nondominated, they are potentially Pareto optimal. While  $F_4, F_5$  and  $F_6$  are nondominated in the higher  $\sigma$ -level, they are also potentially Pareto optimal, according to Remark 1. The problematic hyperinterval is  $H_3$ . By inspecting the tipping vectors (blue dots) and the respective dominance rays we see that the central ray intersects the higher  $\sigma$ -level in a nondominated point. Hence, also  $H_3$  is potentially Pareto optimal

**Proof** For every vector  $\widehat{F} - A\hat{\sigma}$  in the virtual Pareto front, there is (at least) a tipping vector  $T$  such that  $\widehat{F} - A\hat{\sigma} < T$ . Concerning the intersections of the corresponding dominance cones with higher  $\sigma$ -levels, they are  $F_\sigma^A = \widehat{F} + A(\sigma - \hat{\sigma})$  and  $T_\sigma = \widehat{F} + (T - \widehat{F}) \frac{\sigma - \hat{\sigma}}{0 - \hat{\sigma}}$ . Explicit calculation with  $\sigma > \hat{\sigma}$  and  $\widehat{F} - A\hat{\sigma} < T$  shows that  $T_\sigma < F_\sigma^A$  and, therefore, if  $F_\sigma^A$  is nondominated in the  $\sigma$ -level, also  $T_\sigma$  is nondominated.  $\square$

**Example 5** In Example 4, all hyperintervals are potentially Pareto optimal. This appears also in Fig. 6c, where we observe that the dominance rays corresponding to two tipping vectors are nondominated in the higher  $\sigma$ -level (solid blue lines; colors available online). Consider four hyperintervals  $H_1, H_2, H_3, H_4$ , with objective values  $F_1, F_2, F_3, F_4 \in \mathbb{R}^2$ , with  $F_1 < F_2, F_2 < F_3$  and  $F_2 < F_4$  as in Fig. 6d. In this case,  $H_2$  is not potentially Pareto optimal. Indeed, all the dominance rays passing by vectors in the virtual Pareto front are dominated in the higher  $\sigma$ -level (dashed blue lines; colors available online).

**Example 6** Consider six hyperintervals  $H_1, H_2, H_3, H_4, H_5, H_6$ , with  $F_1, F_2 < F_3, F_3 < F_5$  as in Fig. 7, with  $\sigma_1 = \sigma_2 < \sigma_3 < \sigma_4 = \sigma_5 = \sigma_6$ . In this case,  $H_3$  is potentially Pareto optimal, because the dominance ray (solid blue line) corresponding to the central tipping vector does not intersect the dominated region in the higher  $\sigma$ -level. The dominance rays of the remaining tipping vectors are dominated in the higher  $\sigma$ -level (dashed blue lines).

### 4.3 Supplementary condition for potential Pareto optimality

An important condition in the definition of potential Pareto optimality is (4), which depends on a small parameter  $\epsilon$ . However, according to the authors of DIRECT ([20]), the algorithm seems quite insensitive on the magnitude of the parameter, provided it is small when compared to the magnitudes of the ranges of the objective function values. This parameter dependent condition is critical for avoiding that the algorithm wastes excessive computational resources

on exploring regions with very small potential improvement. This condition, therefore, is important for efficiency.

As far as we know, in the literature, condition (4) of DIRECT has never been adapted and implemented for multiobjective cases. If we denote by  $|F|$  the vector of absolute values of the components of  $F$ , i.e.,  $|F| := (|f_1(c)|, \dots, |f_k(c)|)$ , we require the vector lower bound  $\widehat{F}^A$  being nondominated by the objective vectors  $F_i$  diminished by a quantity proportional to  $\epsilon$  and  $|F_i|$ :

$$\begin{aligned} \widehat{F}^A \text{ is nondominated by the set } & \left\{ F_i - \epsilon |F_i| \mid i = 1, \dots, N \right\} \\ := & \left\{ (f_1(c_i) - \epsilon |f_1(c_i)|, \dots, f_k(c_i) - \epsilon |f_k(c_i)|) \mid i = 1, \dots, N \right\}. \end{aligned} \tag{10}$$

