Exact Extension of the DIRECT Algorithm to Multiple Objectives

Alberto Lovison[1,a)] and Kaisa Miettinen[2,b)]

[1] *M3E – Mathematical Methods and Models for Engineering. University of Padova. Italy*
[2] *University of Jyvaskyla, Faculty of Information Technology, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland*

[a)] Corresponding author: alberto.lovison@gmail.com
[b)] kaisa.miettinen@jyu.fi

Abstract. The DIRECT algorithm has been recognized as an efficient global optimization method which has few requirements of regularity and has proven to be globally convergent in general cases. DIRECT has been an inspiration or has been used as a component for many multiobjective optimization algorithms. We propose an exact and as genuine as possible extension of the DIRECT method for multiple objectives, providing a proof of global convergence (i.e., a guarantee that in an infinite time the algorithm becomes everywhere dense). We test the efficiency of the algorithm on a nonlinear and nonconvex vector function.

Exact global search methods are a class of algorithms well developed for single objective optimization. These methods may usually show appreciable speed of convergence for moderate dimensions of the variable space but their specific trait is approximating the global optimum of the function considered with an arbitrary precision in a finite time [1]. At the moment, corresponding exact and global methods for multiobjective optimization have not been developed and employed to the same extent as their single objective counterparts. However, some of these methods have inspired or have been used as a component in a number of multiobjective optimization algorithms [2, 3, 4]. Nevertheless, most of them[1] have the following characteristics:

1. the exact global method is used only after the multiobjective optimization problem has been scalarized, or
2. the exact global method is hybridized at some level with non-deterministic principles.

Therefore, the methods can suffer from one of the following consequences:

1. the resulting Pareto front (in the objective space) is accurately approximated but incompletely represented,
2. the accuracy is scarse and although a systematic covering of the Pareto set is often accomplished, it cannot be guaranteed.

There is a fundamental aspect that distinguishes the usual global optimization of a scalar function and the global optimization of several objective functions (i.e., vector optimization). In scalar optimization,

$$f : S \to \mathbb{R}, \qquad S = [0,1]^d \subseteq \mathbb{R}^d, \qquad \text{minimize}_{x\in S} f(x), \tag{1}$$

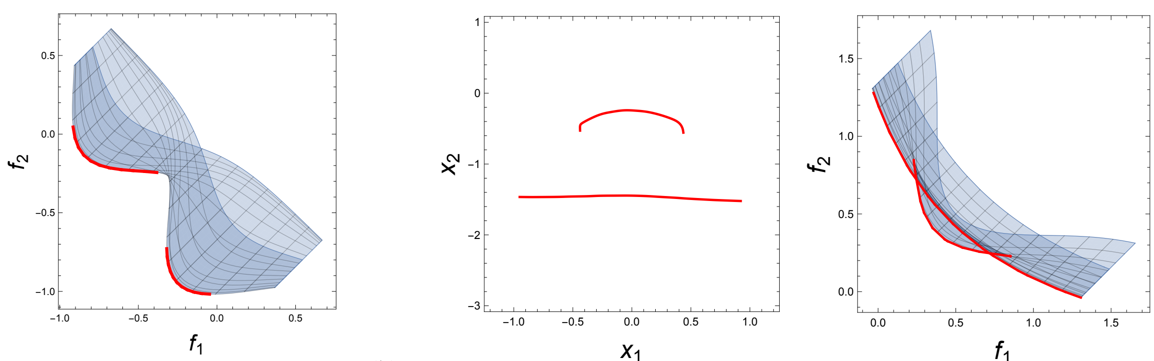
where S is called a *feasible set*, the generic situation is that the set of optimal (objective function) values is a single value that often corresponds to a unique optimal solution. In pathological cases the optimal value could even not exist (unbounded function), or when the optimal value exist it could correspond to a number, even infinite, of optimal solutions. However, this situation can be proved non generic, i.e., for a dense set in a suitable set of functions, there will exist only one optimal value with a corresponding optimal point in the feasible set.[2] Therefore, the global convergence of single objective optimization algorithms is established by requiring that the globally optimal value is being approximated in an infinite time. In the generic situation this determines a unique optimum also in the feasible set S.

[1] For a deeper discussion see [5] As notable exceptions see, e.g., [6, 7].
[2] It is a consequence of Sard's Theorem and Morse's Lemma [8]).

