

# On Set-based Local Search for Multiobjective Combinatorial Optimization

Matthieu Basseur  
Université d'Angers  
LERIA, France  
basseur@info.univ-angers.fr

Arnaud Liefoghe  
Université Lille 1  
LIFL – CNRS – INRIA Lille  
arnaud.liefoghe@lifl.fr

Adrien Goëffon  
Université d'Angers  
LERIA, France  
goeffon@info.univ-angers.fr

Sébastien Verel  
Univ. Nice Sophia-Antipolis  
INRIA Lille  
verel@i3s.u-nice.fr \*

## ABSTRACT

In this paper, we formalize a multiobjective local search paradigm by combining set-based multiobjective optimization and neighborhood-based search principles. Approximating the Pareto set of a multiobjective optimization problem has been recently defined as a set problem, in which the search space is made of all feasible solution-sets. We here introduce a general *set-based local search* algorithm, explicitly based on a set-domain search space, evaluation function, and neighborhood relation. Different classes of set-domain neighborhood structures are proposed, each one leading to a different set-based local search variant. The corresponding methodology generalizes and unifies a large number of existing approaches for multiobjective optimization. Preliminary experiments on multiobjective NK-landscapes with objective correlation validates the ability of the set-based local search principles. Moreover, our investigations shed the light to further research on the efficient exploration of large-size set-domain neighborhood structures.

## Categories and Subject Descriptors

I.2.8 [Computing Methodologies]: Artificial Intelligence—Problem Solving, Control Methods, and Search

## General Terms

Algorithms

## Keywords

Multiobjective optimization, Set-based multiobjective search, Local search, Hypervolume, Set-domain neighborhood.

\*All the authors equally contributed to the paper, and are listed in alphabetical order.

## 1. INTRODUCTION

Since the early 2000s, several approaches incorporated quality indicators within the internal principles of multiobjective randomized search procedures. The *implicit* goal is then to identify a population of solutions that minimizes or maximizes a given indicator value, the most popular examples being the hypervolume and the epsilon indicators [5, 8, 10, 16, 17]. More recently, approximating the Pareto set of a multiobjective optimization problem has been formulated as a *set problem* by Zitzler et al. [18]. The search space is then made of *solution-sets* (and not single solutions). This paradigm allows to better capture the dynamics of multiobjective randomized search algorithms [14]. Following this view, and generalizing the concept of Pareto local optimum set [12], the paper attempts to extend the concept of set-based multiobjective optimization [18] with the aim of formalizing a general-purpose *set-based local search* (SbLS) methodology. The goal of the local search process is *explicitly* defined as identifying a solution-set optimizing a given indicator value. The purpose of the current paper is *not* to introduce a strictly-speaking ‘novel’ algorithm, since existing approaches already share similar principles with SbLS. Among others, they include SEMO [11], PLS [12], PAES [10] or also SMS-EMOA [5]. Instead, the proposed SbLS paradigm synthesizes, abstracts and extends a large class of classical multiobjective approaches by formulating or re-formulating them in terms of local search algorithms. Moreover, it provides a unified framework for the design, analysis and comparison of different approaches from a common terminology and classification.

As in the single-objective case, the evaluation function and the neighborhood relation both play an important role on the design of any local search algorithm. The behavior of the search process and the landscape characteristics are, to a large extent, induced by the definition of these two components. With the aim of generalizing such approaches to set-based multiobjective optimization, we give arguments in favor of using a quality indicator as an evaluation function, and we explicitly define a set-domain neighborhood relation. As a consequence, any single-objective metaheuristic can potentially be applied to identify a Pareto set approximation. We highlight the main properties of set-domain neighborhood structures, and we introduce different classes of such

Table 1: Main notations used in the paper.

Notation	Meaning
solution domain	
$X$	set of feasible element-solutions in the decision/search space
$x$	an <i>element-solution</i> , i.e. a feasible solution for the original problem ( $x \in X$ )
$Z$	set of feasible outcome vectors in the objective space
$z$	a feasible objective vector ( $z \in Z$ )
$M$	number of objective functions
$f$	objective function vector ( $f : X \rightarrow Z$ )
$\mathcal{N}$	a solution-domain neighborhood relation ( $\mathcal{N} : X \rightarrow 2^X$ )
set domain	
$\Sigma$	set of feasible solution-sets ( $\Sigma$ is a power set of $X$ )
$\sigma$	a <i>solution-set</i> ( $\sigma \in \Sigma$ )
$\mathbf{I}$	a <i>set-domain evaluation function</i> , i.e. a unary quality indicator ( $\mathbf{I} : \Sigma \rightarrow \mathbb{R}$ )
$\mathbf{I}_H$	the unary hypervolume indicator
$\mathbf{N}$	a <i>set-domain neighborhood relation</i> ( $\mathbf{N} : \Sigma \rightarrow 2^\Sigma$ )
$\mathcal{C}$	an $\mathcal{N}$ -compliant set-domain neighborhood class

relations, that correspond to different resolution methodologies. Afterwards, we argue that the most challenging issue within SbLS approaches, in terms of computational complexity, is related to the neighborhood exploration, due to the very large neighborhood size. As a result, there is a tradeoff to be found between the exploration of large-size neighborhood structures with a higher effectiveness, and the exploration of smaller neighborhood structures with a smaller cost. At last, we investigate the search abilities of a number of SbLS variants induced by different neighborhood relations to produce improving solution-sets, and we compare the performance of the algorithm with respect to different set-domain neighborhood relations for bi-objective instances of  $\rho$ MNK-landscapes.

The paper is organized as follows. In Section 2, we introduce the basics of multiobjective optimization as well as set-based multiobjective search, and we give the template of a general set-based local search algorithm. Different classes of set-domain neighborhood relation are proposed in Section 3, together with a discussion on the issue of neighborhood exploration in such context. Experimental results are given in Section 4. The last section concludes the paper and discusses a number of open issues.

## 2. TOWARD SET-BASED LOCAL SEARCH

In this section, we give the main definitions for multiobjective optimization and set-based multiobjective search, and we formalize the high-level template of a set-based local search algorithm. The notations are summarized in Table 1.

