

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Tabatabaei, Mohammad; Lovison, Alberto; Tan, Matthias; Hartikainen, Markus; Miettinen, Kaisa

**Title:** ANOVA-MOP : ANOVA Decomposition for Multiobjective Optimization

**Year:** 2018

**Version:** Accepted version (Final draft)

**Copyright:** © 2018, Society for Industrial and Applied Mathematics.

**Rights:** In Copyright

**Rights url:** <http://rightsstatements.org/page/InC/1.0/?language=en>

**Please cite the original version:**

Tabatabaei, M., Lovison, A., Tan, M., Hartikainen, M., & Miettinen, K. (2018). ANOVA-MOP : ANOVA Decomposition for Multiobjective Optimization. *SIAM Journal on Optimization*, 28(4), 3260-3289. <https://doi.org/10.1137/16M1096505>

# ANOVA-MOP: ANOVA DECOMPOSITION FOR MULTIOBJECTIVE OPTIMIZATION\*

MOHAMMAD TABATABAEI<sup>†</sup>, ALBERTO LOVISON<sup>‡</sup>, MATTHIAS TAN<sup>§</sup>, MARKUS HARTIKAINEN<sup>†</sup>,  
AND KAISA MIETTINEN<sup>†</sup>

**Abstract.** Real-world optimization problems may involve a number of computationally expensive functions with a large number of input variables. Metamodel-based optimization methods can reduce the computational costs of evaluating expensive functions, but this does not reduce the dimension of the search domain nor mitigate the *curse of dimensionality* effects. The dimension of the search domain can be reduced by functional ANOVA decomposition involving Sobol' sensitivity indices. This approach allows ranking decision variables according to their impact on the objective function values. On the basis of the *sparsity of effects principle*, typically only a small number of decision variables significantly affects an objective function. Therefore, neglecting the variables with the smallest impact should lead to an acceptably accurate and simpler metamodel for the original problem. This appealing strategy has been applied only to single-objective optimization problems so far. Given a high-dimensional optimization problem with multiple objectives, a method called ANOVA-MOP is proposed for defining a number of independent low-dimensional subproblems with a smaller number of objectives. The method allows to define approximated solutions for the original problem by suitably combining the solutions of the subproblems. The quality of the approximated solutions and both practical and theoretical aspects related to decision making are discussed.

**Key words.** Multiple criteria optimization, Sensitivity analysis, Metamodeling, Dimensionality reduction, Pareto optimality, Decision making

**AMS subject classifications.** 90C29, 90C59

**1. Introduction.** Multiobjective optimization problems (MOPs) arise in many application domains such as engineering and finance. Their characteristic is the co-existence of a number of typically conflicting objective functions. Real-world MOPs may contain black-box functions requiring computationally expensive and/or time-consuming experiments and/or simulations [55]. Because of the conflict among the objectives, MOPs do not usually have a unique optimal solution but several so-called Pareto optimal solutions. Mathematically, such solutions are incomparable and involve different trade-offs among the objectives. Typically, it is the task of a *decision maker* (DM) to identify a preferred solution to be implemented.

As discussed, e.g., in [38, 39, 40], multiobjective optimization methods are classified into interactive and non-interactive methods, and also according to the role of the DM. In non-interactive methods, the preference information of the DM either is not available or it is given only before or after the solution process. E.g., in a posteriori methods, a representative set of Pareto optimal solutions is generated to be analyzed by the DM. On the other hand, in interactive methods, the DM is asked to specify her/his preference information iteratively during the solution process: (s)he can examine the solutions generated on the basis of her/his preferences and can

---

\*Submitted to the editors DATE.

**Funding:** This work was partly funded by the Academy of Finland project No. 287496, Early Career Scheme (ECS) Project No. 21201414, the General Research Fund Project No. 11226716 sponsored by the Research Grants Council of Hong Kong and the KAUTE Foundation.

<sup>†</sup>University of Jyväskylä, Faculty of Information Technology, P.O. Box 35 (Agora), FI-40014 University of Jyväskylä, Finland ([tabatabaei62@yahoo.com](mailto:tabatabaei62@yahoo.com), [hartikainen.markus@gmail.com](mailto:hartikainen.markus@gmail.com), [kaisa.miettinen@jyu.fi](mailto:kaisa.miettinen@jyu.fi)).

<sup>‡</sup>Dipartimento di Matematica, Università degli Studi di Padova, Via Trieste, 63, 35121 Padova, Italy ([alberto.lovison@gmail.com](mailto:alberto.lovison@gmail.com)).

<sup>§</sup>School of Data Science, City University of Hong Kong ([mathtan@cityu.edu.hk](mailto:mathtan@cityu.edu.hk)).

revise them according to the updated understanding. The interaction with the DM is continued until (s)he finds the most preferred solution. An example of preference information is a reference point involving aspiration levels representing desirable values for objective functions and it is employed, e.g., in the reference point method [64].

When a problem has black box objective functions, their behavior is unknown, excluding basic regularity hypotheses. In such cases, heuristic methods may lead to satisfactory solutions, but cannot guarantee convergence in a global and local sense. Indeed, if the dimension of the decision space is high, a satisfactory exploration can only be obtained with an exponentially large number of function evaluations. This can be computationally costly even if function evaluations are computationally inexpensive and becomes virtually impossible if function evaluations involve complex simulations, as e.g., in fluid dynamics. Therefore, large regions of the feasible set may remain unexplored and important parts of the set of Pareto optimal solutions may remain undetected even after thousands of function evaluations. This phenomenon is known as *the curse of dimensionality* [66].

An example of a real-world MOP involving many decision variables and objective functions is the design and optimization of a car restraint system [69]. Assessing the safety of a car requires several tests estimating different types of injuries suffered by the occupants or by pedestrians in case of an accident. Actually, the New Car Assessment Programme (NCAP) tests have tens of injury indicators. The aim of the designers is to have as low injury indicators as possible, or at least go under a predefined threshold. Once the design process has reached a configuration close enough to the set of Pareto optimal solutions, improving an objective function value may impair some other objective function values. Therefore, bringing all the objective functions simultaneously under desired thresholds may be an overwhelming task. Once an accurate virtual model of a car, occupants and driving situations has been prepared involving aerodynamics, mechanics and biomechanics, it is virtually impossible to get some useful insight in the problem because of the many interdependencies. Thus, making a reliable prediction on the model behavior under decision variable changes requires decision support.

The paradigmatic problem described above suggests a possible strategy for tackling the dimensionality issues. Indeed, the restraint system is built by combining subunits that were originally designed to act separately on specific regions of the car, i.e., to mitigate the severity of different injuries possibly occurring in an accident. When these modules are put together, they can interact and produce some extra side effects on each other. Nevertheless, the side effects are likely to be of secondary importance, therefore it seems possible to disassemble a high-dimensional problem with many objectives by grouping together subunits related to similar tasks without losing effects of primal importance. This decomposition leads to a collection of smaller and almost independent subproblems that are tractable with computationally effective optimization methods, as the domains of the subproblems are easier to be explored. Furthermore, it should be easier to get insight on the behavior of objective functions as decision variables are varied in their ranges.

In this paper, we propose a method called ANOVA-MOP for defining essentially independent low-dimensional multiobjective subproblems starting from a multiobjective optimization problem with many decision variables and objective functions. The method is suitable for computationally expensive problems, and can be applied either as a non-interactive or an interactive method. In the latter case, when solving computationally expensive MOPs, the DM does not need to wait for a long time to

see solutions corresponding to her/his preferences. Indeed, at each iteration with a DM, (s)he provides his/her preferences in the form of a reference point for a small number of objective functions rather all of them at the same time. After solving the subproblems (with appropriate optimization algorithms), we describe how to re-assemble the solutions to build an approximated solution for the original high-dimensional problem. Finally, we discuss the trade-off between the dimensionality reduction and the accuracy of the approximation.

## 2. Background.

**2.1. Multiobjective optimization.** Without loss of generality we consider a real vector-valued function  $\mathbf{f}$  defined on a  $d$ -dimensional unit hypercube:

$$\mathbf{f} : S = [0, 1]^d \longrightarrow \mathbb{R}^k, \quad \mathbf{x} = (x_1, \dots, x_d)^\top \longmapsto (f^1(\mathbf{x}), \dots, f^k(\mathbf{x}))^\top,$$

and denote by

$$(1) \quad \underset{\mathbf{x} \in S}{\text{minimize}} \{f^1(\mathbf{x}), \dots, f^k(\mathbf{x})\},$$

the associated *multiobjective optimization problem* (MOP) with  $k$  objective functions. The set  $S$  is called a *feasible set* in the decision space and  $x_1, \dots, x_d$  are called *decision variables or inputs*. The space  $\mathbb{R}^k$  is called an *objective space* and the scalar-valued components  $f^1, \dots, f^k$  of  $\mathbf{f}$  are called *objective functions or outputs*.

We say that the point  $\mathbf{x} \in S$  (correspondingly  $\mathbf{f}(\mathbf{x})$ ) dominates  $\mathbf{y} \in S$  ( $\mathbf{f}(\mathbf{y})$ ) if for all  $\ell = 1, \dots, k$ ,  $f^\ell(\mathbf{x}) \leq f^\ell(\mathbf{y})$  and for some  $1 \leq h \leq k$ ,  $f^h(\mathbf{x}) < f^h(\mathbf{y})$ . Furthermore, a point  $\bar{\mathbf{x}} \in S$  is called a *Pareto optimum* for problem (1), if there does not exist another point  $\mathbf{x} \in S$  which dominates  $\bar{\mathbf{x}}$ . If  $k > 1$ , usually there are many Pareto optima forming a *Pareto optimal set* in the decision space, denoted by  $P \subseteq S$ . The image of  $P$  is a *Pareto frontier* denoted by  $F = \mathbf{f}(P) \subseteq \mathbb{R}^k$ .

**2.2. Multiobjective decomposition.** One way to deal with high-dimensional MOPs is decomposition, i.e., partitioning decision variables and objective functions to form a finite number of lower-dimensional subproblems. For computationally inexpensive MOPs, decomposition has been studied e.g., in [2, 8, 10, 12, 15, 26, 27, 35, 46, 61, 71]. In [14], comprehensive theoretical results related to multiobjective decomposition are provided. Decision making in the context of computationally inexpensive multiobjective decomposition has been studied in [13].

In engineering design, to solve high-dimensional computationally expensive MOPs, multiobjective optimization methods may integrate decomposition approaches such as concurrent subspace optimization [52], collaborative optimization [5], integrated system synthesis [53] and analytical target cascading [24]. For example, the integration of the weighted sum method within the collaborative optimization approach and the linear physical programming and the collaborative optimization approach have been studied in [59] and [36], respectively. The integration of multiobjective optimization methods and the concurrent subspace optimization approach has been considered e.g., in [18, 19, 20]. As mentioned in [68], the final output of these methods based on the concurrent subspace optimization approach is only one solution. To generate a set of solutions, the methods should be rerun with different initial starting points. For more details see e.g., [27, 29, 28, 30, 43, 51, 56, 58, 60, 67, 71, 70]. These methods do not study the theoretical connections between the solutions of subproblems and the original problem.

A common assumption among the above mentioned methods (for both computationally inexpensive and expensive MOPs) is that a given problem has already been decomposed and information of the connections among the subproblems is available. When dealing with computationally expensive functions, the mathematical closed forms of objective functions are typically not available [50, 55]. This means that for a given input, one should conduct time-consuming experiments and/or simulations to get output. Therefore, obtaining information (without proper techniques and strategies) which results in decomposing the original problem can be very demanding and time-consuming or impossible. As mentioned earlier, in this paper, we introduce a method to obtain information to decompose high-dimensional, computationally expensive MOPs and form lower-dimensional simpler subproblems.

**2.3. Metamodeling and curse of dimensionality.** An important element in the ANOVA-MOP method is the use of metamodels for the objective functions. Metamodels are approximating functions defined by using, e.g., polynomial splines, Gaussian processes, neural networks, Fourier expansions and many other methods. They are also known as *response surfaces* [4, 66] when used for approximating physical experiments and as *surrogate models* or *predictors* [48] when applied to computer simulations. Typically, they are used for relieving the computational cost of complex simulations and the resources (or time) costs of physical experiments. (For more details, see, e.g., [1, 16, 17, 65, 72] and [55] for a survey.)

Although metamodels can reduce the cost of function evaluations, the intrinsic complexity in high-dimensional decision and objective spaces is not affected, i.e., the curse of dimensionality is not tamed. Indeed, guaranteeing global convergence towards optima is exponentially complex with respect to the dimension of the decision space, even for single objective optimization (see, e.g., [32, 49]). When we say that the complexity of a problem grows exponentially with the number of decision variables, we mean that the exploration of a  $d$ -dimensional hypercube by means of a full factorial design with  $m$  levels (i.e., discretizing each variable range to  $m$  equally-spaced sections) in each variable requires  $m^d$  points [3].

Furthermore, the exploration and representation of the Pareto frontier also suffers from the high dimension of the objective space. The Pareto frontier is usually a  $(k - 1)$ -dimensional object [31]. Even if it is approximated with a satisfactory accuracy, it is impossible to visualize it in the Euclidean space  $\mathbb{R}^k$  when  $k > 4$ .

**2.4. Sparsity of effects, reducibility and decomposability.** If possible, the dimensions of the problem could be decreased by reducing the number of variables and objectives. We assume that the problem designers have not introduced spurious objectives that increase the problem complexity without bringing essential information. As objectives are usually in conflict with each other, especially close to the Pareto optimal points, we assume that there are no redundant objective functions.