Operatively, this has no effect on the vectors belonging to the  $\sigma$ -level with maximal  $\sigma$ . For the vectors belonging to the intermediate  $\sigma$ -levels, it is necessary to substitute the tipping vectors  $T$  in Theorem 1 with

$$T \longrightarrow T + \epsilon |T|. \tag{11}$$

This does not implement exactly condition (10), but it appears as a valid approximation. Notably, condition (10) implies that also nondominated vectors belonging to the lowest  $\sigma$ -level, which were automatically selected as potentially Pareto optimal, should also be filtered. For these points, however, there is no tipping vector defined, because there are no dominating vectors from which to compute a virtual Pareto front. Therefore, we use the objective vector  $\widehat{F}$  as a tipping vector, then we proceed as above by subtracting a vector proportional to  $|F|$  and  $\epsilon$ , as for the previous case. The dominance ray is the line passing through  $(\widehat{F}, \widehat{\sigma})$  and  $(\widehat{F} - \epsilon |F|, 0)$ . If the dominance ray intersects all the higher  $\sigma$ -levels in a vector that is nondominated for all possible  $\sigma$ , then the hyperinterval  $\widehat{H}$  is potentially Pareto optimal.

As we will see in Sect. 6, this condition becomes more and more influential in the later iterations. There, a large part of the points on the Pareto front, which becomes denser and denser, is excluded from selection because the points do not bring a relevant potential improvement.

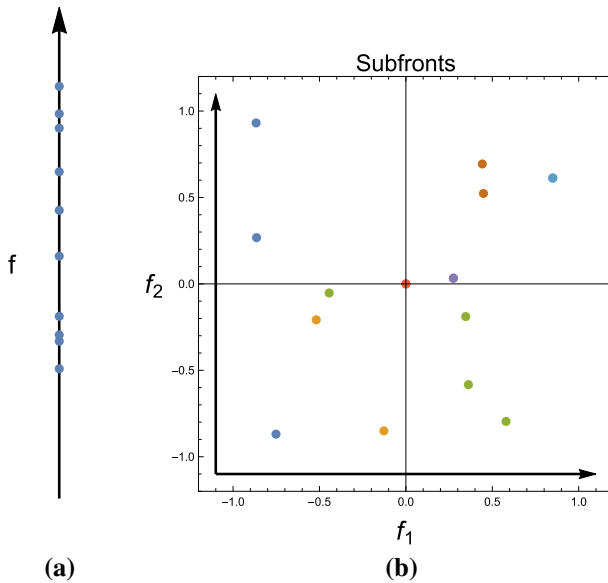
#### 4.4 Discussion and comparison with the selection rules of other algorithms

Alternative extensions of the concept of potential Pareto optimality are considered in [1,4,26]. In [4], a hyperinterval  $H$  is deemed potentially Pareto optimal if it is potentially optimal in the scalar sense (2) for *at least* one of the objective functions.

In [26], an exact extension of the geometrical interpretation of potential optimality is considered as a convex hull of the set of augmented objective vectors. In [1], the nondominated set of the set of the augmented objective vectors is used as the set of potential Pareto optima. The same definition is considered also in [4] as a possible alternative.

By considering Example 6, we notice that for the algorithm of [4], the potential Pareto optima are only  $H_1, H_2, H_4$  and  $H_6$ . Here, although  $H_5$  is a nondominated point in the  $\sigma$ -level, it is not deemed potentially Pareto optimal because it is not optimal for at least one of the objectives. In the generic case, the subset of the Pareto set composed by tradeoff points is a  $(k - 1)$ -dimensional submanifold of the decision space, while points that are optimal for at least one objective considered individually, form a discrete and finite set. Therefore, the definition in [4] excludes a large part of potentially relevant hyperintervals.

Also according to [26],  $H_1, H_2, H_4$  and  $H_6$  are deemed potentially Pareto optimal. In this case,  $H_5$  is excluded because it belongs to a nonconvex part of the nondominated set of the  $\sigma$ -level. Nevertheless, the notion in [26] is less restrictive because it does not exclude at least



**Fig. 8** In DIRECT, a ranking among directions is provided by objective function values (a). In the multiobjective case, a total ordering is not available a priori (b)

the convex parts of the nondominated sets, while [4] can pick as potentially Pareto optimal only the extrema of the subfronts.

On the other hand, if we consider the nondominated subset of the augmented objective vectors as in [1], we can say that all hyperintervals  $H_1, \dots, H_6$  are potentially Pareto optimal, which is in agreement with Definition 5. However, Definition 5 excludes  $H_2$  in Example 4, while for [1] all  $H_1, \dots, H_4$  are potentially Pareto optimal. We can conclude that, when comparing the new definition 5 with the definitions used in the previous algorithms, there is a possibility that [4] and [26] result in slower convergence, because they exclude promising hyperintervals in early iterations. The algorithm of [1] can be slower in convergence, because it could waste some computational resources for hyperintervals that are less likely to contain Pareto optimal points than other hyperintervals. The effectiveness of these different definitions of potential Pareto optimality are compared in Sect. 6.