### 2.1 Multiobjective Optimization

A multiobjective optimization problem can be defined by an objective vector function  $f = (f_1, f_2, \dots, f_M)$  with  $M \geq 2$ , and a set  $X$  of feasible solutions in the *decision space*. In the combinatorial case,  $X$  is a discrete set. To avoid confusion, we call a feasible solution  $x \in X$  an *element-solution*. Let  $Z = f(X) \subset \mathbb{R}^M$  be the set of feasible outcome vectors in the *objective space*. To each element-solution  $x \in X$  is assigned an objective vector  $z \in Z$  on the basis of the vector function  $f : X \rightarrow Z$  with  $z = f(x)$ . In a maximization context, an element-solution  $x' \in X$  is dominated by an element-solution  $x \in X$ , denoted by  $x' \prec x$ , iff  $\forall i \in \{1, 2, \dots, M\}$ ,  $f_i(x') \leq f_i(x)$  and  $\exists j \in \{1, 2, \dots, M\}$  such that  $f_j(x') < f_j(x)$ . An element-solution  $x^* \in X$  is

said to be *Pareto optimal* (or *efficient, non-dominated*), if there does not exist any other element-solution  $x \in X$  such that  $x^* \prec x$ . The set of all Pareto optimal element-solutions is called the *Pareto set* (or the *efficient set*). Its mapping in the objective space is called the *Pareto front*.

In practice, different resolution approaches exist and rely on the cooperation between the resolution process and the decision making process. In any case, the overall goal is often to identify a set of good-quality element-solutions. One of the most challenging issue in multiobjective optimization is to identify a minimal complete Pareto set, i.e. a single Pareto optimal element-solution mapping to each point of the Pareto front. But generating such a set is usually infeasible, due to the complexity of the underlying problem or to the large number of optima. Therefore, the overall goal is often to identify a good *Pareto set approximation*. Evolutionary multiobjective optimization (EMO) algorithms and other population-based metaheuristics have received a growing interest for such purpose since the late eighties [7].

### 2.2 Set-based Multiobjective Search

Recently, finding a Pareto set approximation has been explicitly stated as a *set problem*. In that sense, most existing EMO algorithms can be seen as hill-climbers performing on sets. Zitzler et al. [18] define a set-based multiobjective optimization problem as a couple search space / preference relation (in the space of solution-sets) by focusing on the objective values of element-solutions.

#### 2.2.1 Set-domain Search Space

The search space  $\Sigma \subset 2^X$  is defined as the set of feasible sets of element-solutions (and not single element-solutions). It is often the case that a maximum cardinality is imposed, such that  $|\sigma| \leq \mu$  for all  $\sigma \in \Sigma$ . We denote an element  $\sigma \in \Sigma$  as a *solution-set*. Under such notations,  $x \in X$  is a feasible element-solution, while  $\sigma \in \Sigma$  is a feasible solution-set. Some examples for the definition of the set-domain search space are given below [14].

- The search space of population-based approaches can be defined as  $\Sigma_{=\mu} = \{\sigma \in 2^X : |\sigma| = \mu\}$ , where  $\mu$  is the population size.
- The search space of approaches using a bounded archive can be defined as  $\Sigma_{\leq \mu} = \{\sigma \in 2^X : |\sigma| \leq \mu\}$ , where  $\mu$  is the maximum archive size.

- The search space of a number of existing dominance-based approaches, where solution-sets of mutually non-dominated element-solutions only are considered, can be defined as  $\Sigma^{\neq} = \{\sigma \in 2^X : \forall x, x' \in \sigma, x \not\prec x'\}$ .
- A search space with the two previous restrictive conditions can also be considered as  $\Sigma_{\leq \mu}^{\neq} = \{\sigma \in 2^X : |\sigma| \leq \mu \text{ and } \forall x, x' \in \sigma, x \not\prec x'\}$ .
- A search space without any restriction is  $\Sigma^* = 2^X$ .

### 2.2.2 Set-domain Evaluation Function

Different interpretations of what is a good Pareto set approximation are possible, and the definition of approximation quality strongly depends on the decision maker preferences. Two classes of relations can be defined over  $\Sigma$  to compare solution-sets: *dominance relations*, like the Pareto dominance relation extended to solution-sets, or *quality indicators*; see e.g. [19]. With the aim of fitting state-of-the-art single-objective local search algorithms to the set of feasible solution-sets, we propose to follow the second proposal as an evaluation function. Indeed, the mobility and the heuristic guidance of the search process requires the definition of a unary evaluation function, assigning a scalar value to any feasible solution-set. As in single-objective optimization, an evaluation function enables to compare solution-sets, but with much more precision than a dominance relation. Indeed, not only it gives a total order among solution-sets, but it also quantifies the absolute quality of solution-sets as well as the differences between any two solution-sets. A set-based dominance relation is generally not satisfying enough to measure interesting problem-related features. Quality indicators, as defined in [19], allow to overcome such a limitation by introducing a complete order between solution-sets, and by quantifying their respective quality with respect to the indicator being used. We here assume that the set preference relation is explicitly given in terms of a unary quality indicator  $I : \Sigma \rightarrow \mathbb{R}$ .

Many unary quality indicators exist in the literature [19]. One of them is the *hypervolume indicator*  $I_H$ , that is to be maximized. It gives the portion of the objective space enclosed by a solution-set  $\sigma \in \Sigma$  and an appropriately chosen reference point  $z^* \in \mathbb{R}^M$ , such that  $z^*$  is dominated by all feasible element-solutions. The hypervolume indicator is one of the most commonly used indicator, due to several theoretical properties [3, 8, 16]. In particular, it is dominance preserving, i.e.  $\forall \sigma, \sigma' \in \Sigma$  such that  $\sigma'$  is dominated by  $\sigma$ :  $I_H(\sigma) \geq I_H(\sigma')$ . Let us also notice that a minimal solution-set maximizing the  $I_H$ -value is a subset of the Pareto set. Many recent randomized multiobjective search algorithms are based on the hypervolume indicator, but most of them operates on the solution-domain [5, 16, 17], with the exception of [4] where a hypervolume-based evolutionary algorithm working at the set level is proposed. Apart from the hypervolume indicator, most existing quality indicators are binary indicators and are not dominance preserving [19]. For instance, distance-based indicators are typically not dominance preserving, even if the convergence to the Pareto front can be achieved in some particular cases [13].