The case of multiple objectives, i.e.,

$$f : S^d \to \mathbb{R}^k, \qquad S = [0,1]^d \subseteq \mathbb{R}^d, \qquad \text{minimize}_{x \in S}\, f(x), \tag{2}$$

is completely different: in general there does not exist a global optimum for all the objective functions at the same time, but it is necessary to consider the notion of *dominance* [9, 10]. We say that a point $x \in S$ *dominates* another point $y \in S$ if for all $i = 1, \ldots, k$, $f_i(x) \leqslant f_i(y)$ and there exists j such that $f_j(x) < f_j(y)$. We say that a point x is a *Pareto optimum* if it is not dominated by any other point in the feasible set. The set of the Pareto optimal objective function values, i.e., $\{f(x) : x \text{ is Pareto optimum}\}$ is called the *Pareto front*. In the general nonconvex case the Pareto front is an infinite set and may be composed by a number of separate branches (Figure 1(a)). Furthermore, it can be proved that under reasonable regularity conditions the set of Pareto optima is composed of separate portions of $(k-1)$–dimensional manifolds. These components can superimpose and intersect each other when they are mapped on the Pareto front as in the example represented in Figure 1(b).[3] Only in very special situations, i.e., pathological cases, the set of Pareto optima is a unique vector, or a finite number of vectors [10, 12, 11, 13].



(a) (b)

FIGURE 1. Possible issues occurring for non–convex multiobjective problems. (a) Multiple Pareto fronts. (b) Superimposing fronts. (Left: Pareto optimal set in the feasible set. Right: Pareto front.)

Sometimes the global convergence of multiobjective optimization methods is discussed in the same terms of the single objective cases, i.e., by showing that an algorithm converges towards a (single) global Pareto optimum. We call this approach a *point-wise convergence*. Clearly, this approach completely misses the global features of the set of Pareto optima discussed above. Indeed, any global optimization algorithm can be extended to multiple objectives by simply ignoring all the objectives but one. Such a method confuses the optimum of a single objective function with the set of Pareto optima of the vector function, but what is worse is that this is satisfactory in terms of point-wise convergence. Therefore, it seems more appropriate to approximate the set of Pareto optima by means of a *set-wise approach*, i.e., the set of Pareto optima should not be approached in a-point-at-a-time fashion, as in multi-start methods, but it should be approximated as a whole by means of a sequence of approximating sets of points. Moreover, because it is not unlikely that the same value in the Pareto front can be the image of two or more different points in the feasible set (see the above example in Figure 1(b)), it is important to study the approximating sequences of sets in the feasible set. A suitable concept of distance between sets is needed, and a natural choice is the Hausdorff distance.

We want to develop a multiobjective optimization algorithm that produces at every iteration a candidate set approximating the whole set of Pareto optima. The algorithm can define one or more points at every iteration. In this situation, a candidate set is built at every iteration by picking nondominated points among new and old points. A consistent definition for global convergence can be given as follows:

Definition 1 *Let $P \subseteq [0,1]^d$ be the set of Pareto optima of problem (2). Consider an optimization algorithm producing the following sequence of candidate Pareto optimal sets $S_1, S_2, \ldots$ We will say that the algorithm* globally converges *for the function f if*

$$\lim_{i \to \infty} d_{\mathcal{H}}(S_i, P) = 0, \qquad \textit{where } d_{\mathcal{H}} \textit{ stands for the Hausdorff distance.} \tag{3}$$

[3] The explicit definition and the details of this example, called $L\&H_{2x2}$, are given in [11]. The Pareto set is obtained by the sicon method [12].