As far as decision variables are concerned, the situation is different. Indeed, along with the curse of dimensionality, there is a complementary phenomenon known as the *sparsity of effects principle* ([3], [66, p.173]). It states that in typical cases, i.e., for commonly encountered real-world functions, there are only a limited number of decision variables that are responsible for the largest part of a function variation. Roughly we can say that 80% of the variance of a function is due to 20% of the decision variables. This 80 – 20 empirical rule, popular for a wide range of phenomena, is also known as the *Pareto principle* [44]. In this paper, we assume this principle holds, i.e., that a large fraction of the decision variables of each objective function can be neglected without losing essential information. From now on, we adopt this

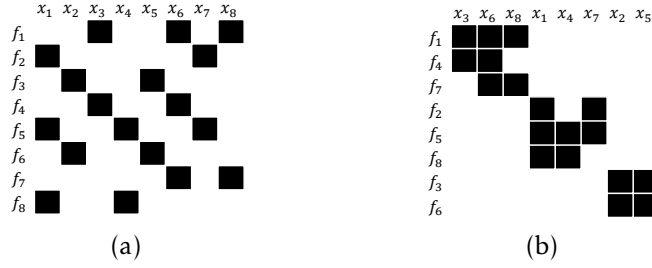


Figure 1: (a) Incidence matrix  $M$ . (b) A decomposition corresponding to  $M$  in (a).

assumption.

If we can neglect 80% of the decision variables in each objective function, we have a large probability that many variables can be neglected in all the objectives at the same time. This leads to the definition of an approximated MOP with the same number of objectives but with a smaller number of decision variables. We call such problems *reducible*.

On the other hand, e.g., in the example of the car restraint system, a variable that is not important for one objective can be very important for another objective. So if every variable is important for at least one objective function there are no negligible variables in the whole problem. This hinders the possibility of reducing the number of variables overall, but enables grouping objectives that depend strongly on the same variables. In that case, essentially independent blocks of objectives and variables can appear, transforming the original high-dimensional problem to a collection of lower-dimensional subproblems. We call such problems *decomposable*.

In the literature, the possibility of exploiting decomposition in multiobjective optimization has been explored for at least thirty years (see, e.g., [2, 15, 26, 27, 71]). Comprehensive mathematical results and a literature review are provided in [13, 14]. Usually, a decomposition is given a priori along with the problem definition. However, this is not usually possible when objective functions are based on physical experiments or complex computer simulations. In the next section, by means of ANOVA, we propose a method to detect the possible decomposable or reducible structure in black box functions. We start by giving some definitions.

**DEFINITION 1.** Let the incidence matrix  $M = [m_i^\ell]$ , with  $1 \leq \ell \leq k$  and  $1 \leq i \leq d$ , be defined as  $m_i^\ell = 1$ , if function  $f^\ell$  does depend on variable  $x_i$  and  $m_i^\ell = 0$ , if function  $f^\ell$  does NOT depend on variable  $x_i$ . If  $m_i^\ell = 0$ , the variable  $x_i$  is called inactive for the objective  $f^\ell$ . Otherwise, it is active.

**DEFINITION 2.** If there exists at least one index  $i$ ,  $1 \leq i \leq d$  such that  $m_i^\ell = 0$  for all  $\ell \in \{1, \dots, k\}$ , problem (1) is said to be reducible.

Figure 1(a) depicts the incidence matrix  $M$  for a MOP with  $k = 8$  and  $d = 8$ , where the rows correspond to objective functions and the columns to decision variables. The black and white squares represent matrix entries with values 1 and 0, respectively. We first define *composition* and then the decomposability of problems.

DEFINITION 3. Let  $\mathbf{f}^{(1)} : [0, 1]^{d_1} \rightarrow \mathbb{R}^{k_1}$  and  $\mathbf{f}^{(2)} : [0, 1]^{d_2} \rightarrow \mathbb{R}^{k_2}$ . The function

$$\mathbf{f}^{(1)} \otimes \mathbf{f}^{(2)} : [0, 1]^{d_1+d_2} \rightarrow \mathbb{R}^{k_1+k_2}, \quad \mathbf{f}^{(1)} \otimes \mathbf{f}^{(2)}(x_1, \dots, x_{d_1+d_2}) := \left( f_1^{(1)}(x_1, \dots, x_{d_1}), \dots, f_{k_1}^{(1)}(x_1, \dots, x_{d_1}), f_1^{(2)}(x_{d_1+1}, \dots, x_{d_1+d_2}), \dots, f_{k_2}^{(2)}(x_{d_1+1}, \dots, x_{d_1+d_2}) \right)^\top$$

is called the composition of  $\mathbf{f}^{(1)}$  and  $\mathbf{f}^{(2)}$ .

DEFINITION 4. Problem (1) is decomposable if, after a reordering of variables and objectives, there exist two subproblems with  $\mathbf{f}^{(1)}$  and  $\mathbf{f}^{(2)}$  such that  $\mathbf{f} = \mathbf{f}^{(1)} \otimes \mathbf{f}^{(2)}$ .

The following proposition is straightforward.

PROPOSITION 5. A problem is decomposable if and only if

1. there exist two subsets of indices  $D_1, D_2 \subseteq \{1, \dots, d\}$  and  $O_1, O_2 \subseteq \{1, \dots, k\}$  such that  $D_1 \cup D_2 = \{1, \dots, d\}$  and  $D_1 \cap D_2 = \emptyset$  and  $O_1 \cup O_2 = \{1, \dots, k\}$  and  $O_1 \cap O_2 = \emptyset$ .
2.  $m_i^\ell \neq 0$  if and only if  $(\ell, i) \in O_1 \times D_1$  or  $(\ell, i) \in O_2 \times D_2$ .

Remark 6.  $M$  is reducible if it has a full column of zeroes and it is decomposable if rows and columns can be reordered such that the matrix becomes block-diagonal. To be decomposable the number of blocks must be at least two. To each block corresponds a subproblem of the original MOP, with a smaller number of variables and objectives. As shown in Figure 1(b), e.g., the first subproblem contains the objective functions  $f^1, f^4$  and  $f^7$  and involves the decision variables  $x_3, x_6$  and  $x_8$ . To find the blocks of the decomposition we consider the matrix  $M$  as the incidence matrix of a directed bipartite graph with two lists of  $d$  and  $k$  nodes, where there exists a connection from the node  $i$  of the first list to the node  $\ell$  of the second list if  $m_i^\ell = 1$ . The matrix has two or more blocks if the graph has two or more connected components. The connected components of the graph are found by a breadth first search.

PROPOSITION 7. To be decomposable, problem (1) must have at least  $k + d - 2$  entries with value 0 in  $M$ .

Let us assume that we have a decomposable MOP as in Definition 4 and let  $d_{1,2} = \text{card}(D_{1,2})$  and  $k_{1,2} = \text{card}(O_{1,2})$ . After reordering the variables we can assume that  $D_1 = \{1, \dots, d_1\}$  and  $D_2 = \{d_1 + 1, \dots, d\}$  and  $O_1 = \{1, \dots, k_1\}$  and  $O_2 = \{k_1 + 1, \dots, k\}$ . We denote  $\mathbf{y} = (x_1, \dots, x_{k_1})$  and  $\mathbf{z} = (x_{k_1+1}, \dots, x_k)$ , such that  $\mathbf{x} = (\mathbf{y}, \mathbf{z}) \in [0, 1]^d$ . Let  $\pi_{1,2} : \mathbb{R}^k \rightarrow \mathbb{R}^{k_{1,2}}$  be canonical projections, i.e.,  $\pi_1(f^1, \dots, f^k) = (f^1, \dots, f^{k_1})$  and  $\pi_2(f^1, \dots, f^k) = (f^{k_1+1}, \dots, f^k)$ . We can then denote subproblems with  $\mathbf{f}^{(1)}$  and  $\mathbf{f}^{(2)}$  as follows:

$$\begin{aligned} \mathbf{f}^{(1)} : [0, 1]^{d_1} &\rightarrow \mathbb{R}^{k_1}, & \mathbf{f}^{(2)} : [0, 1]^{d_2} &\rightarrow \mathbb{R}^{k_2}, \\ \mathbf{y} &\mapsto \pi_1(\mathbf{f}(\mathbf{y}, \bar{\mathbf{z}})), & \mathbf{z} &\mapsto \pi_2(\mathbf{f}(\bar{\mathbf{y}}, \mathbf{z})), \end{aligned}$$

where  $\bar{\mathbf{y}} \in [0, 1]^{d_1}$  and  $\bar{\mathbf{z}} \in [0, 1]^{d_2}$  are vectors,  $f^1, \dots, f^{k_1}$  being independent of  $\mathbf{z}$  and  $f^{k_1+1}, \dots, f^k$  independent of  $\mathbf{y}$ .

There is a straightforward relation between the Pareto optimal sets and frontiers of the original MOP and those of the subproblems with  $\mathbf{f}^{(1)}$  and  $\mathbf{f}^{(2)}$ . The following result has been partially presented in [14].

THEOREM 8. Let us denote the sets of Pareto optima in the decision and objective spaces of the first and the second subproblems by  $P_{1,2}$  and  $F_{1,2}$ , respectively. Then,  $P = P_1 \times P_2 \subseteq [0, 1]^{d_1+d_2} = [0, 1]^d$  and  $F = F_1 \times F_2 \subseteq \mathbb{R}^{k_1+k_2} = \mathbb{R}^k$ , where  $\times$  stands for the Cartesian product.

*Proof.* Let us assume that  $\mathbf{x} = (\mathbf{y}, \mathbf{z}) \in P_1 \times P_2$  is not a Pareto optimum for  $\mathbf{f}$ . Then  $\mathbf{x}$  must be dominated by a different point  $\bar{\mathbf{x}} \in P$ , i.e.,  $f^\ell(\bar{\mathbf{x}}) \leq f^\ell(\mathbf{x})$  for all  $\ell$ ,  $1 \leq \ell \leq k$ , and there exists  $h$  such that  $f^h(\bar{\mathbf{x}}) < f^h(\mathbf{x})$ . The index  $h$  must be either in  $D_1$  or in  $D_2$ . Without loss of generality we assume that  $h \in D_1$ . Then  $\mathbf{y} \notin P_1$ , which contradicts with the assumption. So, we must have  $P_1 \times P_2 \subseteq P$ . Similarly, we assume that  $\mathbf{x} \in P$  but not  $\mathbf{x} = (\mathbf{y}, \mathbf{z}) \in P_1 \times P_2$ . Then, without loss of generality, we say that  $\mathbf{y} \notin P_1$ . Therefore, there exists  $\bar{\mathbf{y}} \in [0, 1]^{d_1}$  dominating  $\mathbf{y}$ , i.e., for some  $1 \leq h \leq k$ , we have  $f^h(\bar{\mathbf{y}}) < f^h(\mathbf{y})$ , and  $f^\ell(\bar{\mathbf{y}}) \leq f^\ell(\mathbf{y})$  for all the remaining  $\ell \in D_1$ . Thus  $(\bar{\mathbf{y}}, \mathbf{z})$  dominates  $(\mathbf{y}, \mathbf{z}) = \mathbf{x}$  for  $\mathbf{f}$ , therefore,  $\mathbf{x}$  is not a Pareto optimum for  $\mathbf{f}$ , which is a contradiction.  $\square$

**2.5. Sensitivity analysis with functional ANalysis Of VAriance.** So far, we have assumed that information about active and inactive variables is known *a priori*. However, no complete knowledge of the functional behavior of the objectives is usually available, unless they are given in an explicit closed form. On the contrary, we often have to deal with black box functions. Therefore, we need a systematic method for determining how *sensitive* an objective function is to variables.

Detecting which variables are important and which are not, is called *global sensitivity analysis*, and one of the most widely used methods is the *functional ANOVA decomposition* [47], where ANOVA stands for *ANalysis Of VAriance*.

Let us consider a single objective function in (1),  $f^\ell$  ( $1 \leq \ell \leq k$ ) and denote it by  $g : [0, 1]^d \rightarrow \mathbb{R}$ ,  $g(\mathbf{x}) = g(x_1, \dots, x_d) := f^\ell(x_1, \dots, x_d)$ . As suggested by Sobol' and Welch<sup>1</sup> we adopt sensitivity indices, main effects and interactions to assess the sensitivity of an objective function to individual variables or to combinations of variables. This analysis is built upon an ANOVA-type decomposition of  $g$ . It is a statistically based, hierarchically ordered, finite expansion of  $g$ . It is analogous to Taylor and Fourier expansions in the sense that the first terms of the expansion should be simple in form and offer a reasonable approximation of the original function under suitable conditions. Given that the terms in the expansion are computed by averaging, the considered form of the feasible set  $D = [0, 1]^d$  simplifies definitions and computations, without being restrictive.

We assume that  $g$  is square integrable and denote the overall mean of  $g(\mathbf{x})$  by  $g_0 := \int_{[0, 1]^d} g(\mathbf{x}) d\mathbf{x} = \int_{[0, 1]^d} g(x_1, \dots, x_d) dx_1 \dots dx_d$ . Sobol' shows that there exists a unique decomposition

$$(3) \quad g(x_1, \dots, x_d) = g_0 + \underbrace{g_1(x_1) + \dots + g_d(x_d)}_{\text{main effects}} + \underbrace{g_{1,2}(x_1, x_2) + \dots + g_{d-1,d}(x_{d-1}, x_d) + \dots}_{(2^{nd} \text{ order}) \text{ interactions}} + \dots + \underbrace{\sum_{1 \leq i_1 < i_2 < \dots < i_s \leq d} g_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) + \dots + g_{1, \dots, d}(x_1, \dots, x_d)}_{(s^{th} \text{ order}) \text{ interactions}}.$$

where the component functions  $g_{i_1, \dots, i_s}$  have a zero mean, i.e.,

$$(4) \quad \int_0^1 g_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) dx_i = 0, \quad \text{for any } i = i_1, \dots, i_s,$$

<sup>1</sup>See [47] for a summary. For the exposition of this topic we follow [48]. See also [11, 54, 57]



and the components are  $L^2$ -orthogonal, i.e., if  $(i_1, \dots, i_s) \neq (j_1, \dots, j_t)$ , then

$$(5) \quad \int_{[0,1]^d} g_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) g_{j_1, \dots, j_t}(x_{j_1}, \dots, x_{j_t}) dx_1 \dots dx_d = 0.$$

The components  $g_{i_1, \dots, i_s}$  are defined iteratively, starting from the *main effects*:

$$(6) \quad g_i(x_i) := \int_0^1 \dots \int_0^1 g(x_1, \dots, x_d) d\mathbf{x}_{-i} - g_0,$$

and proceeding with the *interaction effects* of variables  $x_i$  and  $x_j$ :

$$(7) \quad g_{i,j}(x_i, x_j) := \int_0^1 \dots \int_0^1 g(x_1, \dots, x_d) d\mathbf{x}_{-(i,j)} - g_0 - g_i(x_i) - g_j(x_j),$$

where  $d\mathbf{x}_{-i}$  denotes integration over all variables except  $x_i$  and  $d\mathbf{x}_{-(i,j)}$  denotes integration over all variables except  $x_i$  and  $x_j$ . The main effect of  $x_i$  represents the mean behavior of the function  $g$  when  $x_i$  is varied and all the other variables are averaged. Interactions as  $g_{i,j}$ , on the other hand, represent what happens to  $g$  when  $x_i$  and  $x_j$  are varied *accordingly* or *conversely*, and the other variables are averaged. The plots of main effects and interactions are usually very informative, at least in a qualitative sense.