### 2.2.3 Goal of the Search Process

Usual approaches from multiobjective optimization aim at finding a set of compromise element-solutions within the search space of feasible element-solutions. Here, the search

space is made of all possible sets of element-solutions. Consequently, the overall goal is to find an optimal solution-set, considering the search space of solution-sets and a quality indicator to evaluate them. The goal of the search process is then to find (or approximate) a solution-set  $\sigma \in \Sigma$  that maximizes the indicator value. A *set-based optimization problem* can be formalized as follows.

$$\arg \max_{\sigma \in \Sigma} I(\sigma) \quad (1)$$

Therefore,  $I$  can be seen as a function that assigns, to each solution-set, a scalar value reflecting its quality according to the goal of the search process formulated in Problem (1), i.e. an *evaluation function* defined over solution-sets. Notice that this issue is related to the problem of finding a subset of  $\mu$  element-solutions maximizing  $I$ , i.e. the optimal  $\mu$ -distributions of element-solutions with respect to  $I$  [3].

### 2.2.4 Limitations

In their proposal on set-based multiobjective optimization, Zitzler et al. [18] give the definition of a set-based optimization problem, but not of a set-based local search algorithm. In particular, there is no explicit mention to any set-domain fitness or evaluation function, and there is no definition of any set-domain neighborhood operator, then restricting the application of their approach to some ‘*random set mutation*’, or ‘*heuristic set mutation*’, and where ‘*the neighborhood is in principle the entire search space*’ [18]. However, defining a neighborhood structure on solution-sets allows to distinguish between the properties of the search space, and the heuristics used to explore a solution-set neighborhood. Moreover, the definition of a set-domain neighborhood relation strongly impacts the dynamics of set-based multiobjective search algorithms.

## 2.3 Set-based Local Search

In single-objective optimization, a local search algorithm is based on the definition of a triplet  $(X, \mathcal{N}, h)$ . First,  $X$  is the set of admissible element-solutions (i.e. the search space). Second,  $\mathcal{N} : X \rightarrow 2^X$  is a *neighborhood relation*, i.e. a function that assigns a set of element-solutions  $\mathcal{N}(x) \in 2^X$  to any element-solution  $x \in X$ . The set  $\mathcal{N}(x)$  is called the *neighborhood* of  $x$ , and an element-solution  $x' \in \mathcal{N}(x)$  is called a *neighbor* of  $x$ . At last,  $h : X \rightarrow \mathbb{R}$  is an evaluation function.

Like in single-objective optimization, a multiobjective set-based local search algorithm requires a proper definition of (i) a search space, (ii) a neighborhood structure, and (iii) an evaluation function. In this work, following [14], we propose to define a multiobjective set-based local search as a triplet  $(\Sigma, \mathcal{N}, I)$  such that: (i)  $\Sigma \subset 2^X$  is a set of feasible solution-sets, (ii)  $\mathcal{N} : \Sigma \rightarrow 2^\Sigma$  is a neighborhood relation between solution-sets, and (iii)  $I : \Sigma \rightarrow \mathbb{R}$  is a unary quality indicator, i.e. an evaluation function measuring the quality of solution-sets.  $\Sigma$ ,  $\mathcal{N}$ , and  $I$  still need to be defined for the problem at hand. But this is also the case in single-objective optimization, except that they are here defined on the set-domain.

Algorithm 1 gives the pseudo-code of a general class of set-based local search (SbLS) approaches. If the selection strategy is based on a best-improving solution-set, and if a better neighboring solution-set is always accepted, the SbLS algorithm generalizes the hill-climbing algorithm for set-based multiobjective optimization. In such a case, it is

---

**Algorithm 1** SbLS

---

**Input:**  $\sigma \in \Sigma$   
evaluate  $\sigma$  with respect to  $\mathbf{I}$   
**repeat**  
  select  $\sigma'$  such that  $\sigma' \in \mathbf{N}(\sigma)$   
  evaluate  $\sigma'$  with respect to  $\mathbf{I}$   
  **if**  $\text{accept}(\sigma, \sigma')$  **then**  
     $\sigma \leftarrow \sigma'$   
**until**  $\text{stop}(\sigma)$   
**return** best-found  $\sigma$

---

easy to show that the SbLS algorithm terminates, and that the output of the algorithm consists of a set-domain local optimum.

**DEFINITION 1.** A solution-set  $\sigma \in \Sigma$  is said to be a **set-domain local optimum** with respect to a set-domain neighborhood relation  $\mathbf{N}$  and a set-domain evaluation function  $\mathbf{I}$  iff  $\forall \sigma' \in \mathbf{N}(\sigma), \mathbf{I}(\sigma') \leq \mathbf{I}(\sigma)$ .

Clearly, extensions of the SbLS framework can be designed for advanced local search principles, like tabu search, simulated annealing or iterated local search. In the single-objective case, such advanced mechanisms aim at escaping from local optima, which generally hinder the performance of hill-climbing algorithms. However, it is not obvious that set-based multiobjective optimization problems share similar inhibiting characteristics, even if local fronts are a known issue in NSGA-II [7].

### 3. SET-DOMAIN NEIGHBORHOOD

In this section, we discuss different ideas for the definition of a neighborhood relation between solution-sets  $\mathbf{N} : \Sigma \rightarrow 2^\Sigma$ . For the sake of better understanding, we restrict our analysis by defining set-domain neighborhood relations based on a solution-domain neighborhood operator. First, we propose a number of set-domain neighborhood classes. Then, we highlight the challenging issue of neighborhood selection and exploration in such context. At last, we strengthen the similarities of state-of-the-art algorithms within the SbLS approach.

#### 3.1 Neighborhood Classes

First, let us give the definition of an  $\mathcal{N}$ -compliant set-domain neighborhood relation  $\mathbf{N} : \Sigma \rightarrow 2^\Sigma$ , which is based on a solution-domain neighborhood relation  $\mathcal{N} : X \rightarrow 2^X$ .