### 4.5 Division rule

The challenges of extending the division rule to multiple objectives are related to the fact that in the original division rule for the single objective case, a total ordering criterion among the directions in the decision space is necessary. This total ordering is obtained by testing values computed at both sides of the central point, as illustrated in Figs. 3 and 8a.

Structurally, this is a feature impossible to be reproduced in multiobjective optimization, because test values computed along different directions are vector-valued. Thus, they cannot be compared univocally. This is exhibited in Fig. 8b, where it is clear that comparisons are not always possible. The central point, for instance, can be compared with points in the first and third quadrant of the plane, but not with points in the second and fourth quadrant.

In the literature, some criteria have been proposed to extend the division rule using, for instance, the Euclidean distance of the test point values from the objective functions value of the center point of the ancestor hyperinterval [1,45]. Other choices are possible, e.g.,

using an achievement scalarizing function [43] with the central point value as a reference point. Alternatively, the hypervolume [13] corresponding to each new test point can be used, producing a scalar value useful for ranking. These algorithms, however, make use of a more or less explicit reference to the individual scales at which the objective functions are represented. The Euclidean distance and the hypervolume depend on the function scaling and the achievement scalarizing function explicitly requires that some weights are provided to combine the functions. Instead, a lexicographic ordering could be used, but also in this case an explicit, and arbitrary, ranking among objectives is introduced, which shares the same drawback observed above.

Of course, in a general case, an a priori ranking among the objective functions, function ranges or respective global Lipschitz constants is not provided and does not exist in a universal sense. This means that each of the above-mentioned rankings is somehow arbitrary, i.e., they introduce a bias that in some cases may be negligible, but in principle it can impair the functioning and the convergence features of the algorithm.

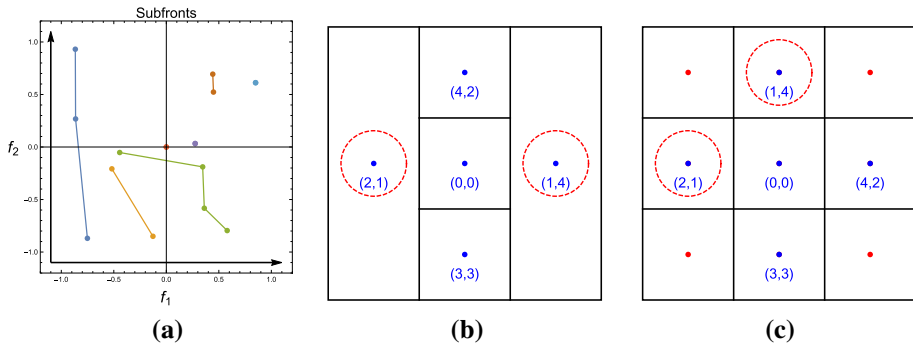
A possible neutral approach could be an iterative estimation of the Lipschitz constants (or a reciprocal scaling) on the basis of the values computed so far, and an attractive choice could be that the directions showing the largest estimated Lipschitz constant could be divided earlier than the remaining directions to follow possible emerging anisotropy in the feasible set. Nevertheless, it is preferable, and conceptually cleaner, that the sequence of newly computed points is completely independent of any ranking or scaling of objective functions. This is in agreement with the issues discussed in the introduction regarding the existing algorithms extending or making use of DIRECT and other exact global optimization algorithms.

A possible approach avoiding a total ordering between objectives is to use nondominated sorting [39], which is a partial ordering. In nondominated sorting, we extract from a finite discrete set of vector test values the nondominated ones, to form a so-called first subfront. After removing the first subfront from the starting set we select again the nondominated vectors from the remaining vectors to form a second subfront, and so on until the original set is completely emptied. Test values belonging to the same subfront are given the same rank. The computation of each subfront relies only on Pareto dominance and, therefore, it does not depend on the function scalings or on any other arbitrary ranking between the objective functions. Nondominated sorting is illustrated in Fig. 9a, where the first subfront is the blue one, the second is the orange, etc.

By adopting such a partial ordering among the directions in the feasible space to be subdivided, we modify the original division rule by splitting *at the same time* directions belonging to the same subfront, i.e., having the same rank. This requires that some extra test point is computed at the center of the extra hyperintervals generated in this way, as can be seen in Fig. 9c.