We obtain quantitative indicators of sensitivity by computing variance based indices from the expansion (3). *Total variance*  $V$  and *partial variances*  $V_{i_1, \dots, i_s}$  are defined as<sup>2</sup>

$$(8) \quad V := \text{Var}(g) = \int_{[0,1]^d} g^2(\mathbf{x}) d\mathbf{x} - g_0^2, \quad V_{i_1, \dots, i_s} := \int_{[0,1]^s} g_{i_1, \dots, i_s}^2(x_{i_1}, \dots, x_{i_s}) dx_{i_1} \dots dx_{i_s}.$$

By squaring and integrating both sides of (3), because of orthogonality (5), we obtain  $V = \sum_{i=1}^d V_i + \sum_{1 \leq i < j \leq d} V_{i,j} + \dots + V_{1, \dots, d}$ , and can define *sensitivity indices* of the  $s^{\text{th}}$  order of  $g$  to the variables  $x_{i_1}, \dots, x_{i_s}$ , as

$$(9) \quad S_{i_1, \dots, i_s} := \frac{V_{i_1, \dots, i_s}}{V} \quad \left( S_i := \frac{V_i}{V}, \quad \text{sensitivity index of the first order} \right).$$

We define the *total sensitivity* to  $x_i$  as the sum of all sensitivity indices involving it as  $T_i = S_i + \sum_{j \neq i} S_{i,j} + \dots + S_{1,2, \dots, d}$ .

*Example 9.* To illustrate the ANOVA method, let us consider the function  $f(x_1, x_2) := \frac{x_1}{2} - \frac{x_1^2}{4} + \sin\left(\frac{3x_1}{7}\right) + \frac{1}{30}(x_2 - 5)^2 \sin\left(\frac{x_1}{30} - 4x_2\right)$ . As the function is in a closed form, it is possible to explicitly compute the main effects and interactions. We have  $f_0 \cong -0.91$ , and  $f_1(x_1)$  and  $f_2(x_2)$  are plotted in Figure 2(b) and (c). It seems that the ANOVA expansion truncated at the first order, i.e.,  $f_0 + f_1(x_1) + f_2(x_2)$  is a very good approximation. The remainder  $f_{1,2}(x, y) = f(x, y) - (f_0 + f_1(x_1) + f_2(x_2))$  is plotted in (d). The total variance, the partial variances, the sensitivity indices and the total sensitivity indices are reported in Table 1.

It is clear that when  $T_i = 0$ ,  $x_i$  does not affect  $g$  and the corresponding entry  $m_i^\ell$  in the incidence matrix is 0. Otherwise, we set  $m_i^\ell = 1$ . So, we have described a

<sup>2</sup>In this subsection,  $g^2$  means  $g$  to the power 2. Note that we do not have to subtract the squared mean of  $g_{i_1, \dots, i_s}$  because the mean is zero (4).

	$V$ and $V_i$	$S_i$	$T_i$
$f(x_1, x_2)$	3.57615	1.	
$f_1(x_1)$	2.95382	0.825977	0.826211
$f_2(x_2)$	0.621495	0.173789	0.174023
$f_{1,2}(x_1, x_2)$	0.00083535	0.000233589	

Table 1: Variances and indices of Example 9.

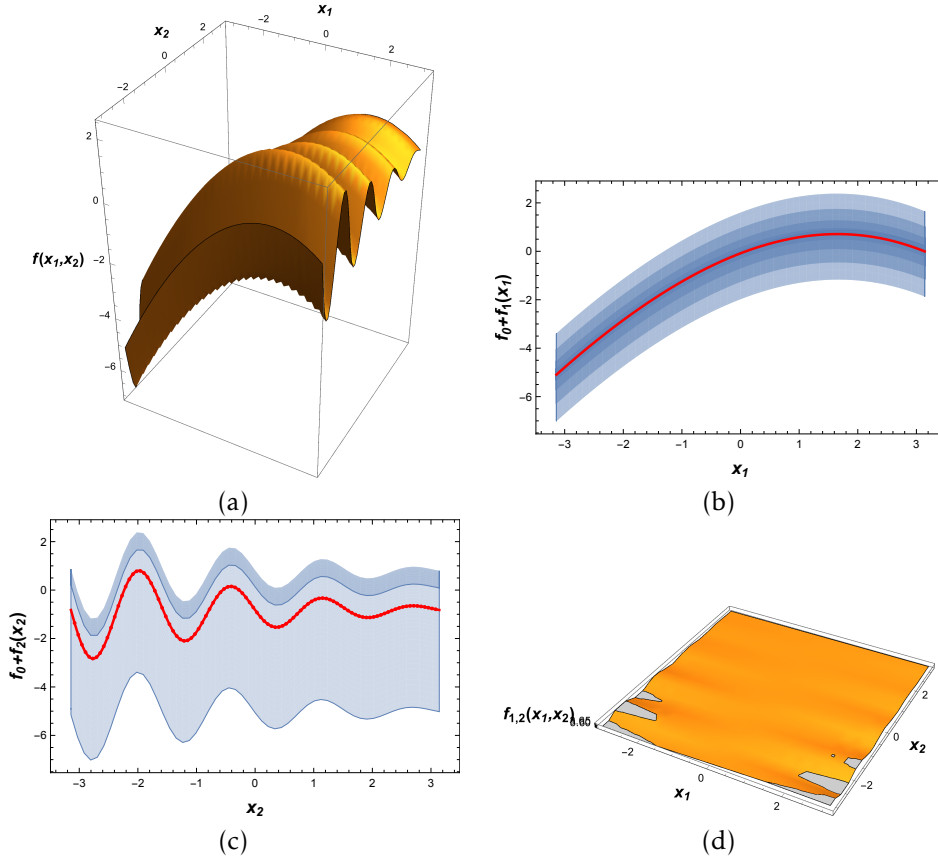


Figure 2: Illustration of Example 9. (a) Plot of the function  $f$ . (b-c) Plots of the main effects  $f_1(x_1) + f_0$  and  $f_2(x_2) + f_0$  (dark line – red in the online version) and values obtained as the other variable is varied. (d) Plot of the interaction of the second order  $f_{1,2}(x, y)$ , equal to  $f(x_1, x_2) - (f_0 + f_1(x_1) + f_2(x_2))$  (color available online).

method for computing an incidence matrix and, thus, for deciding whether a MOP is decomposable at least if it is possible to compute exactly the integrals in (6), (7) and (8).

The integrals can be too difficult to compute exactly or the incidence matrix may not have enough zeroes to be decomposable. Nevertheless, there exist effective and affordable numerical estimates for  $T_i$ s and often a large part of the entries of the incidence matrix can be small (in a suitable sense). We need therefore a method

for dealing with MOPs which are not exactly decomposable but only approximately decomposable.

**2.6. Sensitivity matrix.** In general, there is no evident reason why a particular entry of the incidence matrix  $M$  should be zero. Therefore, we cannot exclude a priori the possibility that  $M$  could be fully non-zero.

Nevertheless, according to the already mentioned *sparsity of effects principle*, decision variables do not usually have the same impact on the values of the objective functions, and the Pareto principle may hold. In other words, there may be a small number of decision variables that are responsible for the largest part of the effects observed and thus, it is reasonable to neglect the remaining variables. Neglecting the variables with a limited impact on the objective functions values leads to an auxiliary approximated problem which is decomposable.

Next we review a selection of numerical methods for estimating the total sensitivity indices  $T_i$  in (9). As already mentioned, computing  $T_i$ s is a matter of estimating integrals over the domain  $S$ . This can usually be approached in one of the following ways. Integrals can be estimated by the Monte Carlo method [21, 41, 54], which usually is reliable and universally applicable, although it requires large samples of function evaluations. Because this may not be affordable for computationally expensive functions, one can use *Latin Hypercube Sampling* which samples at specific points to allow a faster convergence of the integrals when compared to random samples [37].

Alternatively, a moderately large sample of function evaluations can be employed to construct metamodels for the objective functions. There are many methods for selecting the points to evaluate the objective functions. While most are of a *space-filling* type, there are methods that adapt to the function complexity (see [33, 34] for a summary). Then integrals of the metamodels can be estimated again by Monte Carlo or, if the metamodels have a suitable formulation, integrated analytically with a negligible computational cost. This is the case of the *Polynomial Chaos method* [57, 63] where an objective function is expanded in a series with a basis of orthogonal polynomials. This method has also been applied to robust multiobjective optimization [45]. Independently of the method chosen, the estimated total sensitivity indices  $T_i^\ell$  are collected in a sensitivity matrix  $SM := [T_i^\ell]$ ,  $\ell \in \{1, \dots, k\}$  and  $i \in \{1, \dots, d\}$ .

### 3. The ANOVA-MOP method.

**3.1. Approximated problem.** To develop the ANOVA-MOP method, we consider the sensitivity matrix  $SM$  and we fix a *threshold value*  $\delta > 0$ .

DEFINITION 10. Let  $\delta > 0$ . The function  $\bar{\mathbf{f}}_\delta : [0, 1]^d \rightarrow \mathbb{R}^k$ :

$$(10) \quad \bar{x}_i^\ell := \begin{cases} x_i & \text{if } T_i^\ell \geq \delta, \\ \frac{1}{2} & \text{if } T_i^\ell < \delta, \end{cases} \quad \bar{f}_\delta^\ell(x_1, \dots, x_d) := f^\ell(\bar{x}_1^\ell, \dots, \bar{x}_d^\ell),$$

is referred to as the approximated function of  $\mathbf{f}$ , while the corresponding MOP

$$(11) \quad \underset{\mathbf{x} \in S}{\text{minimize}} \{ \bar{f}_\delta^1(\mathbf{x}), \dots, \bar{f}_\delta^k(\mathbf{x}) \},$$

is the approximated problem (MOP) of (1).

In other words, for every objective function  $f^\ell$ ,  $1 \leq \ell \leq k$ , we consider as *active variables* only those  $x_i$  for which  $T_i^\ell \geq \delta$ . The remaining variables are fixed as  $\frac{1}{2}$ .

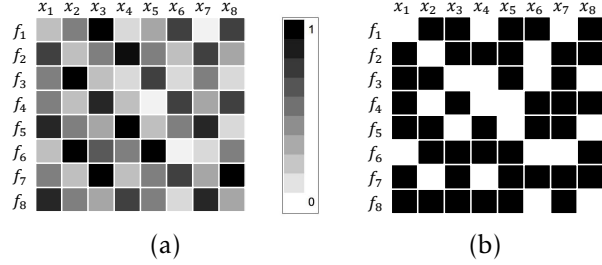


Figure 3: (a) An example of a total sensitivity matrix (SM). The bar shows the range of the total sensitivity indices. (b) A reduced incidence matrix  $M_\delta$  corresponding to a threshold value  $\delta = 0.25$ .

The choice of setting the neglected variables to the midpoint  $\frac{1}{2}$  of the range is non-restrictive and any point in the range  $[0, 1]$  can be chosen in (10). In what follows, we write  $\bar{\mathbf{f}}$  instead of  $\bar{\mathbf{f}}_\delta$  if there is no ambiguity. The incidence matrix  $M_\delta$  for the approximated problem is then  $M_\delta := [m_i^\ell]_{\ell,i}$  with  $m_i^\ell = 1$  if  $T_i^\ell \geq \delta$  and  $m_i^\ell = 0$  if  $T_i^\ell < \delta$ .

**3.2.  $\delta$ -reducible and  $\delta$ -decomposable problems.** Let a MOP be given with sensitivity matrix  $SM$ . It is clear that larger threshold values  $\delta$  imply approximated problems with a *sparser* incidence matrix  $M_\delta$ , because more variables are neglected in the approximated objective functions  $\bar{f}^\ell$ . If  $M_\delta$  is sparse enough, the corresponding approximated problem is reducible or decomposable.

**DEFINITION 11.** Let  $\mathbf{f}$ ,  $\delta$ ,  $\bar{\mathbf{f}}$ ,  $SM$  and  $M_\delta$  be as above. We say that problem (1) is  $\delta$ -reducible if the incidence matrix  $M_\delta$  is reducible and the problem is  $\delta$ -decomposable if  $M_\delta$  is decomposable.

It is obvious that the threshold  $\delta$  cannot be larger than a certain value

$$(12) \quad \omega = \min_{\ell=1,\dots,k} \max_{j=1,\dots,d} (T_j^\ell),$$

because with higher values some of the objective functions will have no active variables. In that case, some of the objective functions will be excluded and we do not want this to happen.

Figure 3(b) illustrates a reduced incidence matrix  $M_\delta$  obtained from the sensitivity matrix  $SM$  in Figure 3(a) with  $\delta = 0.25$ . (For example, for  $f^3$ ,  $T_j^3$ ,  $j = 3, 4, 6, 8$ , are smaller than the threshold.) Therefore, in the corresponding approximated problem, the decision variables  $x_j$ ,  $j = 3, 4, 6, 8$ , are treated as inactive for  $f^3$  and fixed to  $\frac{1}{2}$ . Once the reduced incidence matrix  $M_\delta$  is formed, the decomposability is assessed by using a decomposition method such as [6]. If  $M_\delta$  has too many active entries to be decomposable, the threshold value  $\delta$  is raised and  $M_\delta$  is recomputed. If necessary, the value of  $\delta$  is raised again until the matrix obtained is sparse enough to be decomposable allowing to define the approximated problem  $\bar{\mathbf{f}}_\delta$ .