**DEFINITION 2.** A set-domain neighborhood relation  $\mathbf{N}$  is **compliant** with a solution-domain neighborhood relation  $\mathcal{N}$  ( $\mathcal{N}$ -compliant for short) iff:

$$\forall \sigma \in \Sigma, \forall \sigma' \in \mathbf{N}(\sigma), \forall x' \in (\sigma' \setminus \sigma), \exists x \in \sigma \text{ such that } x' \in \mathcal{N}(x)$$

Let us consider an arbitrary solution-set  $\sigma \in \Sigma$  and an arbitrary neighboring solution-set  $\sigma' \in \mathbf{N}(\sigma)$ . Any element-solution  $x' \in \sigma'$  belongs either to the original solution-set  $\sigma$  or to the neighborhood  $\mathcal{N}(x)$  of an element-solution  $x \in \sigma$ . Note that this relation is not invertible in the general case, since it is not required that all element-solutions from  $\sigma$  are connected, with respect to  $\mathcal{N}$ , with an element-solution from  $\sigma'$ . In particular, the empty solution-set  $\{\}$  may be a neighbor of every solution-set for some definitions of the set-domain search space. Now, let us define the following  $\mathcal{N}$ -compliant set-domain general neighborhood relations.

1.  $\sigma' \in \mathbf{N}^{(1,1)}(\sigma) \Rightarrow |\sigma' \setminus \sigma| \leq 1$  and  $\forall x' \in \sigma' \setminus \sigma, \exists x \in \sigma, x' \in \mathcal{N}(x)$ .
2.  $\sigma' \in \mathbf{N}^{(1,\star)}(\sigma) \Rightarrow \exists x \in \sigma, \forall x' \in \sigma' \setminus \sigma, x' \in \mathcal{N}(x)$ .
3.  $\sigma' \in \mathbf{N}^{(\star,1)}(\sigma) \Rightarrow \exists \sigma_0 \subset \sigma, \exists \phi : \sigma_0 \rightarrow \sigma' \setminus \sigma$  with  $\forall x' \in \sigma' \setminus \sigma, \exists! x \in \sigma_0$  such that  $\phi(x) = x'$  and  $x' \in \mathcal{N}(x)$ .
4.  $\sigma' \in \mathbf{N}^{(\star,\star)}(\sigma) \Rightarrow \forall x' \in \sigma' \setminus \sigma, \exists x \in \sigma$  such that  $x' \in \mathcal{N}(x)$ .

These four generic set-domain neighborhood relations describe the main ways of exploring a set-domain search space. Interestingly, for any solution-set  $\sigma \in \Sigma$ :

$$\mathbf{N}^{(1,1)}(\sigma) \subset (\mathbf{N}^{(1,\star)}(\sigma) \cup \mathbf{N}^{(\star,1)}(\sigma)) \subset \mathbf{N}^{(\star,\star)}(\sigma)$$

By extension, we define the concept of set-domain neighborhood class as follows.

**DEFINITION 3.** A neighborhood relation  $\mathbf{N}$  belongs to the set-domain neighborhood class  $\mathcal{C}(\xi_1, \xi_2)$  with  $\xi_i \in \{1, \star\}$  iff  $\forall \sigma \in \Sigma, \forall \sigma' \in \mathbf{N}(\sigma), \sigma' \in \mathbf{N}^{(\xi_1, \xi_2)}(\sigma)$ .

In other words, a set-domain relation  $\mathbf{N}$  belonging to the  $\mathcal{C}(\xi_1, \xi_2)$  class implies that, for any solution-set  $\sigma \in \Sigma, \mathbf{N}(\sigma) \subset \mathbf{N}^{(\xi_1, \xi_2)}(\sigma)$ . Within set-domain neighborhood relations that belong to the  $\mathcal{C}(1, 1)$  class, at most one element-solution can be added. Neighborhood relations from  $\mathcal{C}(1, \star)$  allow to add multiple element-solutions belonging to the neighborhood of the same element-solution. The  $\mathcal{C}(\star, 1)$  class allows to add neighboring element-solutions from multiple element-solutions, but each neighboring element-solution belongs to the neighborhood of distinct element-solutions. At last, the neighborhood relation  $\mathbf{N}^{(\star, \star)}$ , from the  $\mathcal{C}(\star, \star)$  class, corresponds to the unrestricted definition of an  $\mathcal{N}$ -compliant set-domain neighborhood relation.

Let us notice that, given that solution-sets can be seen as populations of element-solutions, the previous definitions can be extended to population-based operators like recombination, by considering any  $n$ -ary operator  $\mathcal{O}_{n,p} : \Sigma^n \rightarrow 2^{\Sigma^p}$ .

#### 3.2 Neighborhood Selection

At each iteration of the SbLS algorithm, the problem of finding the best neighbor of a solution-set  $\sigma \in \Sigma$  can be formalized as follows.

$$\arg \max_{\sigma' \in \mathbf{N}(\sigma)} \mathbf{I}(\sigma') \quad (2)$$

The goal is to find a solution-set (generally with a fixed or maximum size  $\mu$ ) that achieves the best indicator value among all possible neighboring solution-sets. We denote this problem as the *set-based neighborhood selection* (SbNS) problem. It is obvious to see that the neighborhood structure of  $\mathbf{N}^{(1,\star)}$ ,  $\mathbf{N}^{(\star,1)}$ , and  $\mathbf{N}^{(\star,\star)}$  is too large to be enumerated exhaustively. In particular, considering the bounded set-domain search space  $\Sigma_{\leq \mu}$ , the size of these neighborhood relations are exponential in the solution-domain neighborhood size and the maximal solution-set size  $\mu$ . Moreover, since the  $\mathbf{N}^{(1,1)}$  neighborhood of a solution-set  $\sigma$  contains all its own subsets, even  $\mathbf{N}^{(1,1)}$  is exponential in  $\mu$ . The exhaustive neighborhood exploration is particularly unfeasible for the  $\Sigma^\star$  set-domain search space, since feasible solution-sets themselves can reach an exponential size. As a consequence, there is a compromise to be found between the size and the quality of the neighborhood relation and the computational complexity to explore it.