In addition to avoiding the use of any ranking between directions or scaling between objectives, this partial ordering avoids introducing complicated or arbitrary adaptive rules. On the other hand, it involves some computational burden, that has to be taken into account. The computational cost of extra points may be acceptable because of the following reasons. First of all, in typical applications, the dimension of the feasible space, which corresponds to the number of directions to be divided, cannot be very large, because an exhaustive search requiring global convergence, i.e., an everywhere dense sampling in an infinite time, cannot be efficiently performed in high-dimensional decision spaces. Rarely we are considering decision spaces with dimensions larger than five.

Secondly, not all directions in the decision space are candidates for division but only the subset corresponding to the longest edges of the hyperinterval, as in the original DIRECT



**Fig. 9** Division rule. **a** Test values are grouped in subfronts according to nondominance. **b** Sometimes the partial ordering is sufficient for establishing a total ranking among input directions and the classical division rule can be applied. Numbers represent objective function values. Test points corresponding to the same direction belong to the same subfront and obtain the same rank. **c** When a total ranking is not available, a complete division rule is adopted. Test points corresponding to different directions belong to the same subfront

division rule. Finally, nondominated sorting produces a total ordering among directions in a considerable number of cases, as can be seen in Fig. 9b.

### 5 multiDIRECT algorithm

Now we can collect and summarize the ideas discussed in the previous sections about the possible issues of extending the DIRECT algorithm for multiobjective optimization problems. Based on the discussion we can design an algorithm to be called *multiDIRECT*, which is presented in Algorithm 1.

---

#### Algorithm 1 *multiDIRECT*

---

- Input:** Let  $f$  be a vector-valued function as in (MOP). Let  $nIter$  be the number of iterations to be performed. Let  $\epsilon > 0$  be a small parameter.
- 1: Initialize the set of hyperintervals covering the domain  $D = [0, 1]^d$  by  $S := \{H\}$  with  $H := (c, F, \sigma) = ((0.5, \dots, 0.5), F(0.5, \dots, 0.5), (1, \dots, 1))$ .
  - 2: **for**  $iter = 1, iter \leq nIter, iter ++$ , **do**
  - 3:   Group the hyperintervals according to their radius in  $\sigma$ -levels  $\sigma L := \{H = (c, F, s) \mid s = \sigma\}$ .
  - 4:   For each  $\sigma$ -level, extract the nondominated set  $ND(\sigma L)$ .
  - 5:   Initialize the set of candidate hyperintervals  $Cand$  with the nondominated set of the  $\sigma$ -level with  $\sigma = \sigma_{max}$ , i.e.,  $Cand := ND(\sigma_{max} L)$ .
  - 6:   In the nondominated set of  $\sigma$ -levels with a nonmaximal  $\sigma$ , add the potentially Pareto optimal hyperintervals according to Theorem 1 and condition (11) to  $Cand$ .
  - 7:   For every hyperinterval in  $Cand$ , select the input directions with a maximal radius, divide in thirds along each of these directions and define new test points at the center of each third.
  - 8:   Evaluate objective functions on the test points.
  - 9:   Determine the partial ordering among directions according to the nondominated sorting.
  - 10:   Apply the division rule described in Section 4.5 evaluating new test points where necessary.
  - 11:   Update the list  $S$  by removing the candidate hyperintervals and adding the newly computed subintervals.
  - 12: **end for**
  - 13: Extract the nondominated set of the hyperintervals  $S$  as the approximation of the Pareto optimal set.
- return**  $ND(S)$ .
-

We notice that the algorithm is globally convergent.

**Theorem 2** *The sequence of points generated by multiDIRECT for any function  $f : D \rightarrow \mathbb{R}^k$  becomes infinitely dense everywhere in  $D$  in an infinite time.*

**Proof** By recalling Remark 1, we note that for every iteration, the nondominated vectors in the  $\sigma$ -level with a maximal  $\sigma$  are regarded potentially Pareto optimal. Therefore, in a finite number of iterations, the maximal  $\sigma$ -level will be emptied, i.e.,  $\sigma_{max} \rightarrow 0$  in a infinite time. Then, also the radius of a ball contained in the feasible set and not containing any of the sampled points, i.e., the centers of the hyperintervals, is tending to zero, because this radius is smaller than the diameter of the largest hyperinterval multiplied by  $\sqrt{d}$ . Therefore, the sampling consisting in the centers of the hyperintervals is deemed to become everywhere dense.  $\square$