It is not needed to test infinite values of  $\delta$ . Indeed, it is sufficient to sort in an increasing order the values of the sensitivity indices  $T_j^\ell$  for all  $\ell \in \{1, \dots, k\}$  and  $j \in \{1, \dots, d\}$  and test the values of  $\delta$  from the  $(n + k - 2)^{th}$  position to the largest among the values smaller than  $\omega$ . This is because of Proposition 7 and the reason explained above.

Thus, by raising sufficiently the value of the threshold we are likely to obtain a decomposable approximated problem, amenable to our analysis. Therefore, the question now is how much in accuracy we lose by gaining in decomposability.

**3.3.  $\epsilon$ -approximations of Pareto optimal sets and frontiers.** An important question in any optimization involving metamodels is how good the approximations of the optimal sets obtained by optimizing the approximated functions are. There is a (usually tacit) assumption that the operations of optimizing and approximating do commute, but this holds only if the functions behave well (in a suitable sense). Indeed, theoretical results of convergence can only be achieved by assuming some regularity of the functions, usually in the form of the existence of a global Lipschitz constant.

**DEFINITION 12.** We say that  $\bar{\mathbf{f}}$  is an  $\epsilon$ -approximation of  $\mathbf{f}$ , if and only if  $|\mathbf{f}(\mathbf{x}) - \bar{\mathbf{f}}(\mathbf{x})| < \epsilon$  for all  $\mathbf{x} \in [0, 1]^d$ .

In what follows, we assume that we have sampled the feasible set with  $N$  points  $\{\mathbf{x}^n, n = 1, \dots, N\}$ , and that the maximum radius of a ball not containing sample points is  $\rho := \max_{\mathbf{x} \in D, n=1, \dots, N} |\mathbf{x} - \mathbf{x}_n| > 0$ .

**THEOREM 13.** Let  $f : D \rightarrow \mathbb{R}$  be globally Lipschitz with constant  $L > 0$ . Let  $\bar{f}$  be an interpolation with the same or a smaller Lipschitz constant, defined on the dataset  $\{(\mathbf{x}^n, f(\mathbf{x}^n)), n = 1, \dots, N\}$ . Then  $\bar{f}$  is an  $\epsilon$ -approximation for  $f$  with  $\epsilon = 2L\rho$ .

*Proof.* For every  $\mathbf{x} \in D$  there exists  $\mathbf{x}_n$  such that  $|\mathbf{x} - \mathbf{x}_n| \leq \rho$ . Thus  $|\bar{f}(\mathbf{x}) - f(\mathbf{x})| = |\bar{f}(\mathbf{x}) - f(\mathbf{x}_n) + f(\mathbf{x}_n) - f(\mathbf{x})| \leq |\bar{f}(\mathbf{x}) - \bar{f}(\mathbf{x}_n)| + |f(\mathbf{x}) - f(\mathbf{x}_n)| \leq \bar{L}|\mathbf{x} - \mathbf{x}_n| + L|\mathbf{x} - \mathbf{x}_n| = (\bar{L} + L)|\mathbf{x} - \mathbf{x}_n| \leq 2L\rho$ , because for any interpolating method  $f(\mathbf{x}_n) = \bar{f}(\mathbf{x}_n)$ .<sup>3</sup>  $\square$

Under the given hypothesis, the Pareto optimal set of problem (1) can be approximated by the Pareto optimal set of the approximated problem. We give below more precise definitions and results.

**DEFINITION 14.** Following [25, 62], for a tolerance  $\epsilon > 0$ ,  $\mathbf{x}$   $\epsilon$ -dominates  $\mathbf{y}$  if and only if for all  $\ell \in \{1, \dots, k\}$ ,  $f^\ell(\mathbf{x}) \leq f^\ell(\mathbf{y}) + \epsilon$ , and there exists an index  $\bar{\ell} \in \{1, \dots, k\}$  such that  $f^{\bar{\ell}}(\mathbf{x}) < f^{\bar{\ell}}(\mathbf{y}) + \epsilon$ .

**DEFINITION 15.** Let  $P \subset [0, 1]^d$  be the Pareto optimal set for problem (1). For  $\epsilon > 0$ , the set  $\bar{P} \subset [0, 1]^d$  is called a Pareto optimal set approximation (or an  $\epsilon$ -Pareto optimal set), if and only if for all  $\mathbf{x} \in P$ , there exist  $\bar{\mathbf{x}} \in \bar{P}$  such that  $\bar{\mathbf{x}}$   $\epsilon$ -dominates  $\mathbf{x}$  [42]. The image  $\mathbf{f}(\bar{P})$  of the Pareto optimal set approximation is called a Pareto frontier approximation (or an  $\epsilon$ -Pareto frontier) for [42].

**THEOREM 16.** If  $\bar{\mathbf{f}}$  is an  $\epsilon$ -approximation for  $\mathbf{f}$ , then the Pareto optimal set for  $\bar{\mathbf{f}}$  is an  $\epsilon$ -Pareto optimal set for (1).

*Proof.* Let  $\mathbf{x} \in P$ , then  $\bar{\mathbf{f}}(\mathbf{x}) < \mathbf{f}(\mathbf{x}) + \epsilon$  because  $\bar{\mathbf{f}}$  is an  $\epsilon$ -approximation of  $\mathbf{f}$ . If  $\mathbf{x}$  is not a Pareto optimum for  $\bar{\mathbf{f}}$ , then there exists  $\bar{\mathbf{x}} \in \bar{P}$  such that  $\bar{\mathbf{x}}$  dominates  $\mathbf{x}$ . Thus  $\bar{\mathbf{x}}$   $\epsilon$ -dominates  $\mathbf{x}$  for  $\mathbf{f}$ .  $\square$

*Remark 17.* The converse also is true, i.e., the Pareto optimal set  $P$  is an  $\epsilon$ -approximation of the Pareto optimal set of the approximated MOP, because the above proof works also swapping the roles of  $\mathbf{f}$  and  $\bar{\mathbf{f}}$ .

<sup>3</sup>One should note that not all metamodels have controllable Lipschitz constants. However, one can define such metamodels, e.g., considering a Delaunay tessellation of the domain based on the sample points, and using linear interpolations on the simplices.

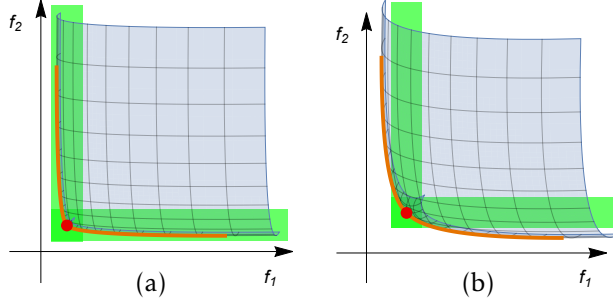


Figure 4: (a) Pareto frontier (the thick curve — orange in the online version) and the Pareto frontier approximation (the dark dot — red online) for a bi-objective optimization problem. The light strips (green online) represent the tolerance zone of width  $2\epsilon$ , where the Pareto frontier should be located for the  $\epsilon$ -approximation to be acceptable. (b) If the functional dependence is not enough weak to be neglected, the Pareto frontier violates the tolerance zone. Pareto optimal sets are traced by singular continuation [31, 32] (color available online).

A straightforward consequence of the previous theorem and remark is that if (1) is  $\delta$ -decomposable and the approximation is an  $\epsilon$ -approximation, then  $\bar{P}$  has a special form of the Cartesian product of the Pareto optimal sets of the subproblems, and  $P$  is  $\epsilon$ -dominated by the Cartesian product. In particular, the Pareto frontier  $\mathbf{f}(P)$  is contained in a set of a special form.

**COROLLARY 18.** *If problem (1) is  $\delta$ -decomposable and  $\bar{\mathbf{f}}_\delta$  is an  $\epsilon$ -approximation of  $\mathbf{f}$  with  $\epsilon, \delta > 0$ , then the Pareto frontier of problem (1) is contained in a tolerance zone:*

$$\bigcup_{\substack{y \in P^{(1)} \\ \bar{z} \in P^{(2)}}} \left\{ (u_1, u_2) : \left( |u_1 - \bar{\mathbf{f}}^{(1)}(y)| < 2\epsilon, u_2 > \bar{\mathbf{f}}^{(2)}(z) - 2\epsilon \right) \text{ or } \left( |u_2 - \bar{\mathbf{f}}^{(2)}(z)| < 2\epsilon, u_1 > \bar{\mathbf{f}}^{(1)}(y) - 2\epsilon \right) \right\}.$$

*Remark 19.* For bi-objective  $\delta$ -decomposable problems with continuous objective functions, the Pareto frontier approximation is a single solution because of the Weierstrass theorem. In non-degenerate cases, the Pareto frontier is a curve. In the assumption of Corollary 18, the tolerance zone is the union of two orthogonal half-infinite strips of width  $2\epsilon$ , as illustrated in Figure 4(a). In (b), the approximated problem is not  $\epsilon$ -accurate, and the Pareto frontier violates the tolerance zone.

**3.4.  $\delta$ -Decomposability versus  $\epsilon$ -accuracy through Lipschitz constants.** We know that for having a sufficiently large number of zero entries in the incidence matrix  $M_\delta$  to guarantee the  $\delta$ -decomposability of a problem we must use large values of  $\delta$ . Nevertheless, large values of  $\delta$  mean that when building the decomposable problem we have to neglect variables with not so small sensitivity indices. Therefore, a  $\delta$ -decomposition may not be an  $\epsilon$ -approximation. One may ask whether there is any quantitative relation between  $\delta$  and  $\epsilon$ ? In other words, is it possible to choose  $\delta$  small enough for getting an  $\epsilon$ -approximation?

If the sensitivity index  $T_i^\ell$  is smaller than a prescribed  $\delta$ , this does not automatically imply that the variations of  $f^\ell$  corresponding to the variable  $x_i$  are uniformly small. As  $T_i^\ell$  is obtained by averaging, it is possible that arbitrarily large differences  $\left| f_i^\ell(x_i^{(1)}) - f_i^\ell(x_i^{(2)}) \right|$  occur (it suffices that such differences occur in small regions of the domain). This means that a function must oscillate quite wildly, i.e., must have

a large, if existing, Lipschitz constant. Therefore, a relation between  $\delta$  and  $\epsilon$  can be obtained by assuming the existence of a Lipschitz constant for  $\mathbf{f}$ . We exhibit such a relation and prove it for a simple case.

LEMMA 20. Let  $g : [0, 1] \rightarrow \mathbb{R}$  be Lipschitz continuous with constant  $L > 0$ . If  $\text{Var}(g) < \delta$  with  $\delta > 0$ , then  $|g(b) - g(a)| \leq \sqrt[3]{12L\delta}$  for all  $a, b$ ,  $0 < a < b < 1$ . Equivalently,  $\bar{g}(\mathbf{x}) := g_0 = g(\frac{1}{2})$  is an  $\epsilon$ -approximation for  $g$  with  $\epsilon = \sqrt[3]{12L\delta}$ .

*Proof.* In the worst case,  $g$  is monotonically increasing (or decreasing) from  $a$  to  $b$ , and  $g$  is growing at the maximum speed allowed by the assumption, i.e.,  $g(x) = g(a) + L(x - a)$ . If so, with  $g_0$  as the mean value of  $g$ , and  $\text{Var}(g)$  the variance of  $g$ , we have  $\int_a^b |g(x) - g_0|^2 dx \leq \text{Var}(g) < \delta$ . Due to the mean value theorem,  $g_0 = \int_0^1 g(x) dx = g(x_0)$  for some  $x_0 \in [0, 1]$ . Thus,  $\int_a^b |g(x) - g_0|^2 dx = \int_a^b |g(x) - g(x_0)|^2 dx = L^2 \int_a^b |x - x_0|^2 dx = L^2 \left( \frac{(b-x_0)^3}{3} - \frac{(a-x_0)^3}{3} \right) \geq L^2 \frac{(b-a)^3}{12}$ , because  $\frac{(b-x_0)^3}{3} - \frac{(a-x_0)^3}{3}$  has its minimum at  $x_0 = \frac{b+a}{2}$ . This implies that  $|b - a| \leq \sqrt[3]{\frac{12\delta}{L^2}}$  and  $|g(b) - g(a)| \leq L(b - a) \leq \sqrt[3]{12L\delta}$ .  $\square$

It is lengthy but not difficult to prove the following.

THEOREM 21. Let  $f : [0, 1]^2 \rightarrow \mathbb{R}$  be Lipschitz continuous with constant  $L > 0$  and  $\delta$ -reducible. Then  $\bar{f}_\delta$  is an  $\epsilon$ -approximation for  $f$ , with  $\epsilon = \sqrt[6]{96\delta \text{Var}(f)L^4}$ .

**3.5. The relation between decomposability and tolerance thresholds.** Next we establish, for some low-dimensional cases, the probability for a problem to be  $\delta$ -decomposable given a value of  $\delta > 0$ . This gives a measure of the applicability of ANOVA-MOP to many real life applications. Let us consider a MOP and assume that the Pareto principle holds, i.e., globally the sensitivity indices follow a Pareto distribution. This implies that, for each objective function, 20% of the decision variables produce 80% of the variance. Recursively, this also implies that the fraction of the 20% most active decision variables among the 20% already selected is accountable for 80% of 80% of the variance, which allows to state that 4% of the decision variables are responsible for 64% of the variance, and so on. This can be summarized by saying that a fraction  $r$ ,  $0 \leq r \leq 1$ , of the most relevant decision variables produce a fraction equal to  $(\frac{4}{5})^{\log_5(r)}$  of the total variance, see Figure 5(a). As a result, we can establish a relation between the tolerance threshold  $\delta$  and the corresponding variance of the functions that we neglect when we build the approximation of problem (1).

Let us consider a bi-objective problem with  $d > 2$  decision variables. We assume for the sake of simplicity that the objective functions are *perfectly Paretian* in the sense that they follow exactly the behavior described in the previous paragraph, i.e., given any  $\delta > 0$ , both objectives have  $1 \leq \alpha \leq d$  non-negligible entries. According to the previous discussion, we assume that by passing to  $\bar{\mathbf{f}}_\delta$  we lose  $100 \times \left( 1 - (\frac{4}{5})^{\log_5(\frac{\alpha}{d})} \right)$ % of the total variance of each objective function. Let us denote by  $r = \frac{\alpha}{d}$  the corresponding fraction of active variables.