**Table 2: Several state-of-the-art algorithms for multiobjective optimization as instances of SbLS framework.**

		$\Sigma$	N	I
(1+1)-PAES	[10]	$\Sigma_{\leq \mu}^{\mathcal{N}}$	$\mathcal{N}$ -compliant, $\mathcal{C}(1, 1)$	$\mathbf{I}_H$ (among other variants)
SEMO	[11]	$\Sigma_{\leq \mu}^{\mathcal{N}}$	$\mathcal{N}$ -compliant, $\mathcal{C}(1, 1)$	none (compliant with $\mathbf{I}_H$ )
PLS	[12]	$\Sigma^{\mathcal{N}}$	$\mathcal{N}$ -compliant, $\mathcal{C}(1, \star)$	none (compliant with $\mathbf{I}_H$ )
SMS-EMOA	[5]	$\Sigma_{=\mu}$	not necessarily $\mathcal{N}$ -compliant	$\mathbf{I}_H$ (combined with dominance-depth ranking)

### 3.3 Neighborhood Exploration

A first option to reduce the computational complexity is to evaluate the fitness value of a neighboring solution-set *incrementally*. In particular, the computation of the hypervolume of a solution-set from scratch is known to be  $\#P$ -complete with respect to the number of objective functions [6], whereas the complexity can be reduced in practice by computing the hypervolume contribution of element-solutions, in a similar way than [5]. Still, due to the large neighborhood size, it can remain an intractable task to enumerate the neighborhood exhaustively. As a consequence, a basic best-improvement local search strategy should be avoided for very unrestricted neighborhood relations. Then, in such large neighborhoods, the use of specific neighborhood exploration techniques have to be defined, as often made in single-objective very large-scale neighborhood search [2]. Such procedures aim at identifying an improving neighbor or the best neighbor without enumerating the whole neighborhood. For instance, one can consider a random neighborhood exploration as in a (1 + 1)-EA, a first-improvement local search strategy, or a more problem-specific approach.

When dealing with the hypervolume indicator, a common approach in EMO consists of the following *hypervolume-based greedy heuristic (hgh)*. Let us assume that we are given a number of parent and offspring element-solutions merged into a population that needs to be reduced to a population size  $\mu$ . The *hgh* procedure removes the worst element-solutions with respect to the hypervolume contribution one-by-one until the population size shrinks to  $\mu$  element-solutions [4, 18]. This can easily be adapted to SbNS as follows. First, an unconstrained-size solution-set is generated with element-solutions from the initial solution-set, together with element-solutions from their neighborhood. Then, a subset of element-solutions is selected by applying *hgh*. The way neighboring element-solutions are produced is related to the set-domain neighborhood relation, whereas the selection of the best  $\mu$  element-solutions is related to the neighborhood exploration. However, it is obvious to see that the neighboring solution-set returned by such a procedure is generally not optimal with respect to the SbNS problem given in Eq. (2).

### 3.4 Similarities with Existing Approaches

In this section, we show how a number of state-of-the-art algorithms conveniently fit into the SbLS paradigm proposed in the paper. Indeed, existing approaches can be redefined in terms of set-based local search. In Table 2, we give the set-domain search components of the following algorithms in terms of the SbLS model: SEMO [11], PLS [12], PAES [10] and SMS-EMOA [5].

The set-domain search space of SEMO and PLS consists of mutually non-dominated element-solutions only, without any restriction on the cardinality of solution-sets. Since

there is no bounding mechanism, the goal is to improve the solution-set in terms of Pareto dominance. Adding new non-dominated element-solutions is always better. As a consequence, even if the optimization goal is not explicitly given in terms of a quality indicator, this is similar than optimizing any Pareto-compliant indicator, like the hypervolume. The set-domain neighborhood relation of SEMO belongs to  $\mathcal{C}(1, 1)$ , whereas the one of PLS belongs to  $\mathcal{C}(1, \star)$ . The PAES set-domain search space is made of solution-sets with mutually non-dominated element-solutions and a maximum cardinality. Moreover, one of the PAES variant explicitly aims at improving the hypervolume [10]. The set-domain neighborhood relation of the (1+1)-PAES belongs to  $\mathcal{C}(1, 1)$ . At last, SMS-EMOA also shares similarities with the SbLS framework. The set-domain search space is defined within fixed-size solution-sets, and the set-domain evaluation function is close to  $\mathbf{I}_H$ , except that a dominance-depth ranking [7] is additionally applied to element-solutions. However, the set-domain neighborhood relation is not necessarily  $\mathcal{N}$ -compliant since a recombination operator can be used. Still, it is similar to a  $\mathcal{N}^{(1,1)}$  set-domain neighborhood relation in case of mutations only. Finally, let us remind that many original methods can potentially be designed within the flexible SbLS framework by combining existing set-domain search components in an innovative way, or by designing original ones.

## 4. EXPERIMENTAL ANALYSIS

This section presents an experimental analysis of different approaches from the proposed SbLS framework on a number of bi-objective  $\rho$ MNK-landscape instances of different structures and sizes. As this study presents a different way of tackling multiobjective combinatorial optimization problems, we mostly focus on a proof-of-principles of the SbLS framework by measuring the ability of SbLS algorithms to produce improving solution-sets. We also analyze the behavior of different SbLS formulations rather than comparing their absolute efficiency against state-of-the-art algorithms.

### 4.1 $\rho$ MNK-landscapes

The family of NK-landscapes constitutes a model for constructing single-objective multimodal landscapes [9]. Feasible element-solutions are binary strings of size  $N$ , *i.e.* the decision space is  $X = \{0, 1\}^N$ .  $N$  refers to the problem size (*i.e.* the bit-string length), and  $K$  to the number of variables that influence a particular position from the bit-string (*i.e.* the epistatic interactions). In single-objective NK-landscapes, the objective function  $f : \{0, 1\}^N \rightarrow [0, 1]$  to be maximized is defined as  $f(x) = \frac{1}{N} \sum_{i=1}^N c_i(x_i, x_{i_1}, \dots, x_{i_K})$ , where  $c_i : \{0, 1\}^{K+1} \rightarrow [0, 1]$  defines the component function associated with each variable  $x_i$ ,  $i \in \{1, \dots, N\}$ , and where  $K < N$ . By increasing the number of variable interactions