## 6 Some numerical comparisons

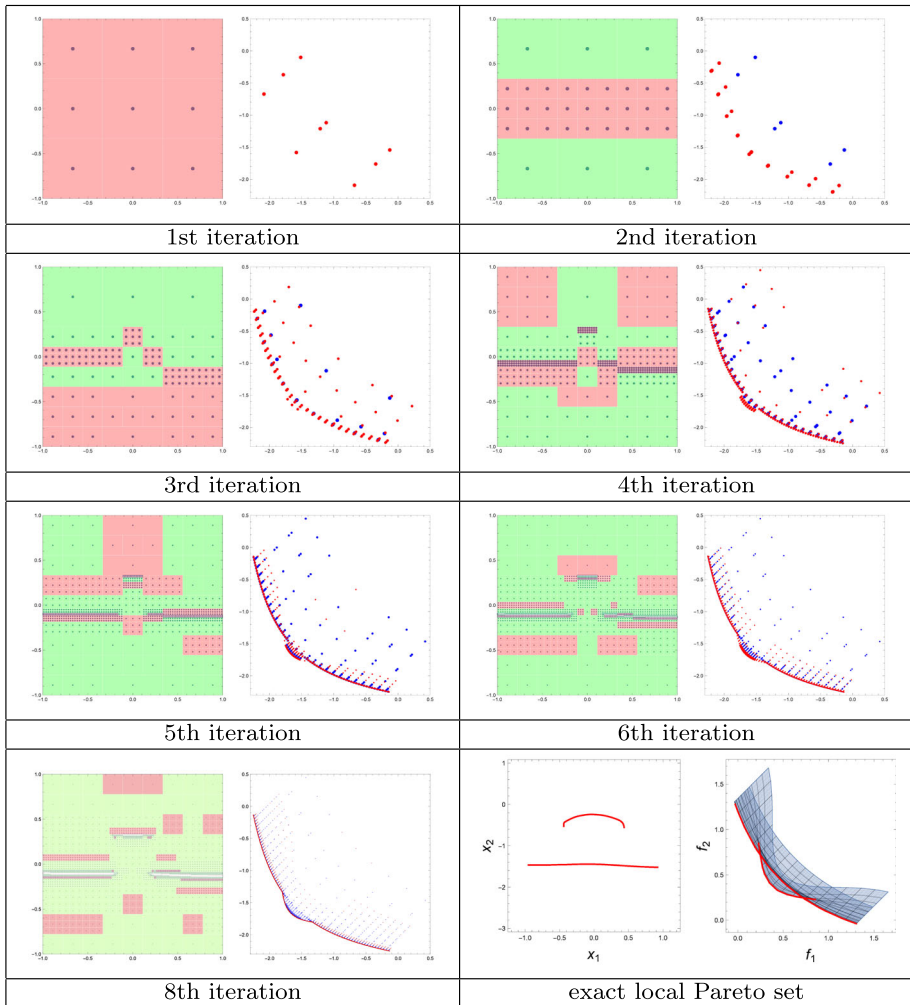
In this section, we demonstrate the numerical behaviour of the proposed *multiDIRECT* algorithm and compare its performance to other deterministic algorithms. In the sequence of visualizations in Fig. 10, we present the application of the *multiDIRECT* algorithm on a biobjective problem with two variables named  $L\&H_{2 \times 2}$ . The problem is defined in [25] and its Pareto set and front are illustrated in Fig. 1a. We have used the output of the SICON algorithm as a surrogate for the exact representation of the Pareto set, because SICON produces a set-wise approximation with a quadratic precision. Furthermore, it was possible to produce a grid of points of desired density in the Pareto set. On the other hand, we did not compare SICON with the remaining methods because it is not of the same class as *multiDIRECT*. Indeed SICON makes use of first and second derivatives and it is not designed for being globally convergent.

We have decided to compare *multiDIRECT* with two other algorithms because they were the only globally convergent and completely fully deterministic ones available. In addition, we have also tested a version of *multiDIRECT* with a division rule borrowed from one of the other algorithms. More precisely, we consider:

1. MO-DIRECT-ND with total ordering in the division rule as in the original paper [1] (green line),
2. *multiPS* with the Lipschitz constants computed from the analytical definition of the functions (orange line),
3. *multiDIRECT* with the same total ordering for the division rule in MO-DIRECT-ND (yellow line),
4. *multiDIRECT* with partial ordering as in Algorithm 1 and  $\epsilon = 0$  (blue line).