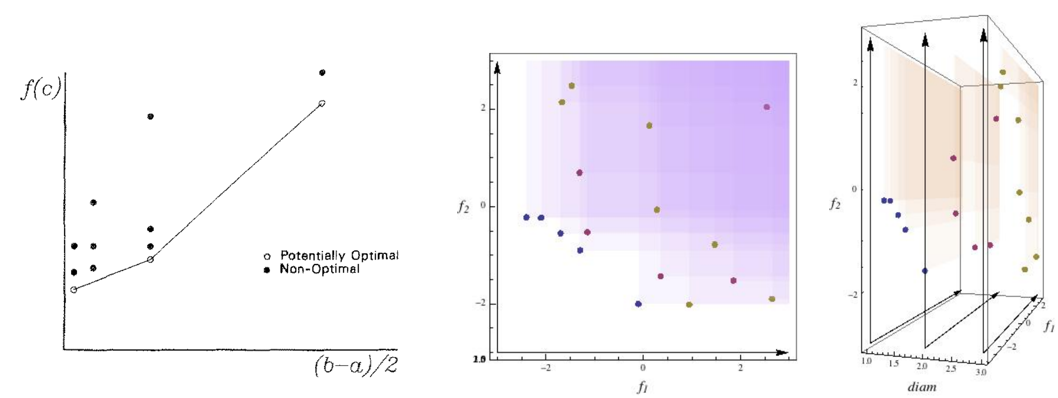
In this framework, we have developed in [4] *multi*SHUBERT, a multiobjective extension of the Piyavskii-Shubert algorithm [14, 15], and we have shown the global convergence of the algorithm in the sense explained above. A quite natural generalization of the Piyavskii-Shubert algorithm is DIRECT [16], which outperforms its ancestor in several senses. First, DIRECT does not require the definition of a global Lipschitz constant, which usually is not known, and second, DIRECT exhibits a faster convergence. In this contribution we present an extension of DIRECT for multiple objectives fulfilling all the requirements discussed above: the *multi*DIRECT is exact, i.e., it is not hybridized with heuristic strategies, it shows a global convergence in the set-wise sense and, finally, exhibits an appreciable speed of convergence when compared with *multi*SHUBERT.

Next, we describe the idea of DIRECT and after that we can describe how *multi*DIRECT extends these concepts. Thus, we first consider an optimization problem with a Lipschtiz continuous single objective function f.The DIRECT algorithm divides the feasible set defined as hyperintervals in smaller hyperintervals on the basis of a Lipschitz criterion without fixing a Lipschitz constant. Assuming that the point in the center of a hyperinterval has been evaluated, it is possible to estimate a lower bound for f by fixing a Lipschitz constant $\alpha > 0$ by setting $\ell_j := f(x_j) - \alpha d_j$, where x_j is the point at the center of the j-th hyperinterval and d_j is the radius of the hyperinterval. By ordering the sequence of ℓ_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 Lipschitz constant are chosen. Indeed small d_j with low values for $f(x_j)$ will be preferred with small values of α while large radii d_j will be preferred by larger values of α.

If we call an interval whose lower bound is optimal for a certain constant α *potentially optimal*, we can show that all potentially optimal intervals correspond to the points in the lower face of the convex hull of the following set of points as the Lipschitz constant α varies from 0 to ∞ (see Figure 2(a)):

$$\left\{\left(d_j, f(x_j)\right) \,\middle|\, j = 1, 2, \ldots\right\} \subseteq \mathbb{R}^2. \tag{4}$$

At each iteration, DIRECT picks these hyperintervals on the convex hull and promotes them for further subdivision in smaller hyperintervals. For each new candidate hyperinterval the point at the center is evaluated, and a new diagram as in Figure 2(a) is produced by updating the list of radii and function values by removing the splitted intervals and adding the newly produced ones.



(a) (b) (c)

FIGURE 2. (a) DIRECT selection rule. (b) *multi*DIRECT objectives values. (c) *multi*DIRECT selection rules.

The extension of the algorithm to multiple objectives is proposed in an as straightforward manner as possible by focussing on the geometrical interpretation of the selection rule of DIRECT. We first assume that the (hyper)-rectangles composing the feasible set for the vector-valued function f have a point in their center where the objective functions have been evaluated. By augmenting the dimension of the objective space by adding the radius of the hyperintervals, we consider the following set

$$\left\{\left(d_j, f_1(x_j), \ldots, f_k(x_j)\right) \,\middle|\, j = 1, 2, \ldots\right\} \subseteq \mathbb{R}^{k+1}, \tag{5}$$

illustrated in Figure 2(b), which defines the multiobjective analogue of the diagram in Figure 2(a).

Definition 2 *A hyperinterval I_j is* potentially optimal *if there exists a sequence $\{\alpha_1, \ldots, \alpha_k\}$ of Lipschitz constants such that the virtual lower bound $(f_1(x_j) - \alpha_1 d_j, \ldots, f_k(x_j) - \alpha_k d_j)$ is not dominated by the current set of values $\left\{(f_1(x_j), \ldots, f_k(x_j)) \middle| \, j = 1, 2, \ldots\right\}$.*

Proposition 1 *Every point $(d_j, f_1(x_j), \ldots, f_k(x_j))$ belonging to faces whose outward normal vector is pointing downwards of the convex hull of the set (5) corresponds to a potentially optimal interval I_j.*