**Table 3: Parameter setting.**

problem instances		
$\rho$	$\in$	$\{-0.5, 0.0, +0.5\}$ , s.t. $\rho \geq -1/(M-1)$
$N$	$\in$	$\{256, 512\}$
$K$	$\in$	$\{2, 4\}$
problem domain		
$X$	$=$	$\{0, 1\}^N$
$f$	$=$	analytic objective functions
$\mathcal{N}$	$=$	1-bit flip neighborhood
set domain		
$\Sigma$	$=$	$\Sigma_{\leq \mu}$ , with $\mu = 100$
$\mathbf{I}$	$=$	$\mathbf{I}_H$ , with $z^{ref} = (0, 0)$
$\mathbf{N}$	$\in$	$\{\hat{\mathbf{N}}_{hgh}^{(1,1)}, \hat{\mathbf{N}}_{hgh}^{(1,*)}, \hat{\mathbf{N}}_{hgh}^{(*,1)}, \hat{\mathbf{N}}_{hgh}^{(*,*)}\}$

$K$  from 0 to  $(N-1)$ , NK-landscapes can be gradually tuned from smooth to rugged. In this work, we set the position of these interactions at random. Component values are uniformly distributed in the range  $[0, 1]$ .

Multiojective NK-landscapes with a set of  $M$  independent objective functions are defined in [1]. More recently, multiojective NK-landscapes with correlated objective functions have been proposed [15]. Component values follow a multivariate uniform law of dimension  $M$ , defined by a correlation matrix. We here consider the same correlation between all pairs of objective functions, given by a correlation coefficient  $\rho > \frac{-1}{M-1}$ . The same epistasis degree  $K_m = K$  is used for all  $m \in \{1, \dots, M\}$ . An instance generator and the problem instances under study in this paper can be found at the following URL: <http://mocobench.sf.net/>. For more details on  $\rho$ MNK-landscapes, the reader is referred to [15].

## 4.2 Experimental Setup

In the following, we conduct an experimental study on the influence of the problem size ( $N$ ), the non-linearity ( $K$ ), and the objective correlation ( $\rho$ ) for bi-objective  $\rho$ MNK-landscapes ( $M = 2$ ) on the performance the SbLS algorithm used in conjunction with the set-domain neighborhood relations under study in the paper. The parameters defining the problem instances under investigation are given in Table 3. One instance, generated at random, is considered *per* parameter setting.

The most simple local search variant of the SbLS framework given in Algorithm 1 is experimented with different set-domain neighborhood relations. In particular, a set-based  $(1+1)$ -EA is considered. We evaluate the ability of such a simple SbLS algorithm to improve the current solution-set along the iterations. The set-domain search space is  $\Sigma_{\leq \mu}$ . It is made of solution-sets with mutually non-dominated element-solutions only, and a variable size with at most  $\mu$  element-solutions. The maximum solution-set size is set to  $\mu = 100$ . Initial solution-sets are generated at random, each element-solution from the solution-set being a random binary string, of size  $N$ . The set-domain evaluation function is the hypervolume  $\mathbf{I}_H$ . The reference point used for the hypervolume calculation is set to the origin  $(0, 0)$ . The solution-domain neighborhood structure  $\mathcal{N}$  is the 1-bit flip neighborhood. In addition, a comparison with a conventional EMO algorithm, namely NSGA-II [7], is also provided in order to facilitate the result analysis. Let us notice that both SbLS and NSGA-II share equivalent parameter settings,  $\mu = 100$  being an appropriate population size for NSGA-II [7].

The process of the SbLS algorithm under consideration

works as follows. At each iteration, a random neighboring solution-set is evaluated, and is accepted only if its hypervolume set-domain fitness value is strictly improving over the current solution-set. Each neighborhood operator adds a number of neighboring element-solutions (drawing with replacement), and then reduces the solution-set to  $\mu$  element-solutions following to the hypervolume-based greedy heuristic (*hgh*) described in Section 3.3. Notice that, whenever there exists a dominated element-solution, it is removed from the solution-set. We consider four set-domain neighborhood relations, each one being derived from a set-domain neighborhood class defined in Section 3.1. Thus,  $\forall \sigma \in \Sigma$ :

1.  $\hat{\mathbf{N}}_{hgh}^{(1,1)}(\sigma) = \{hgh(\sigma \cup \{x'\}) : x \in \sigma, x' \in \mathcal{N}(x)\}$ .
2.  $\hat{\mathbf{N}}_{hgh}^{(1,*)}(\sigma) = \{hgh(\sigma \cup \mathcal{N}(x)) : x \in \sigma\}$ .
3.  $\hat{\mathbf{N}}_{hgh}^{(*,1)}(\sigma) = \{hgh(\sigma \cup_{x \in \sigma} \{x'_x\}) : \forall x \in \sigma, x'_x \in \mathcal{N}(x)\}$ .
4.  $\hat{\mathbf{N}}_{hgh}^{(*,*)}(\sigma) = \{hgh(\sigma \cup_{x \in \sigma} \mathcal{N}(x))\}$ .

Let us notice that the set-domain neighborhood relation  $\hat{\mathbf{N}}_{hgh}^{(\xi_1, \xi_2)}$  belongs to the  $\mathcal{C}(\xi_1, \xi_2)$  class. Each set-domain neighborhood relation tries to add 1 or  $N$  neighboring element-solution(s) from 1 or  $\mu$  element-solution(s) with respect to the set-domain neighborhood relation under consideration. As discussed in Section 3.4, those different SbLS variants share similar principles with existing state-of-the-art algorithms like SEMO, PAES, PLS and SMS-EMOA. The CPU time limit used as a stopping condition is set to  $10^4$  seconds, a voluntarily-high value in order to ensure the convergence of all the algorithms to a set-domain local optimum.

## 4.3 Computational Results

Table 4 reports the average hypervolume value obtained for each couple ( $\rho$ MNK-landscape, neighborhood class) over a set of 30 independent simulation runs. We use 30 different random seeds, each one leading all the algorithms to start with the same initial solution-set. For a given instance, the values in bold correspond to algorithms which are not statistically outperformed by any other approach with respect to a Mann-Whitney signed rank statistical test with a  $p$ -value of 0.05. Let us remind that both the solution-domain neighborhood relation and the overall method remain unchanged, whereas only the set-domain neighborhood relations differ.