To compare the performance of the four algorithms, we have measured at each iteration the Hausdorff distance between the Pareto set approximation produced by the algorithm under examination and the surrogate exact Pareto set produced by SICON. In Fig. 11, we plot the Hausdorff distances versus the number of function evaluations used to produce the current approximation. In Fig. 11, we also report the mesh size used in SICON, the red horizontal line, to represent the best approximation level reachable using this surrogate for the Pareto set.

As can be observed, the algorithms using the partial ordering require initially a larger number of function evaluations but demonstrate better performance in the last iterations. By observing the two versions of *multiDIRECT*, i.e., the yellow and blue lines in the online paper, there are no differences in performance for the first iterations. Actually the blue line

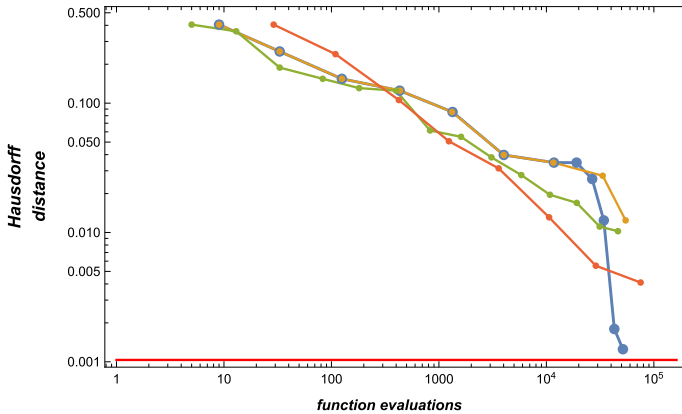


**Fig. 10** Application of the *multiDIRECT* algorithm to  $L&H_{2,x^2}$  [19]. For every subpicture, the left panel depicts the decision space while the right panel represents the objective space. The last subpicture is a representation of the local Pareto set obtained by *SICON* [23]

is hidden behind the yellow line. When the selection rule based on the potential Pareto optimality starts to filter some of the hyperintervals accepted by the ND rule, the number of function evaluations for every iteration gets smaller while the performances do not deteriorate sensibly. This results in with an overall better efficiency.

Although the partial ordering division rule used by *multiDIRECT* appears more costly in the first iterations, it leads to smaller distances from the global Pareto set in later iterations.

The *multiPS* algorithm appears overall more efficient than the other algorithms, at least with  $10^3$  to  $3 \times 10^4$  function evaluations, but it must be taken into account that global information of the analytical Lipschitz constant has been used and it is usually not available. Using larger constants implies wasting computational resources in unpromising regions, while smaller Lipschitz constants may miss to detect some non-negligible portion of the Pareto set.



**Fig. 11** Global convergence rates for the extensions of the DIRECT algorithm. Green line: MO-DIRECT-ND. Orange line: *multiPS* algorithm with analytical estimate of the global Lipschitz constant. Yellow line: *multiDIRECT* with total ordering as in MO-DIRECT-ND. Blue line: *multiDIRECT* as in Algorithm 1 and  $\epsilon = 0$ . The red horizontal line represents the mesh size with which the exact Pareto set has been approximated, therefore the Hausdorff distance cannot go under this value

We can eventually conclude that the overall best performance has been obtained by the *multiDIRECT* method, although it required a considerable number of function evaluations. In particular, it seems that the selection rule based on the new notion of potential Pareto optimality is a crucial aspect in the efficiency of the algorithm proposed.

## 7 Conclusions

We have provided a conceptual and theoretical investigation on the notion of global convergence in multiobjective optimization, with a particular emphasis on the notion of potential Pareto optimality. It is a crucial notion involved in extending the global single objective optimization algorithm DIRECT to multiple objectives. We have discussed advantages and disadvantages of the extensions available in the literature and proposed a new extension called *multiDIRECT*.

Compared to Lipschitz optimization algorithms, *multiDIRECT* guarantees global convergence even if Lipschitz constants are not known or do not exist. The introduced notion of potential Pareto optimality reduces the number of hyperintervals selected for subdivision and speeds up the convergence when compared to algorithms using more elementary selection rules. A preliminary numerical investigation confirms the effectiveness of this algorithm. The combination of potential Pareto optimality with different single objective optimization algorithms and further numerical investigations will be subjects of future research.

**Acknowledgements** This research is related to the thematic research area DEMO (Decision Analytics utilizing Causal Models and Multiobjective Optimization, [juu.fi/demo](http://juu.fi/demo)) of the University of Jyväskylä.