The probability that a bi-objective problem is  $\delta$ -decomposable, when for each objective function we have kept  $\alpha$  active variables is given by the probability that the selection of  $\alpha$  active variables of the second objective function has an empty intersection with  $\alpha$  active variables of the first objective function. We have  $\text{Prob} =$

$\frac{\binom{d-\alpha}{\alpha}}{\binom{d}{\alpha}} \cong \left(1 + \frac{\alpha}{d-2\alpha}\right)^{d-2\alpha+\frac{1}{2}} \left(1 - \frac{\alpha}{d}\right)^{d+\frac{1}{2}} = \left((1-r)\left(\frac{1-r}{1-2r}\right)^{1-2r}\right)^n \left(\frac{(1-r)^2}{1-2r}\right)^{\frac{1}{2d}}$ , by Stirling's formula. Given any  $r$ ,  $0 < r < \frac{1}{2}$ , this probability tends to zero as  $d \rightarrow \infty$ . However, the probability remains high for a moderately large number of  $d$ , see Figure 5(b). It is interesting then to compute the maximum ratio  $r = \frac{\alpha}{d}$ , i.e., the maximum variance preserved in the approximation  $\bar{f}_\delta$ , such that the probability of picking a decomposable MOP is higher than 50%. This computation is reported in Figure 5(c) and (d). E.g., we claim that for having a probability higher than 50% to randomly pick a decomposable bi-objective problem we have to neglect at least 80% of the variables (corresponding to 20% of the variance) for  $d = 10$  and 88% of the variables (25% of the variance) if  $d = 25$ . (A corresponding analysis can be carried on for 3 objectives, although less immediately.)

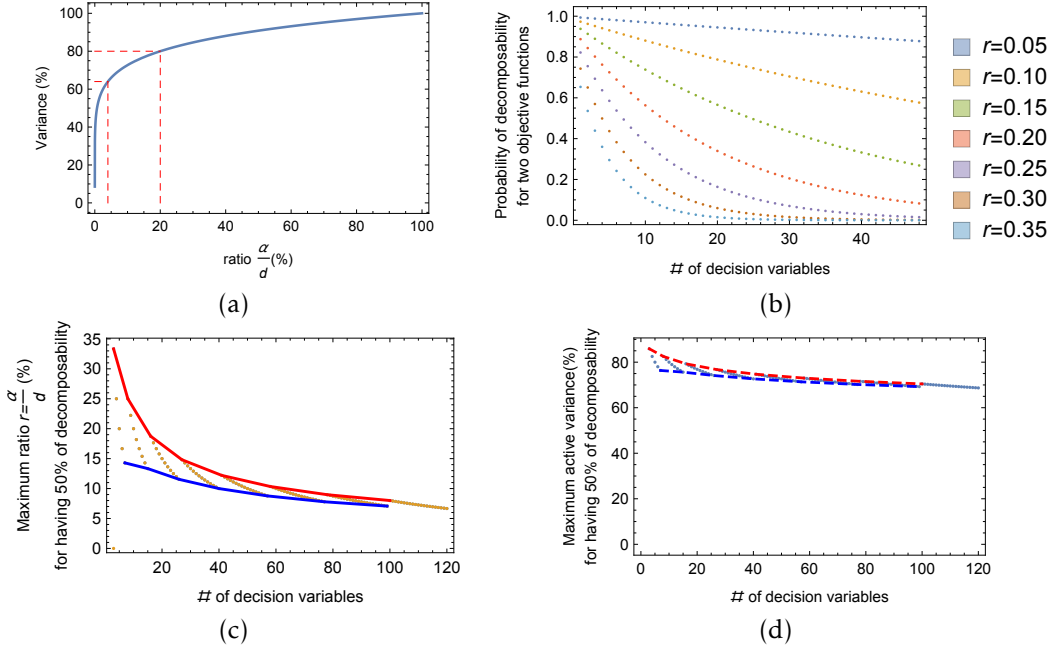


Figure 5: (a) Percentage of the variance of a “perfectly Paretian” objective function maintained by keeping an active fraction  $r = \frac{\alpha}{d}$  of the most representative variables. (b) Probability of  $\delta$ -decomposability for a bi-objective problem with a fraction  $r = \frac{\alpha}{d}$  of active variables as  $d$  varies from 3 to 50. (c) Maximum ratio  $r$  allowing for a 50% probability of having decomposability as  $d$  varies. (d) Maximum variance  $V(r)$  corresponding to perfectly Paretian objective functions allowing a 50% probability of having decomposability as  $d$  varies (color available online).

On the other hand, the probability that a bi-objective problem is  $\delta$ -reducible is the probability there exists at least one common neglected variable for both the objectives. This probability is equal to  $\sum_{h=0}^{d-\alpha-1} \binom{\alpha-h}{h} \binom{d-\alpha}{h} / \binom{d}{\alpha}$ , which automatically is 100% if  $\alpha < \frac{d}{2}$  and remains higher than 50% for  $\alpha \leq 91$ .

This result can be generalized to any number of objective functions. Let us denote by  $A_1$  the event that  $x_1$  is active in at least one of the objectives. Then  $A_1 \cap \dots \cap A_d$  represents the fact that each variable is active in at least one objective. We want to



compute the probability of  $\left(\bigcap_{i=1}^d A_i\right)^c$ , which by the generalized De Morgan law becomes:

$$(13) \quad \text{Prob}\left(\left(\bigcap_{i=1}^d A_i\right)^c\right) = 1 - \text{Prob}\left(\bigcup_{i=1}^d A_i^c\right) = 1 - \left[ \binom{d}{1} \text{Prob}(A_1^c) - \binom{d}{2} \text{Prob}(A_1^c \cap A_2^c) + \right. \\ \left. + \binom{d}{3} \text{Prob}(A_1^c \cap A_2^c \cap A_3^c) - \dots \pm \binom{d}{d-\alpha} \text{Prob}(A_1^c \cap \dots \cap A_{d-\alpha}^c) \right].$$

We observe that the probability that the first variable is neglected for all  $k$  objectives is given by  $\text{Prob}(A_1^c) = \left(\frac{(d-1)\dots(d-\alpha)}{d(d-1)\dots(d-\alpha+1)}\right)^k = \left(\frac{d-\alpha}{d}\right)^k$ , and by generalizing we obtain  $\text{Prob}(A_1^c \cap A_2^c) = \left(\frac{(d-2)\dots(d-\alpha-1)}{d(d-1)\dots(d-\alpha+1)}\right)^k = \left(\frac{(d-\alpha)(d-\alpha-1)}{d(d-1)}\right)^k$  and  $\text{Prob}(A_1^c \cap \dots \cap A_{d-\alpha}^c) = \left(\frac{(d-\alpha)\dots(d-\alpha-(d-\alpha)+1)}{d\dots(d-(d-\alpha)-1)}\right)^k = \left(\frac{(d-\alpha)!}{d\dots(\alpha+2)(\alpha+1)}\right)^k$ . Therefore, (13) becomes  $1 - \text{Prob}\left(\bigcup_{i=1}^d A_i^c\right) = 1 - \left[ \sum_{r=1}^{d-\alpha} (-1)^{r-1} \binom{d}{r} \prod_{h=0}^{r-1} \left(\frac{d-\alpha-h}{d-h}\right)^k \right]$ . This probability is automatically 1 for  $\alpha < \frac{d}{k}$ , but remains higher than 50% for large values of  $\alpha$ , e.g.,  $\alpha = 51$  when  $k = 7$ . The overall behavior is reported in Figure 6. This discussion demonstrates the applicability of

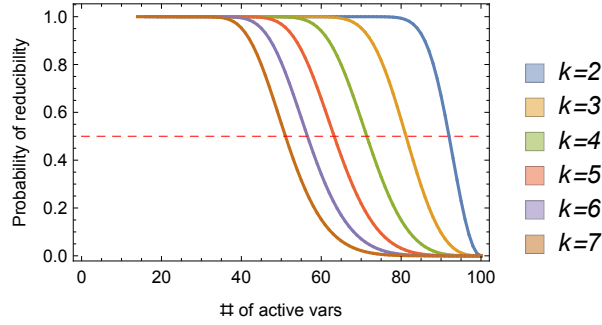


Figure 6: Probability of reducibility for a MOP with 100 decision variables and  $m$  active variables for each objective functions as the number of objectives  $k$  varies from 2 to 7. The lines are in an increasing order of  $k$  from right to left (color available online).

ANOVA-MOP to a wide range of applications, and explains why we do not consider any special method for cases when the column of common variables among the subproblems is of non-zero size.

Nevertheless, sometimes it can happen that for achieving a desired level of decomposability we may obtain an unacceptably inaccurate approximation of the objective functions. Then, one can consider a partially decomposable problem, but a specific strategy should be devised to deal with non-zero interaction columns. If it appears that the whole range of the variables in the interaction columns contains potentially interesting values, it is nontrivial to develop an efficient method for solving accurately and effectively such a partially decomposed problem. The possible advantages and computational costs should be carefully studied, and this is an interesting subject for future research.

**3.6. The ANOVA-MOP algorithm.** Now that we have presented the backgrounds that the ANOVA-MOP method is based on, we formalize it in Algorithm 1. ANOVA-MOP

converts the original problem into subproblems of fewer objectives and variables and results with approximated Pareto optima and information about the approximation quality. The method is modular and contains ingredients that can be defined with a large flexibility. To be as general as possible, we present the algorithm for black box functions. Naturally, the method can also be applied when dealing with other than black box functions.

ANOVA-MOP needs a training dataset to construct metamodels for the black box objective functions  $f^1, \dots, f^k$ . The dataset can be provided by a user or, e.g., a maximin Latin Hypercube Design [22] can be applied to get a dataset of  $n$  points. Alternatively, one can use a method that builds the metamodels and generates the dataset simultaneously as mentioned below.

1. Build a metamodel (e.g., Gaussian processes, radial basis functions, neural networks or splines) for each individual black box objective functions followed by estimating the total sensitivity indices. In our implementation, we have adopted the Bayesian Polynomial Chaos method (BPC) [57], but ANOVA-MOP is not limited to it.

BPC generates incrementally a training dataset and builds a metamodel iteratively to meet some goodness-of-fit criteria mentioned in [57]. In BPC, it is possible to choose a maximal interaction in the ANOVA decomposition, e.g., one can compute the full expansion or limit the analysis to the main effects. We opt for the complete expansion and use polynomials with a degree  $p = 4$ . The output of the method consists of the metamodels for the objective functions and a 95% confidence interval containing the estimated total sensitivity indices.

2. For decomposition, we use the implementation of Chen [6]. In this method, one can set the number of subproblems, the maximum number of common variables between subproblems and the minimum number of active variables within each subproblem.
3. ANOVA-MOP can be applied as a non-interactive or as an interactive method. The former case requires a multiobjective optimization method while the latter uses a single objective method, denoted by SOLVER. In Section 4 we adopt respectively RVEA [7] and DIRECT [23]. We use RVEA because it is a general purpose solver designed for problems with more than three objectives. However, these are only examples and any appropriate solver can be applied.

If ANOVA-MOP is used as an interactive method, it resembles the reference point method [64, 38], where an achievement scalarizing function (ASF) involving a reference point consisting of aspiration levels  $\bar{z} = (\bar{z}_1, \dots, \bar{z}_k)^T$ , specified by the DM is employed. As mentioned earlier, aspiration levels represent desirable objective function values. We consider the (ASF) [64] of the form  $\text{ASF}(\mathbf{x}, \bar{z}) = \max_{i \in O} (w_i (f^i(\mathbf{x}) - \bar{z}_i))$ , where  $w_i \geq 0$  for  $i = 1, \dots, k$  are non-negative fixed weights which set a direction where  $\bar{z}$  is projected onto the Pareto frontier. The weights can be used to normalize objectives. By solving the problem

$$(14) \quad \underset{\mathbf{x} \in S}{\text{minimize}} \quad \text{ASF}(\mathbf{x}, \bar{z})$$

with an appropriate single objective solver, we get a (weak) Pareto optimum best reflecting the preferences. (To avoid weak Pareto optimality, one can add an augmentation term to ASF, see e.g., [38].)

As discussed in [9], to accomplish cognitively demanding tasks, one requires a

sufficient ability to hold information as it is processed. For a high number of objective functions, the DM may find that providing corresponding reference points iteratively may impose a heavy cognitive load. This is the motivation for applying ANOVA-MOP in an interactive way, where the DM gives aspiration levels to subsets of the original objectives, one subproblem at a time, leading to reducing the cognitive load. In what follows, we discuss how subproblems of ANOVA-MOP are utilized to interact with the DM.

Once the  $m$  subproblems have been formed, they are considered one by one. The DM is informed which objectives are included in each subproblem. For the  $a^{\text{th}}$  subproblem, the DM is asked to specify corresponding aspiration levels and an ASF is formulated with objective functions of that subproblem. The corresponding reference point is denoted by  $\bar{\mathbf{z}}^{a*}$ . The  $a^{\text{th}}$  subproblem (which is derived from (14))

$$(15) \quad \underset{\mathbf{x}^a \in S^a}{\text{minimize}} \quad \text{ASF}(\mathbf{x}^a, \bar{\mathbf{z}}^{a*}) \quad S^a := [0, 1]^{d_a}$$

is solved and the solution shown to the DM. The interaction with the DM is continued until the most preferred solution for the objectives in the  $a^{\text{th}}$  subproblem is found. This is repeated for all subproblems. The following theorem shows that the solutions obtained with the subproblems form a preferred solution to problem (1).