First of all, the hypervolume values reached by the four neighborhood classes are quite close with each other. Globally, one see that the results obtained by SbLS, whatever the neighborhood structure, validate the accuracy of the approach. Indeed, when compared against NSGA-II, there exists a significant superiority of SbLS for conflicting and independent objective functions. On the contrary, when the objective functions are positively correlated ( $\rho = 0.5$ ), NSGA-II shows comparable or superior results. However, let us notice that such a correlation degree is less representative of the multiojective nature of the optimization problem. Indeed, the higher the objective correlation, the closer the multiojective optimization problem to a single-objective one. As a consequence, in such a case, Pareto-based approaches are not necessarily appropriate from a decision making point of view.

Considering the four set-domain neighborhood classes, one can observe that SbLS- $\hat{\mathbf{N}}_{hgh}^{(*,*)}$  is never statistically outperformed by any other approach. This result validates the interest of defining a wide set-domain neighborhood relation.

**Table 4: Average  $I_H$ -value obtained by the different variants of the SbLS algorithm and NSGA-II.**

			SbLS- $\hat{N}_{hgh}^{(1,1)}$	SbLS- $\hat{N}_{hgh}^{(1,*)}$	SbLS- $\hat{N}_{hgh}^{(*,1)}$	SbLS- $\hat{N}_{hgh}^{(*,*)}$	NSGA-II
$\rho = -0.5$	$N = 256$	$K = 2$	<b>0.5068</b>	0.5019	<b>0.5074</b>	<b>0.5074</b>	0.4438
		$K = 4$	0.5189	0.5124	0.5203	<b>0.5231</b>	0.4806
	$N = 512$	$K = 2$	<b>0.4915</b>	0.4880	0.4897	<b>0.4905</b>	0.4012
		$K = 4$	<b>0.5186</b>	0.5106	<b>0.5180</b>	<b>0.5187</b>	0.4440
$\rho = 0.0$	$N = 256$	$K = 2$	0.5161	0.5129	0.5169	<b>0.5193</b>	0.5032
		$K = 4$	<b>0.5453</b>	0.5395	0.5444	<b>0.5433</b>	0.5373
	$N = 512$	$K = 2$	<b>0.5155</b>	0.5095	<b>0.5156</b>	<b>0.5159</b>	0.4791
		$K = 4$	0.5417	0.5396	<b>0.5441</b>	<b>0.5430</b>	0.5201
$\rho = +0.5$	$N = 256$	$K = 2$	0.5264	0.5231	0.5276	0.5296	<b>0.5333</b>
		$K = 4$	0.5518	0.5494	<b>0.5539</b>	<b>0.5535</b>	<b>0.5593</b>
	$N = 512$	$K = 2$	0.5213	0.5161	0.5208	<b>0.5228</b>	0.5184
		$K = 4$	0.5526	0.5444	0.5528	0.5538	<b>0.5589</b>

The  $\hat{N}_{hgh}^{(1,*)}$  neighborhood seems to be the weakest one here: it is always statistically outperformed by at least one other method for all the instances experimented. We attribute this to the neighborhood selection operator, the same for the four neighborhood classes, which is penalizing SbLS- $\hat{N}_{hgh}^{(1,*)}$ .

Another point of interest deals with the convergence rate of the SbLS approaches. Figure 1 shows the evolution of the average hypervolume value over the 30 runs throughout the search process. Since the shape of the convergence rates mainly depends on  $\rho$ , only the plots for  $N = 256$  and  $K = 2$  are pictured. For a better view of the earliest stages of the search process, the running time is displayed in a logarithmic scale. Unsurprisingly, the iterations of SbLS- $\hat{N}_{hgh}^{(*,*)}$  are time-consuming. Its global convergence is then the slowest one. On the contrary, using the  $\hat{N}_{hgh}^{(*,1)}$  neighborhood allows to reach good-quality solution-sets rapidly, and clearly converges faster, but to a lower hypervolume value.

## 5. CONCLUSIONS

This paper extends the idea of set-based multiobjective optimization [18] by introducing a general set-based local search (SbLS) methodology. This view is motivated by the fact that the expected output of a multiobjective approach consists of a set of (approximate) non-dominated element-solutions. The flexible SbLS framework is based on a set-domain search space that consists of potential solution-sets rather than single element-solutions. The set-domain evaluation function consists of a unary quality indicator, like the hypervolume. In addition, we defined a number of neighborhood relations between solution-sets. The corresponding SbLS variants formalize and share similarities with state-of-the-art multiobjective approaches, but also enables the design of original methodologies. Our preliminary experiments on bi-objective  $\rho$ MNK-landscapes reveal the ability of basic SbLS approaches to produce improving solution-sets, and to provide competitive results in terms of problem solving. Different definitions of a set-domain neighborhood relation leads to different search process dynamics. The set-domain neighborhood relation to be used needs to be carefully selected with respect to the computational resources available.

As a future work, we plan to make a stronger link between the SbLS framework and state-of-the-art algorithms. Hopefully, this will allow a better understanding of the search process behavior, and of the characteristics of multiobjective optimization problems [14]. In particular, the set-domain neighborhood relations introduced in the paper can be ex-

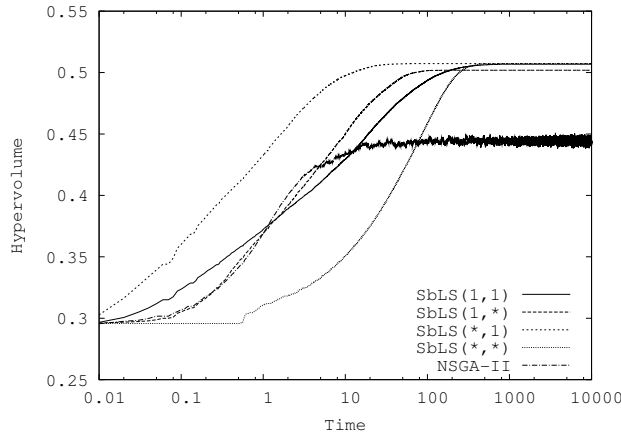
tended with higher degrees of reduction or expansion over the solution-set cardinality, and with more general solution-domain operators like recombination. However, one of the main computational issue within such very-large set-domain neighborhood structures deals with their exploration, where there is a clear need for improvement over the algorithm efficiency. Furthermore, SbLS variants could largely be improved with more advanced metaheuristic principles like tabu search or simulated annealing. In particular, restart mechanisms or iterated local search extensions would allow to appreciate the benefit of each set-domain neighborhood class with an opportunity to escape from (set-domain) local optima. At last, extending the experimental analysis would give further insights on solving different multiobjective optimization problem classes, specifically for problems with more than two objective functions and against more recent indicator-based algorithms like SMS-EMOA [5] or HypE [4].