**Funding** Open access funding provided by Università degli Studi di Padova within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence,



and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Al-Dujaili, A., Suresh, S.: Dividing rectangles attack multi-objective optimization. In: 2016 IEEE Congress on Evolutionary Computation (CEC), pp. 3606–3613 (2016)
2. Allgower, E.L., Schmidt, P.H.: An algorithm for piecewise-linear approximation of an implicitly defined manifold. *SIAM J. Numer. Anal.* **22**, 322–346 (1985)
3. Audet, C., Savard, G., Zghal, W.: A mesh adaptive direct search algorithm for multiobjective optimization. *Eur. J. Oper. Res.* **204**(3), 545–556 (2010)
4. Campana, E.F., Diez, M., Liuzzi, G., Lucidi, S., Pellegrini, R., Piccialli, V., Rinaldi, F., Serani, A.: A multi-objective direct algorithm for ship hull optimization. *Comput. Optim. Appl.* **71**(1), 53–72 (2018)
5. Cao, P., Qi, S., Tang, J.: Structural damage identification using piezoelectric impedance measurement with sparse inverse analysis. *Smart Mater. Struct.* **27**(3), 035020 (2018)
6. Cao, P., Shuai, Q., Tang, J.: A multi-objective DIRECT algorithm toward structural damage identification with limited dynamic response information. *J. Nondestruct. Eval. Diagn. Prognos. Eng. Syst.* **1**(2), 021004–021004–12 (2017)
7. Custódio, A.L., Madeira, J.F.A., Vaz, A., Vicente, L.: Direct multisearch for multiobjective optimization. *SIAM J. Optim.* **21**, 1109–1140 (2011)
8. Custódio, A.L., Madeira, J.F.A.: GLODS: global and local optimization using direct search. *J. Global Optim.* **62**, 1–28 (2015)
9. Custódio, A.L., Madeira, J.F.A.: MultiGLODS: global and local multiobjective optimization using direct search. *J. Global Optim.* **72**(2), 323–345 (2018)
10. Das, I., Dennis, J.E.: Normal-Boundary Intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM J. Optim.* **8**, 631–657 (1998)
11. Deb, K.: *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley, Chichester (2001)
12. Dellnitz, M., Schütze, O., Hestermeyer, T.: Covering Pareto sets by multilevel subdivision techniques. *J. Optim. Theory Appl.* **124**, 113–136 (2005)
13. Emmerich, M., Beume, N.: An EMO algorithm using the hypervolume measure as selection criterion. In: Coello Coello, C.A., Hernández Aguirre, A., Zitzler, E. (eds.) *Evolutionary Multi-Criterion Optimization*, pp. 62–75. Springer, Berlin, Heidelberg (2005)
14. Evtushenko, Y., Potapov, M.: A nondifferentiable approach to multicriteria optimization. In: Demyanov, V.F., Pallaschke, D. (eds.) *Nondifferentiable Optimization: Motivations and Applications*, pp. 97–102. Springer, Berlin, Heidelberg (1985)
15. Evtushenko, Y.G., Posypkin, M.A.: A deterministic algorithm for global multi-objective optimization. *Optim Methods Softw.* **29**(5), 1005–1019 (2014)
16. Gergel, V., Kozinov, E.: Efficient multicriterial optimization based on intensive reuse of search information. *J. Global Optim.* **71**(1), 1–18 (2018)
17. Guillemin, V., Pollack, A.: *Differential Topology*. Prentice-Hall Inc., Englewood Cliffs, N.J. (1974)
18. Hartikainen, M., Miettinen, K., Wiecek, M.M.: PAINT: Pareto front interpolation for nonlinear multiobjective optimization. *Comput. Optim. Appl.* **52**, 845–867 (2012)
19. Hartikainen, M.E., Lovison, A.: PAINT-SiCon: constructing consistent parametric representations of Pareto sets in nonconvex multiobjective optimization. *J. Global Optim.* **62**(2), 243–261 (2015)
20. Jones, D.R., Perttunen, C.D., Stuckman, B.E.: Lipschitzian optimization without the Lipschitz constant. *J. Optim. Theory Appl.* **79**, 157–181 (1993)
21. Knowles, J.: ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Trans. Evol. Comput.* **10**, 50–66 (2006)
22. Lera, D., Sergeyev, Y.D.: GOSH: derivative-free global optimization using multi-dimensional space-filling curves. *J. Global Optim.* **71**(1), 193–211 (2018)
23. Lovison, A.: Singular Continuation: generating piecewise linear approximations to Pareto sets via global analysis. *SIAM J. Optim.* **21**, 463–490 (2011)
24. Lovison, A.: Global search perspectives for multiobjective optimization. *J. Global Optim.* **57**, 385–398 (2013)