**THEOREM 22.** *Suppose the incidence matrix  $M$  can be decomposed into  $m$  subproblems. Let  $\bar{\mathbf{z}}^{a*}$ , for  $a = 1, \dots, m$ , be a reference point given by the DM for the  $a^{\text{th}}$  subproblem (15) and denote by  $\bar{\mathbf{x}}^{a*}$  an optimal solution. Then  $\mathbf{x}^* = (\mathbf{x}^{1*}, \dots, \mathbf{x}^{m*}) \in \underset{\mathbf{x} \in S}{\text{argmin}} \text{ASF}(\mathbf{x}, \bar{\mathbf{z}}^*)$ , where  $\bar{\mathbf{z}}^* = (\bar{\mathbf{z}}^{1*}, \dots, \bar{\mathbf{z}}^{m*})$ .*

*Proof.* We show that

$$(16) \quad \text{ASF}(\mathbf{x}^*, \bar{\mathbf{z}}^*) \leq \text{ASF}(\mathbf{x}, \bar{\mathbf{z}}^*), \quad \text{for all } \mathbf{x} \in S.$$

Suppose  $\text{ASF}(\mathbf{x}^*, \bar{\mathbf{z}}^*) = (w_h(f^h(\mathbf{x}^*) - \bar{z}_h^*))$ . Then, there exists an index  $e$  such that  $h \in O_e$ . The optimal solution of the  $e^{\text{th}}$  subproblem is  $\mathbf{x}^{e*}$ . This means that  $(w_h(f^h(\mathbf{x}^*) - \bar{z}_h^*)) = \max_{j \in O_e} (w_j(f^j(\mathbf{x}^{e*}, \underline{\mathbf{x}}^e) - \bar{z}_j^{e*}))$ . For all  $\mathbf{x}^e \in S^e$ , we have  $\max_{j \in O_e} (w_j(f^j(\mathbf{x}^{e*}, \underline{\mathbf{x}}^e) - \bar{z}_j^{e*})) \leq \max_{j \in O_e} (w_j(f^j(\mathbf{x}^e, \underline{\mathbf{x}}^e) - \bar{z}_j^{e*}))$ .

On the other hand,  $\max_{j \in O_e} (w_j(f^j(\mathbf{x}^e, \underline{\mathbf{x}}^e) - \bar{z}_j^{e*})) \leq \max_{a=1, \dots, n} (\max_{b \in O_a} (w_b(f^b(\mathbf{x}^a, \underline{\mathbf{x}}^a) - \bar{z}_b^a)))$  for all  $\mathbf{x}^e \in S^e$ . Since,  $\max_{a=1, \dots, n} (\max_{b \in O_a} (w_b(f^b(\mathbf{x}^a, \underline{\mathbf{x}}^a) - \bar{z}_b^a))) = \text{ASF}(\mathbf{x}, \bar{\mathbf{z}}^*)$ , inequality (16) results.  $\square$

One should remember that if the functions involved are computationally expensive, the possibility of using metamodels reduces the waiting time set on the DM for obtaining solutions corresponding to her/his preferences.

**4. Benchmarks.** In what follows, we demonstrate how ANOVA-MOP can be applied. <sup>4</sup> By writing  $k \times d = (k_1 \times d_1) \otimes \dots \otimes (k_n \times d_n)$  we mean decomposing a problem with  $k$  objectives and  $d$  variables to  $n$  subproblems with  $k_i$  objectives and  $d_i$  variables. We consider two problems defined by algebraic functions. This allows us to know in advance which are the subproblems composing the full problem and perform exhaustive searches in the feasible sets. This naturally cannot be done with expensive black box functions but enables a transparent demonstration of the capabilities of the method.

<sup>4</sup>The code for ANOVA-MOP is publicly available at <http://www.mit.jyu.fi/optgroup/desdeo.html>

---

**Algorithm 1** ANOVA-MOP

---

**Input:** Let  $\mathbf{f} : [0, 1]^d \rightarrow \mathbb{R}^k$  be a black box vector function in (1). The user has an option of selecting a threshold value  $\delta$ .

- 1: Find a training dataset, apply a metamodeling technique to build the metamodels  $\tilde{\mathbf{f}}$  for  $\mathbf{f}$  and estimate the total sensitivity indices. Alternatively, employ BPC [57], which finds the dataset, builds the metamodels and estimates the total sensitivity indices.

- 2: Perform the ANOVA analysis on  $\tilde{\mathbf{f}}$  and build the  $k \times d$  sensitivity matrix  $SM$ .
- 3: Define a sorted list  $E = \{e_1, \dots, e_{d \cdot k}\}$  of all the entries of  $SM$  in an increasing order.
- 4: Determine  $\omega$  according to (12) and find the maximum  $r$  such that  $e_r \leq \omega$ .

▷ Here  $\omega$  is also a valid value for  $\delta$  because the entries equal to  $\delta$  remain active.

- 5: Extract the list  $\Delta$  by picking the elements of  $E$  from the  $(d + 1)$ -th to  $\omega$ , i.e.,  $\Delta := \{e_{d+1}, \dots, e_r\}$ .  
▷ We need at least  $d$  entries equal to zero for having reducibility and  $d+k-2$  for decomposability.

6: **if**  $\delta$  has been chosen by the user **then**

- 7: Build the reduced incidence matrix  $M_\delta$ .

8: **else** ▷ Pick the smallest value of  $\delta$  for which  $M_\delta$  is reducible or decomposable, if existing.

9: **repeat**

- 10: Pick as tolerance value  $\delta$  the smallest element in  $\Delta$ .

- 11: Build the reduced incidence matrix  $M_\delta$ .

- 12: Remove  $\delta$  from  $\Delta$ .

- 13: **until**  $M_\delta$  is reducible or decomposable or  $\Delta$  is empty.

14: **end if**

15: **if**  $M_\delta$  is decomposable **then**

- 16: Decompose problem with  $\tilde{\mathbf{f}}_\delta$  in approximated subproblems with  $\tilde{\mathbf{f}}^{(1)}, \dots, \tilde{\mathbf{f}}^{(m)}$  as described in Remark 6.

- 17: Solve the subproblems separately by calling SOLVER, which returns the sets of Pareto optima  $\bar{P}^{(1)}, \dots, \bar{P}^{(m)}$  and corresponding Pareto frontiers  $\bar{F}^{(1)}, \dots, \bar{F}^{(m)}$

- 18: Build the Pareto optimal set approximation for (1)  $\bar{P} := \bar{P}^{(1)} \times \dots \times \bar{P}^{(m)}$ :

$$\bar{P} = \left\{ \left( x_1^{(1)}, \dots, x_{d_1}^{(1)}, x_1^{(2)}, \dots, x_{d_{m-1}}^{(m-1)}, x_1^{(m)}, \dots, x_{d_m}^{(m)} \right) \mid \mathbf{x}^{(1)} \in \bar{P}^{(1)}, \dots, \mathbf{x}^{(m)} \in \bar{P}^{(m)} \right\}$$

and, analogously, the corresponding Pareto frontier approximation  $\bar{F} := \bar{F}^{(1)} \times \dots \times \bar{F}^{(m)}$ .

19: **else if**  $M_\delta$  is reducible **then**

- 20: Solve the approximated reduced problem  $\tilde{\mathbf{f}}_\delta$  by calling SOLVER, which returns the sets of solutions  $\bar{P}$  and  $\bar{F}$ .

21: **else**

- 22: Set  $\bar{P} = \bar{F} = \emptyset$ .

▷ The method is not applicable

23: **end if**

- 24: Select randomly a validation sample  $V := \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(v)}\} \subseteq \bar{P}$  and estimate the approximation quality  $\epsilon$  of the solutions found as follows

$$(17) \quad \epsilon \simeq \max_{\mathbf{x} \in V} \|\mathbf{f}(\mathbf{x}) - \tilde{\mathbf{f}}_\delta(\mathbf{x})\|_\infty = \max_{v=1, \dots, v} \max_{\ell=1, \dots, k} \left| f^\ell(\mathbf{x}^{(v)}) - \tilde{f}_\delta^\ell(\mathbf{x}^{(v)}) \right|.$$

- 25: **return**  $(\bar{P}, \bar{F}, V, \epsilon)$ .

To demonstrate the capabilities of the ANOVA-MOP method, we repeat the following for both the problems:

1. We perform an almost exhaustive search in the feasible set of the original problem by using a full factorial design of experiment if the dimension of the decision space is moderate or by using a large random sample if the dimension is large. We then evaluate all these points with the original functions, and extract the nondominated subset which is considered as the exact Pareto frontier of the problem denoted by  $F_{exh}$ .
2. We apply the multiobjective optimization method RVEA on the original problem. This serves as a comparison for the ANOVA-MOP method.
3. We start applying the ANOVA-MOP algorithm by defining the metamodels and following Algorithm 1 to obtain subproblems. For the decomposition, we apply the Chen method by setting the number of subproblems to 2 for problem (18) and to 4 for the high dimensional problem in Section 4.2.
4. We run RVEA separately on each subproblem to obtain  $(P^{(i)}, F^{(i)})$ ,  $i = 1, \dots, m$ , where  $m$  is the number of subproblems.
5. We build the Cartesian products of the Pareto optimal sets of the subproblems  $\bar{P} = P^{(1)} \times \dots \times P^{(m)}$  and the Pareto frontiers  $\bar{F} = F^{(1)} \times \dots \times F^{(m)}$ . They are the approximated Pareto optimal set and frontier obtained, respectively.
6. We estimate the width of the  $2\epsilon$ -strips considered in Corollary 18 and Remark 19 by computing a set-wise distance between the “exact” Pareto frontier  $F_{exh}$  defined in step 1 and the candidate Pareto frontier  $F_{cand}$  obtained in steps 2 and 5 as  $\text{loss}(F_{exh}, F_{cand}) := \max_{p \in F_{exh}} \min_{q \in F_{cand}} \max_{\ell=1, \dots, k} \max\{q^\ell - p^\ell, 0\}$ . In particular, we estimate  $\epsilon \simeq \text{loss}(F_{exh}, \bar{F})$ . Since  $\bar{F}$  is the Cartesian product of the Pareto frontiers of the subproblems,  $\bar{F} = F^{(1)} \times \dots \times F^{(m)}$ , the global loss equals the maximum of the losses computed for the subproblems separately, i.e.,  $\text{loss}(F_{exh}, F^{(1)} \times \dots \times F^{(m)}) = \max_{i=1, \dots, m} \text{loss}(\pi^{(i)}(F_{exh}), F^{(i)})$ , where  $\pi^{(i)} : \mathbb{R}^k \rightarrow \mathbb{R}^{k_i}$  is the projection from the objective space to the subspace of the objective functions of the  $i$ -th subproblem.
7. We compare this value with the value obtained in the ANOVA-MOP method, i.e., we extract a random subset  $V \subseteq \bar{P}$  of the points in the approximated Pareto frontier and evaluate the original functions on these points. Then we consider as a possible estimate for  $\epsilon$  the infinity norm of the difference between the approximated value and the exact value on these points, as defined in (17).
8. We compare the results of RVEA on the un-decomposed problem and on the decomposed problem in terms of the loss computed at the previous step and the number of function evaluations used in the two cases.

**4.1.  $\delta$ -decomposable problem.** We consider the following five-objective parametrized optimization problem  $5 \times 5 \simeq (3 \times 3) \otimes (2 \times 2)$  with  $\lambda = 1.3$ :

$$(18) \quad \underset{\mathbf{x} \in [-\lambda, \lambda]^5}{\text{minimize}} \{f^1(\mathbf{x}), \dots, f^5(\mathbf{x})\} \quad \text{where} \quad \mathbf{x} = (x_1, \dots, x_5) \quad \text{and}$$

$$\begin{aligned}
f^1(\mathbf{x}) &= g_1(\mathbf{x}) + \gamma g_4(\mathbf{x}), & g_1(\mathbf{x}) &= \left| (x_1, x_2, x_3)^T - P_1^T \right|^2, & P_1 &= (1, 1, 1)^T, \\
f^2(\mathbf{x}) &= g_2(\mathbf{x}) + \gamma g_5(\mathbf{x}), & g_2(\mathbf{x}) &= \left| (x_1, x_2, x_3)^T - P_2^T \right|^2, & P_2 &= (1, -1, -1)^T, \\
f^3(\mathbf{x}) &= g_3(\mathbf{x}) + \gamma(g_4(\mathbf{x}) + g_5(\mathbf{x})), & g_3(\mathbf{x}) &= \left| (x_1, x_2, x_3)^T - P_3^T \right|^2, & P_3 &= (1, 1, -1)^T, \\
f^4(\mathbf{x}) &= g_4(\mathbf{x}) + \gamma g_1(\mathbf{x}), & g_4(\mathbf{x}) &= \left| (x_4, x_5)^T - P_4^T \right|^2, & P_4 &= (1, -1)^T, \\
f^5(\mathbf{x}) &= g_5(\mathbf{x}) + \gamma(g_1(\mathbf{x}) + g_2(\mathbf{x})), & g_5(\mathbf{x}) &= \left| (x_4, x_5)^T - P_5^T \right|^2, & P_5 &= (-1, 1)^T,
\end{aligned}$$

with  $\gamma = 7.000 \times 10^{-3}$ . The BPC method managed to build metamodels by generating iteratively a sequence of points, stopping with 47 points, when a suitable goodness-of-fit criterion was fulfilled [57]. BPC also estimated the total sensitivity matrix  $SM$  given below based on which we form the incidence matrix  $M$  as

$$SM = \begin{bmatrix} 0.333 & 0.333 & 0.333 & 0.001 & 0.001 \\ 0.333 & 0.333 & 0.333 & 0.001 & 0.001 \\ 0.333 & 0.333 & 0.333 & 0.001 & 0.001 \\ 0.001 & 0.001 & 0.001 & 0.499 & 0.499 \\ 0.001 & 0.001 & 0.001 & 0.499 & 0.499 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

By taking as  $\delta$  any value  $0.001 < \delta \leq 0.333$ , we obtain the reduced incidence matrix  $M_\delta$  and a corresponding auxiliary decomposable problem  $\bar{\mathbf{f}}_\delta$  as follows:

$$(19) \quad M_\delta = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad \bar{\mathbf{f}}_\delta(\mathbf{x}) := \bar{\mathbf{f}}^{(1)} \otimes \bar{\mathbf{f}}^{(2)}(\mathbf{x}) = \begin{pmatrix} f^1(x_1, x_2, x_3, 0, 0) \\ f^2(x_1, x_2, x_3, 0, 0) \\ f^3(x_1, x_2, x_3, 0, 0) \\ f^4(0, 0, 0, x_4, x_5) \\ f^5(0, 0, 0, x_4, x_5) \end{pmatrix},$$

where  $\mathbf{f}^{(1)}$  and  $\mathbf{f}^{(2)}$  are the following three- and bi-objective subproblems:

$$(20) \quad \underset{(x_1, x_2, x_3) \in [-\lambda, \lambda]^3}{\text{minimize}} \quad \bar{\mathbf{f}}^{(1)}(x_1, x_2, x_3) := \begin{pmatrix} f^1(x_1, x_2, x_3, 0, 0) \\ f^2(x_1, x_2, x_3, 0, 0) \\ f^3(x_1, x_2, x_3, 0, 0) \end{pmatrix},$$

$$(21) \quad \underset{(x_4, x_5) \in [-\lambda, \lambda]^2}{\text{minimize}} \quad \bar{\mathbf{f}}^{(2)}(x_4, x_5) := \begin{pmatrix} f^4(0, 0, 0, x_4, x_5) \\ f^5(0, 0, 0, x_4, x_5) \end{pmatrix}.$$

Figure 7(a) and (b) depict the Pareto frontiers of these subproblems, respectively. According to Theorem 8, the set of Pareto optima for  $\bar{\mathbf{f}}_\delta = \bar{\mathbf{f}}^{(1)} \otimes \bar{\mathbf{f}}^{(2)}$  is the Cartesian product of the sets of Pareto optima for subproblems (20) and (21).