At every iteration the *multi*DIRECT algorithm will select for subdivision of all those potentially optimal hyperintervals. In the sequence of Figure 3 we present the iterations of the *multi*DIRECT algorithm on two variables biobjective problem (see [4] and Figure 1(b)). Analogously, it is not difficult to prove the following global convergence result:

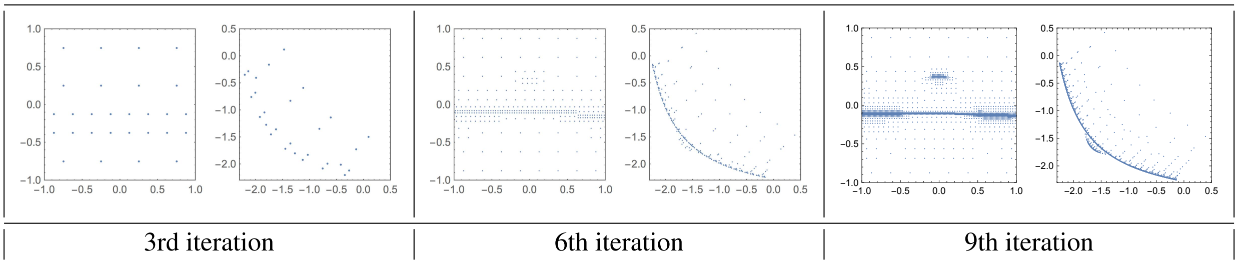


FIGURE 3. Application of the *multi*DIRECT algorithm to $L\&H_{2x2}$ [11]. Left panel: feasible set. Right panel: objectives space.

Theorem 1 *The sequence of points generated by multi*DIRECT *for any function $f : [0,1]^d \to \mathbb{R}^k$ becomes infinitely dense everywhere in the domain $[0,1]^d$ in an infinite time.*

By taking the sequence $S_j := \left\{x_\iota \middle| \, f(x_\iota) \text{ is nondominated by } f(x_1), \ldots, f(x_j)\right\}$ the previous proposition guarantees the global convergence in the sense of Definition 1.

REFERENCES

[1] C. Stephens and W. Baritompa, Journal of Optimization Theory and Applications **96**, 575–588 (1998).
[2] J. Knowles, IEEE Transactions on Evolutionary Computation **10**, 50 – 66 (2006).
[3] A. Lovison, Journal of Global Optimization **57**, 385–398 (2013).
[4] A. Lovison and M. E. Hartikainen, "On generalizing Lipschitz global methods for multiobjective optimization," in *Evolutionary Multi-Criterion Optimization*, edited by A. Gaspar-Cunha, C. Henggeler Antunes, and C. C. Coello (Springer International Publishing, 2015), pp. 264–278.
[5] H. P. Benson and S. Sayin, Naval Research Logistics **44**, 47–67 (1997).
[6] A. Al-Dujaili and S. Suresh, "Dividing rectangles attack multi-objective optimization," in *2016 IEEE Congress on Evolutionary Computation (CEC)* (2016), pp. 3606–3613.
[7] C. S. Y. Wong, A. Al-Dujaili, and S. Sundaram, "Hypervolume-based DIRECT for Multi-Objective Optimisation," in *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*, GECCO '16 Companion (ACM, New York, NY, USA, 2016), pp. 1201–1208.
[8] V. Guillemin and A. Pollack, *Differential topology* (Prentice-Hall Inc., Englewood Cliffs, N.J., 1974), pp. xvi+222.
[9] K. Miettinen, *Nonlinear Multiobjective Optimization* (Kluwer Academic Publishers, Boston, 1999).
[10] M. Hartikainen, K. Miettinen, and M. M. Wiecek, Computational Optimization and Applications **52**, 845–867 (2012).
[11] M. E. Hartikainen and A. Lovison, Journal of Global Optimization **62**, 243–261 (2015).
[12] A. Lovison, SIAM Journal on Optimization **21**, 463–490 (2011).
[13] A. Lovison and F. Pecci, "Hierarchical Stratification of Pareto Sets," http://arxiv-web3.library.cornell.edu/abs/1407.1755 .
[14] S. Piyavskii, {USSR} Computational Mathematics and Mathematical Physics **12**, 57 – 67 (1972).
[15] B. Shubert, SIAM Journal on Numerical Analysis **9**, 379–388 (1972).
[16] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, Journal of Optimization Theory and Applications **79**, 157–181 (1993).