25. Lovison, A., Hartikainen, M.E.: On generalizing Lipschitz global methods for multiobjective optimization. In: Gaspar-Cunha, A., Henggeler Antunes, C., Coello, C.C. (eds.) *Evolutionary Multi-Criterion Optimization*, pp. 264–278. Springer, Cham (2015)
26. Lovison, A., Miettinen, K.: Exact extension of the DIRECT algorithm to multiple objectives. *AIP Conf. Proc.* **2070**(1), 020053 (2019)
27. Lovison, A., Pecci, F.: Hierarchical stratification of Pareto sets (2014). [arXiv:1407.1755](https://arxiv.org/abs/1407.1755) Accessed 15 June 2020
28. de Melo, W.: On the structure of the Pareto set of generic mappings. *Boletim da Sociedade Brasileira de Matemática* **7**, 121–126 (1976)
29. Miettinen, K.: *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Boston (1999)
30. Neumaier, A.: Complete search in continuous global optimization and constraint satisfaction. *Acta Numerica* **13**, 271–369 (2004)
31. Pardalos, P.M., Žilinskas, A., Žilinskas, J.: *Non-Convex Multi-Objective Optimization*. Springer, Cham (2017)
32. Paulavičius, R., Sergeyev, Y.D., Kvasov, D.E., Žilinskas, J.: Globally-biased BIRECT algorithm with local accelerators for expensive global optimization. *Expert Syst. Appl.* **144**, 113052–27 (2020)
33. Piyavskii, S.: An algorithm for finding the absolute extremum of a function. *USSR Comput. Math. Math. Phys.* **12**, 57–67 (1972)
34. Schütze, O., Dell’Aere, A., Dellnitz, M.: On continuation methods for the numerical treatment of multi-objective optimization problems. In: *Practical Approaches to Multi-Objective Optimization*. IBFI, Schloss Dagstuhl, Germany (2005)
35. Sergeyev, Y.D., Kvasov, D.E.: Global search based on efficient diagonal partitions and a set of Lipschitz constants. *SIAM J. Optim.* **16**(3), 910–937 (2006)
36. Sergeyev, Y.D.: On convergence of “divide the best” global optimization algorithms. *Optimization* **44**(3), 303–325 (2007)
37. Shubert, B.: A sequential method seeking the global maximum of a function. *SIAM J. Numer. Anal.* **9**, 379–388 (1972)
38. Smale, S.: Global analysis and economics I: Pareto optimum and a generalization of Morse theory. In: Peixoto, M.M. (ed.) *Dynamical Systems*, pp. 531–544. Academic Press, (1973)
39. Srinivas, N., Deb, K.: Multiobjective optimization using nondominated sorting in genetic algorithms. *Evol. Comput.* **2**(3), 221–248 (1994)
40. Stephens, C., Baritomba, W.: Global optimization requires global information. *J. Optim. Theory Appl.* **96**, 575–588 (1998)
41. Strongin, R.G., Sergeyev, Y.D.: *Global Optimization with Non-Convex Constraints*. Springer, US (2000)
42. Wang, L., Ishida, H., Hiroyasu, T., Miki, M.: Examination of multi-objective optimization method for global search using DIRECT and GA. In: *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 2446–2451. IEEE (2008)
43. Wierzbicki, A.P.: A mathematical basis for satisficing decision making. *Math. Model.* **3**, 391–405 (1982)
44. Wierzbicki, A.P.: On the completeness and constructiveness of parametric characterizations to vector optimization problems. *OR Spectrum* **8**, 73–87 (1986)
45. Wong, C.S.Y., Al-Dujaili, A., Sundaram, S.: Hypervolume-based DIRECT for multi-objective optimization. In: *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion, GECCO ’16 Companion*, pp. 1201–1208. ACM, New York (2016)
46. Žilinskas, A.: A one-step worst-case optimal algorithm for bi-objective univariate optimization. *Optim. Lett.* **8**(7), 1945–1960 (2013)
47. Žilinskas, A., Gimbutienė, G.: On one-step worst-case optimal trisection in univariate bi-objective Lipschitz optimization. *Commun. Nonlinear Sci. Numer. Simul.* **35**, 123–136 (2016)