In other words, all the solutions of problem (19) are obtained by joining a solution  $(x_1, x_2, x_3)$  of subproblem (20) and a solution  $(x_4, x_5)$  of subproblem (21). The Pareto optimal set of (19) serves as an approximation for the Pareto optimal set of problem (18).

**4.1.1. Approximation quality.** Before applying ANOVA-MOP, we estimate the quality of the decomposed problem, represented by  $\epsilon$ , and test whether the claim of Corollary 18 is compatible with such an estimated value. Given the moderate complexity of the problem we can execute an almost exhaustive exploration of the domain, i.e., we set  $A_{exh}$  to be a full factorial in 5 dimensions with 20 levels for each

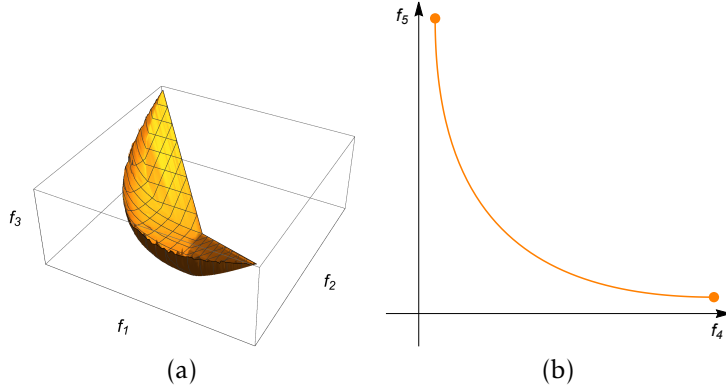


Figure 7: (a) Pareto frontier of subproblem (20). (b) Pareto frontier of subproblem (21) (color available online).

variable. Then  $\#\{A_{exh}\} = 20^5$  and we evaluate the objective functions at the points in  $A_{exh}$ , denoting  $B_{exh} = \mathbf{f}(A_{exh})$ . We extract the “exact” Pareto optimal set  $(P_{exh}, F_{exh}) := ND(A_{exh}, B_{exh})$ , where with  $ND(A, B)$  we denote the operation of removing the dominated solutions from  $(A, B)$ , where  $A \subset [-\lambda, \lambda]^d$  is any sample and  $B = f(A)$ .

**4.1.2. ANOVA-MOP test.** Next, we analyze the behavior of RVEA [7], when applied to the original problem versus the application of RVEA coupled with ANOVA-MOP. We select analogous settings compared to the dimensions of the problems and compare the results in terms of accuracy and computational costs. In particular, we set  $3^{dim}$  as the population size and 10 as the number of generations. We estimate the method performance by computing the previous set-wise distance between  $F_{exh}$  and the optimal set obtained. We compute also the number of function evaluations required by the three methods. When we use ANOVA-MOP, we must consider the computational overhead, i.e., the function evaluations required to build the metamodels used for the decomposition, here 47 function evaluations, and the evaluations performed to validate the optimal set found on the metamodel  $\mathbf{f}_\delta$ , here depending on the optimization results.

1. We run RVEA on problem (18). We call  $P_1$  the final Pareto optimal set obtained and  $F_1$  the corresponding Pareto frontier (even though evolutionary algorithms cannot guarantee optimality).
2. We run RVEA separately on the two subproblems and generate the Pareto optimal set for the whole problem as the Cartesian product of the Pareto optimal sets found in the separate subproblems and validate the function values. We call  $P_2$  and  $F_2$  these Pareto optimal set and frontiers, respectively.

We summarize the results in Table 2, where

$$(22) \quad \text{loss}_1(F_{exh}, F_i^{(1)}) := \max_{p \in F_{exh}} \min_{q \in F_i^{(1)}} \max_{\ell=1,2,3} \max(q^\ell - p^\ell, 0),$$

$$(23) \quad \text{loss}_2(F_{exh}, F_i^{(2)}) := \max_{p \in F_{exh}} \min_{q \in F_i^{(2)}} \max_{\ell=4,5} \max(q^{\ell-3} - p^\ell, 0),$$

$$(24) \quad \text{loss}_{tot}(F_{exh}, F_i) := \max_{p \in F_{exh}} \min_{q \in F_i} \max_{\ell=1,\dots,5} \max(q^\ell - p^\ell, 0),$$

Dataset	$\text{loss}_1(F_{exh}, F_i^{(1)})$	$\text{loss}_2(F_{exh}, F_i^{(2)})$	$\text{loss}_{tot}(F_{exh}, F_i)$	$N_{evals}$
$(P_1, F_1)$	0.3509	0.11132	<b>0.97362</b>	$3^5 \times 10 = \mathbf{2430}$
$(P_2, F_2)$	0.537	0.7275	<b>0.7275</b>	$47 + (3^3 + 3^2) \times 10 + 247 = \mathbf{654}$

Table 2: Summary of the results of the comparison for the first example.

and we observe that only when  $F = F^{(1)} \times F^{(2)}$  then

$$\text{loss}_{tot}(F_{exh}, F_i) = \max\left(\text{loss}_1(F_{exh}, F_i^{(1)}), \text{loss}_2(F_{exh}, F_i^{(2)})\right),$$

which is only the case for the Pareto optimal sets obtained by ANOVA-MOP. We notice that the accuracy of the Pareto optimal sets is comparable in the two cases while the number of function evaluations is sensibly smaller when the decomposition is used, while keeping the settings of the algorithm scaled to the problem dimension. Therefore, we can claim that the main scope of the method, i.e., mitigating the effects of the curse of dimensionality, is accomplished.

In the case of expensive functions, we could not have explored exhaustively the problem domain and could not estimate the width of the tolerance zone as we did. In that case, we could estimate  $\epsilon$  by evaluating the original functions on a random subset  $V$  of the Pareto optimal set approximation and by computing the maximum difference between the values of  $\mathbf{f}$  and of  $\mathbf{f}_\delta$  on  $V$ . In our case, we can compute this value on the entire set  $\bar{P}$  because it is still of a moderate size and our functions are inexpensive and get  $\epsilon \simeq \max_{\mathbf{x} \in \bar{P}} \|\mathbf{f}(\mathbf{x}) - \bar{\mathbf{f}}_\delta(\mathbf{x})\|_\infty = 0.054045$ .

**4.2. High-dimensional  $\delta$ -decomposable problem.** Next we consider the following parametrized  $12 \times 10 \simeq (3 \times 3) \otimes (3 \times 3) \otimes (2 \times 2) \otimes (4 \times 2)$  problem involving twelve variables with  $S = [-\lambda, \lambda]^{12}$ , where  $\lambda = 1.3$ . We minimize ten objectives of



component functions defined as follows with  $\gamma = 7.000 \times 10^{-3}$ :

$$\begin{aligned}
f^1(\mathbf{x}) &= g_1(\mathbf{x}) + \gamma g_4(\mathbf{x}), & g_1(\mathbf{x}) &= |(x_1, x_2, x_3)^T - P_1^T|^2, & P_1 &= (1, 1, 1)^T, \\
f^2(\mathbf{x}) &= g_2(\mathbf{x}) + \gamma g_5(\mathbf{x}), & g_2(\mathbf{x}) &= |(x_1, x_2, x_3)^T - P_2^T|^2, & P_2 &= (1, -1, -1)^T, \\
f^3(\mathbf{x}) &= g_3(\mathbf{x}) + \gamma(g_4(\mathbf{x}) + g_6(\mathbf{x})), & g_3(\mathbf{x}) &= |(x_1, x_2, x_3)^T - P_3^T|^2, & P_3 &= (1, 1, -1)^T, \\
f^4(\mathbf{x}) &= g_4(\mathbf{x}) + \gamma(g_1(\mathbf{x}) + g_7(\mathbf{x})), & g_4(\mathbf{x}) &= |(x_4, x_5, x_6)^T - P_4^T|^2, & P_4 &= (-1, -1, -1)^T, \\
f^5(\mathbf{x}) &= g_5(\mathbf{x}) + \gamma(g_2(\mathbf{x}) + g_8(\mathbf{x})), & g_5(\mathbf{x}) &= |(x_4, x_5, x_6)^T - P_5^T|^2, & P_5 &= (-1, 1, -1)^T, \\
f^6(\mathbf{x}) &= g_6(\mathbf{x}) + \gamma(g_3(\mathbf{x}) + g_9(\mathbf{x})), & g_6(\mathbf{x}) &= |(x_4, x_5, x_6)^T - P_6^T|^2, & P_6 &= (-1, -1, 1)^T, \\
f^7(\mathbf{x}) &= g_7(\mathbf{x}) + \gamma(g_2(\mathbf{x}) + g_5(\mathbf{x})), & g_7(\mathbf{x}) &= |(x_7, x_8)^T - P_7^T|^2, & P_7 &= (1, -1)^T, \\
f^8(\mathbf{x}) &= g_8(\mathbf{x}) + \gamma(g_1(\mathbf{x}) + g_4(\mathbf{x}) + g_9(\mathbf{x})), & g_8(\mathbf{x}) &= |(x_7, x_8)^T - P_8^T|^2, & P_8 &= (-1, 1)^T, \\
f^9(\mathbf{x}) &= g_9(\mathbf{x}) + \gamma(g_3(\mathbf{x}) + g_6(\mathbf{x}) + g_8(\mathbf{x})), & g_9(\mathbf{x}) &= \sum_{i=9}^{12} \sin(x_i) + \sum_{i=9}^{12} \cos(x_i) \\
f^{10}(\mathbf{x}) &= g_{10}(\mathbf{x}) + \gamma(g_4(\mathbf{x}) + g_5(\mathbf{x})), & g_{10}(\mathbf{x}) &= \sum_{i=9}^{12} \sin(-x_i) + \sum_{i=9}^{12} \cos(-x_i).
\end{aligned}$$

This problem is defined as a weak perturbation, controlled by the small parameter  $\gamma$ , of the composition of the following low-dimensional subproblems:

$$\begin{aligned}
(x_1, x_2, x_3) &\mapsto (g_1(\mathbf{x}), g_2(\mathbf{x}), g_3(\mathbf{x})), & (x_4, x_5, x_6) &\mapsto (g_4(\mathbf{x}), g_5(\mathbf{x}), g_6(\mathbf{x})), \\
(x_7, x_8) &\mapsto (g_7(\mathbf{x}), g_8(\mathbf{x})), & (x_9, x_{10}, x_{11}, x_{12}) &\mapsto (g_9(\mathbf{x}), g_{10}(\mathbf{x})).
\end{aligned}$$

A sample of 104 points seems sufficiently accurate to perform the sensitivity analysis revealing the underlying structure of the weakly correlated subproblems. Because of the high dimension of the problem, we cannot consider a full factorial for building a reference Pareto optimal set. Therefore, we simply use a large random sample  $RS_{exh}$  ( $\# \{RS_{exh}\} = 10^6$ ) of points in the feasible set, evaluate the objective functions and extract the nondominated set, i.e.,  $(P_{exh}, F_{exh}) = ND(RS_{exh}, \mathbf{f}(RS_{exh}))$ ,  $\# \{P_{exh}\} = 359701$ . We proceed by comparing the results of RVEA for the original problem and the results of applying RVEA separately to the subproblems formed by ANOVA-MOP. We consider 10 generations for a starting dataset of 1000 points, such that the total number of function evaluations for solving the original problem is 10000. The nondominated set  $(P_1, F_1)$  contains 7473 points.

We apply then ANOVA-MOP and get four subproblems. We run 10 generations for each subproblem starting from a dataset of 250, 250, 100, 500 points, respectively, given that their decision space dimensions are 3, 3, 2 and 4, respectively. The total number of function evaluations is then  $(250 + 250 + 100 + 500) \times 10 = 11000$ . The nondominated sets contain, respectively, 220, 245, 98 and 26 points and we call them  $(P_{2,1}, F_{2,1}), \dots, (P_{2,4}, F_{2,4})$ , respectively. We proceed by evaluating the quality of the solutions found by the two instances of RVEA. For the decomposed problem we consider separately the loss for the subproblems and take the maximum loss as a global indicator of the performance. As can be seen in Table 3, the loss between the nondominated sets and their corresponding canonical projections of  $F_{exh}$  is considerably small. It means that ANOVA-MOP has managed to provide nondominated solutions close enough to  $F_{exh}$ .

dataset	loss( $F_{exh}$ , dataset)	$N_{evals}$
$(P_1, F_1)$	loss( $F_{exh}, F_1$ ) = 1.8978	$10^5$
$(P_{2,1}, F_{2,1})$	loss( $\pi^{(1)}(F_{exh}), F_{2,1}$ ) = 0.222087	2500
$(P_{2,2}, F_{2,2})$	loss( $\pi^{(2)}(F_{exh}), F_{2,2}$ ) = 0.249897	2500
$(P_{2,3}, F_{2,3})$	loss( $\pi^{(3)}(F_{exh}), F_{2,3}$ ) = 0.0866798	1000
$(P_{2,4}, F_{2,4})$	loss( $\pi^{(4)}(F_{exh}), F_{2,4}$ ) = 0.301691	5000

Table 3: Summary of comparison for the second example.