## Acknowledgements.

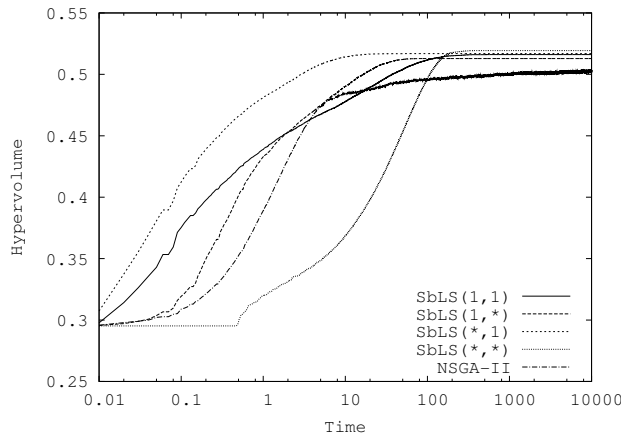
This work was partially supported by the "Pays de la Loire" Region within the LigeRO project.

## 6. REFERENCES

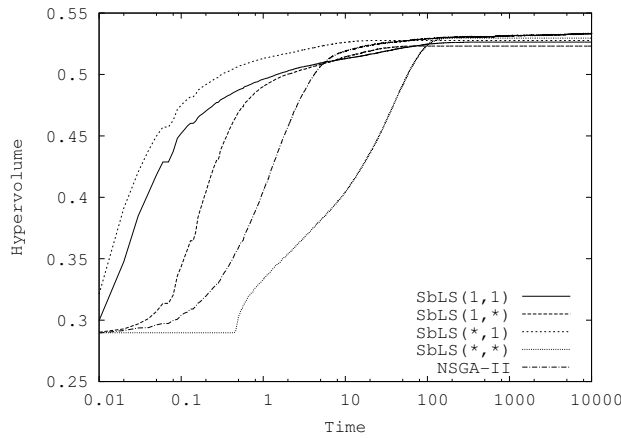
- [1] H. E. Aguirre and K. Tanaka. Working principles, behavior, and performance of MOEAs on MNK-landscapes. *European Journal of Operational Research*, 181(3):1670–1690, 2007.
- [2] R. K. Ahuja, O. Ergun, J. B. Orlin, and A. P. Punnen. A survey of very large-scale neighborhood search techniques. *Discrete Applied Mathematics*, 123(1-3):75–102, 2002.
- [3] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Hypervolume-based multiobjective optimization: Theoretical foundations and practical implications. *Theoretical Computer Science*, 425:75–103, 2012.
- [4] J. Bader and E. Zitzler. HypE: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation*, 19(1):45–76, 2011.
- [5] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007.
- [6] K. Bringmann and T. Friedrich. Approximating the volume of unions and intersections of high-dimensional geometric objects. *Computational Geometry - Theory and Applications*, 43:601–610, 2010.



(a)  $\rho = -0.5$



(b)  $\rho = 0.0$



(c)  $\rho = +0.5$

**Figure 1: Evolution of the average  $I_H$ -value for  $N=256$ ,  $K=2$ ,  $\rho \in \{-0.5, 0, +0.5\}$ .**

- [7] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, Chichester, UK, 2001.
- [8] M. Fleischer. The measure of Pareto optima: Applications to multi-objective metaheuristics. In *2nd International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, volume 2632 of *Lecture Notes in Computer Science*, pages 519–533, Faro, Portugal, 2003. Springer.
- [9] S. A. Kauffman. *The Origins of Order*. Oxford University Press, New York, USA, 1993.
- [10] J. D. Knowles. *Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization*. PhD thesis, University of Reading, UK, 2002.
- [11] M. Laumanns, L. Thiele, E. Zitzler, E. Welzl, and K. Deb. Running time analysis of multi-objective evolutionary algorithms on a simple discrete optimization problem. In *Conference on Parallel Problem Solving From Nature (PPSN VII)*, volume 2439 of *Lecture Notes in Computer Science*, pages 44–53, Granada, Spain, 2002. Springer.
- [12] L. Paquete, T. Schiavinotto, and T. Stützle. On local optima in multiobjective combinatorial optimization problems. *Annals of Operations Research*, 156(1):83–97, 2007.
- [13] O. Schütze, X. Esquivel, A. Lara, and C. A. Coello Coello. Using the averaged Hausdorff distance as a performance measure in evolutionary multi-objective optimization. *IEEE Transactions on Evolutionary Computation*, 16(4):504–522, 2012.
- [14] S. Verel, A. Liefooghe, and C. Dhaenens. Set-based multiobjective fitness landscapes: a preliminary study. In *13th Genetic and Evolutionary Computation Conference (GECCO 2011)*, pages 769–776, Dublin, Ireland, 2011. ACM.
- [15] S. Verel, A. Liefooghe, L. Jourdan, and C. Dhaenens. Analyzing the effect of objective correlation on the efficient set of MNK-landscapes. In *Learning and Intelligent Optimization (LION 5)*, volume 6683 of *Lecture Notes in Computer Science*, pages 238–252, Rome, Italy, 2011. Springer.
- [16] E. Zitzler, D. Brockhoff, and L. Thiele. The hypervolume indicator revisited: On the design of Pareto-compliant indicators via weighted integration. In *4th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, volume 4403 of *Lecture Notes in Computer Science*, pages 862–876, Matsushima, Japan, 2007. Springer.
- [17] E. Zitzler and S. Künzli. Indicator-based selection in multiobjective search. In *Conference on Parallel Problem Solving from Nature (PPSN VIII)*, volume 3242 of *Lecture Notes in Computer Science*, pages 832–842, Birmingham, UK, 2004. Springer.
- [18] E. Zitzler, L. Thiele, and J. Bader. On set-based multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 14(1):58–79, 2010.
- [19] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.