As for the previous benchmark, we denote by  $\pi^{(l)} : \mathbb{R}^k \rightarrow \mathbb{R}^{k_l}$  the canonical projection on the subspace of the objective function values of the  $l^{th}$  subproblem. Therefore, based on Table 3, the performance of the decomposed approach is 0.3 to be compared with 1.9, which comes from the original problem while the number of function evaluation is approximately the same. This means that with a fixed number of function evaluations, ANOVA-MOP has obtained solutions closer to the true Pareto frontier in comparison with the ones obtained by RVEA. We recall that this value of the loss can be used as an estimate of the thickness  $\epsilon$  of the tolerance zone in Corollary 18.

We further give another estimate of  $\epsilon$  by computing the maximum difference between the values of  $\mathbf{f}$  and of  $\mathbf{f}_\delta$  on a random subset  $V$  of the Pareto optimal set approximation as  $\epsilon \simeq \max_{\mathbf{x} \in V} \|\mathbf{f}(\mathbf{x}) - \bar{\mathbf{f}}_\delta(\mathbf{x})\|_\infty = 0.130699$ .

**4.3. Using ANOVA-MOP as an interactive method.** As discussed in Section 3.6, when using ANOVA-MOP as an interactive method, the solution obtained is the most preferred solution for the DM. In what follows, we demonstrate how this solution can be obtained by considering the problem discussed in Subsection 4.2.

We first informed the DM that the problem was decomposed into 4 subproblems and then told which objective functions were in the first subproblem. The DM provided the first reference point  $\bar{\mathbf{z}}_1^{1*} = (7, 5, 3)^T$ . The corresponding ASF problem including metamodels was solved. The solution  $\bar{\mathbf{f}}_1^{1*} = (4.5308, 1.0639, 1.0921)^T$  obtained was shown to him. He wanted to iterate and gave a new reference point  $\bar{\mathbf{z}}_2^{1*} = (2, 1.0639, 1.0921)^T$  to improve the value of the first objective function. The corresponding solution  $\bar{\mathbf{f}}_2^{1*} = (2.0458, 2.0461, 2.0749)^T$  was shown to the DM, who accepted it and thus the first subproblem was solved.

For the second subproblem, the DM provided a reference point  $\bar{\mathbf{z}}_1^{2*} = (4, 2.5, 6)^T$ . The DM accepted the solution  $\bar{\mathbf{f}}_1^{2*} = (2.0708, 2.0708, 2.07)^T$ . The solution process continued to the third subproblem. The DM provided a reference point  $\bar{\mathbf{z}}_1^{3*} = (5, 6)^T$  and obtained the solution  $\bar{\mathbf{f}}_1^{3*} = (2.0553, 2.094)^T$ . Here he wanted to improve the first objective value and provided a reference point  $\bar{\mathbf{z}}_2^{3*} = (1, 5)^T$ . The corresponding solution  $\bar{\mathbf{f}}_2^{3*} = (1.0735, 4.5764)^T$  was accepted.

As far as the fourth subproblem is concerned, the DM gave a reference point  $\bar{\mathbf{z}}_1^{4*} = (2, 6)^T$ . The solution  $\bar{\mathbf{f}}_1^{4*} = (1.1332, 4.1547)^T$  was shown to him and he provided  $\bar{\mathbf{z}}_2^{4*} = (1.1332, 3)^T$  to improve the second objective function value. The solution obtained  $\bar{\mathbf{f}}_2^{4*} = (2.2508, 3.7472)^T$  was satisfactory for the DM, therefore the process terminated.

Thanks to the decomposition, the DM could concentrate on two or three objec-

tive functions at a time. Because of this, it was easier to follow the changes in the solutions obtained after having specified a reference point and the solution process was not cognitively as demanding as when dealing with all objective functions simultaneously. Visualizations are also easy to analyze in these low dimensions.

Finally, the solution to the original problem is composed by joining the solutions of the subproblems. We recall that the composition of the values of such a solution is only an approximation of the real function values, therefore the composition of the decision variables must be evaluated on the original functions and proposed to the *DM* for his/her final decision. Here, the final approximated solution  $\bar{\mathbf{f}}$  and the solution  $\mathbf{f}$  evaluated with the original objective functions are:

$$\begin{aligned}\bar{\mathbf{f}} &= (2.0458, 2.0461, 2.0749, 2.0708, 2.0708, 2.0700, 1.0735, 4.5764, 2.2508, 3.7472)^T, \\ \mathbf{f} &= (2.0319, 2.0319, 2.0460, 2.0390, 2.0633, 2.0472, 1.0460, 4.5280, 2.2396, 3.7325)^T.\end{aligned}$$

Based on (17), we have,  $\epsilon \simeq \max_{\mathbf{x} \in V} \|\mathbf{f}(\mathbf{x}) - \bar{\mathbf{f}}_\delta(\mathbf{x})\|_\infty = 0.0484$ .

**5. Conclusions.** In this paper, we have developed a metamodel-based method called ANOVA-MOP to solve computationally expensive multiobjective optimization problems with high-dimensional decision and objective spaces. In this method, first a metamodel for each individual objective function is built. Total sensitivity indices are then estimated. Then, a set of thresholds is considered by using these indices. Each threshold leads to identifying a set of active decision variables which have influence on the objective function values. By increasing the threshold, the number of active variables changes. The threshold is increased until the objective functions with the same active variables can be grouped into a low-dimensional subproblem in the decision and objective spaces, or at least the number of decision variables of the original problem is decreased. This leads to decomposing the original problem into a limited number of simpler subproblems with low-dimensional decision and objective spaces. The solutions of the subproblems are composed with a Cartesian product to obtain an approximated Pareto optimal set for the original high-dimensional problem.

ANOVA-MOP can be applied as an interactive or a non-interactive method to solve computationally expensive MOPs with any number of objective functions. As an interactive method, due to the decomposition, in each step of the interaction with a DM, (s)he focuses on a lower number of objective functions at a time rather than all objective functions simultaneously and eventually finds a preferred solution. As a non-interactive method, a representative set of Pareto optima can be obtained and subproblems can be solved by parallel computing. This boosts the solution process considerably. In both interactive and non-interactive cases, since inactive variables are neglected in the subproblems, the dimensions of the decision spaces in the subproblems are reduced. This allows applying any appropriate optimization method for low-dimensional problems.

An interesting situation occurs when we do not wish to raise the tolerance threshold until the problem becomes decomposable, e.g., because we need to preserve the accuracy of the approximation. In this case it is possible to define subproblems with a lower dimension but with some common decision variables. It is clear that if these variables are fixed to any value, the remaining decision variables and the objective functions will represent independent subproblems. Thus, it seems promising to extend ANOVA-MOP to such, more general situations. This will be the subject of future research.

## REFERENCES

- [1] C. AUDET, A. IANNI, S. LE DIGABEL, AND C. TRIBES, *Reducing the number of function evaluations in mesh adaptive direct search algorithms*, SIAM Journal on Optimization, 24 (2014), pp. 621–642.
- [2] H. P. BENSON AND E. SUN, *Outcome space partition of the weight set in multiobjective linear programming*, Journal of Optimization Theory and Applications, 105 (2000), pp. 17–36.
- [3] G. E. P. BOX AND R. D. MEYER, *An analysis for unreplicated fractional factorials*, Technometrics, 28 (1986), pp. 11–18.
- [4] G. E. P. BOX AND K. B. WILSON, *On the experimental attainment of optimum conditions*, Journal of the Royal Statistical Society, Series B, XIII (1951), pp. 1–45.
- [5] R. D. BRAUN AND I. M. KROO, *Development and application of the collaborative optimization architecture in a multidisciplinary design environment*, in Proceedings of the ICASE/NASA Langley Workshop on Multidisciplinary Design Optimization, 1995, pp. 98–116.
- [6] L. CHEN, Z. DING, AND S. LI, *A formal two-phase method for decomposition of complex design problems*, Journal of Mechanical Design, 127 (2005), pp. 184–195.
- [7] R. CHENG, Y. JIN, M. OLHOFFER, AND B. SENDHOFF, *A reference vector guided evolutionary algorithm for many-objective optimization*, IEEE Transactions on Evolutionary Computation, 20 (2016), pp. 773–791.
- [8] F. CIUCCI, T. HONDA, AND M. C. YANG, *An information-passing strategy for achieving Pareto optimality in the design of complex systems*, Research in Engineering Design, 23 (2012), pp. 71–83.
- [9] N. COWAN, *The magical mystery four: How is working memory capacity limited, and why?*, Current Directions in Psychological Science, 19 (2010), pp. 51–57.
- [10] B. DANDURAND AND M. WIECEK, *Distributed computation of Pareto sets*, SIAM Journal on Optimization, 25 (2015), pp. 1083–1109.
- [11] B. EFRON AND C. STEIN, *The jackknife estimate of variance*, The Annals of Statistics, 9 (1981), pp. 586–596.
- [12] M. EHRGOTT AND S. NICKEL, *On the number of criteria needed to decide Pareto optimality*, Mathematical Methods of Operations Research, 55 (2002), pp. 329–345.
- [13] A. ENGAU AND M. M. WIECEK, *Interactive coordination of objective decompositions in multiobjective programming*, Management Science, 54 (2008), pp. 1350–1363.
- [14] M. GARDENGI, T. GÓMEZ, F. MIGUEL, AND M. M. WIECEK, *Algebra of efficient sets for multiobjective complex systems*, Journal of Optimization Theory and Applications, 149 (2011), pp. 385–410.
- [15] Y. Y. HAIMES AND D. LI, *Hierarchical multiobjective analysis for large-scale systems: Review and current status*, Automatica, 24 (1988), pp. 53–69.
- [16] L. HANNAH, W. POWELL, AND D. DUNSON, *Semiconvex regression for metamodeling-based optimization*, SIAM Journal on Optimization, 24 (2014), pp. 573–597.
- [17] M. HARTIKAINEN AND A. LOVISON, *Paint-SiCon: constructing consistent parametric representations of Pareto sets in nonconvex multiobjective optimization*, Journal of Global Optimization, 62 (2015), pp. 243–261.
- [18] C. H. HUANG, *Development of multi-objective concurrent subspace optimization and visualization methods for multidisciplinary design*, PhD thesis, Department of Mechanical and Aerospace Engineering, State University of New York at Buffalo, 2003.
- [19] C. H. HUANG AND C. L. BLOEBAUM, *Incorporation of preferences in multiobjective concurrent subspace optimization for multidisciplinary design*, in Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2004, pp. AIAA–2004–4548.
- [20] C. H. HUANG, J. GALUSKI, AND C. L. BLOEBAUM, *Multi-objective Pareto concurrent subspace optimization for multidisciplinary design*, AIAA Journal, 45 (2007), pp. 1894–1906.
- [21] M. J. W. JANSEN, *Analysis of variance designs for model output*, Computer Physics Communications, 117 (1999), pp. 35–43.
- [22] M. E. JOHNSON, L. M. MOORE, AND D. YLVIKAKER, *Minimax and maximin distance designs*, Journal of Statistical Planning and Inference, 26 (1990), pp. 131–148.
- [23] D. R. JONES, C. D. PERTTUNEN, AND B. E. STUCKMAN, *Lipschitzian optimization without the Lipschitz constant*, Journal of Optimization Theory and Applications, 79 (1993), pp. 157–181.
- [24] H. M. KIM, N. F. MICHELENA, P. Y. PAPALAMBROS, AND T. JIANG, *Target cascading in optimal system design*, Journal of Mechanical Design, 125 (2003), pp. 474–480.
- [25] S. KUTATELADZE, *Convex e-programming*, vol. 20, Soviet Mathematics - Doklady, 1979.
- [26] R. LAZIMY, *Solving multiple criteria problems by interactive decomposition*, Mathematical Programming, 35 (1986), pp. 334–361.
- [27] S. M. LEE AND B. H. RHO, *A multicriteria decomposition model for two-level decentralized organizations*, International Journal on Policy and Information, 9 (1985), pp. 119–133.
- [28] F. LEWIS, K.; MISTREE, *Collaborative, sequential, and isolated decisions in design*, Journal of Mechanical Design, 120 (1998), pp. 643–652.

- [29] K. LEWIS AND F. MISTREE, *Modeling interactions in multidisciplinary design: a game theoretic approach*, AIAA Journal, 35 (1997), pp. 1387–1392.
- [30] T. LONG, L. LIU, S. ZHOU, J. WANG, AND L. MENG, *Multi-objective multidisciplinary optimization of long-endurance UAV wing using surrogate model in modelcenter*, in 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, AIAA, 2008, pp. AIAA 2008–5918.
- [31] A. LOVISON, *Singular continuation: Generating piecewise linear approximations to Pareto sets via global analysis*, SIAM Journal on Optimization, 21 (2011), pp. 463–490.
- [32] A. LOVISON, *Global search perspectives for multiobjective optimization*, Journal of Global Optimization, 57 (2013), pp. 385–39.
- [33] A. LOVISON AND E. RIGONI, *Adaptive sampling with a Lipschitz criterion for accurate metamodeling*, Communications in Applied and Industrial Mathematics, 1 (2010), pp. 110–126.
- [34] A. LOVISON AND E. RIGONI, *Extracting optimal datasets for metamodeling and perspectives for incremental samplings*, Mathematics and Computers in Simulation, 81 (2010), pp. 681–692.
- [35] C. MALIVERT AND N. BOISSARD, *Structure of efficient sets for strictly quasi-convex objectives*, Journal of Convex Analysis, 1 (1994), pp. 143–150.
- [36] C. McALLISTER, T. SIMPSON, K. HACKER, K. LEWIS, AND A. MESSAC, *Integrating linear physical programming within collaborative optimization for multiobjective multidisciplinary design optimization*, Structural and Multidisciplinary Optimization, 29 (2005), pp. 178–189.
- [37] M. MCKAY, W. CONOVER, AND R. BECKMAN, *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics, 21 (1979).
- [38] K. MIETTINEN, *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers, Boston, 1999.
- [39] K. MIETTINEN, *Introduction to multiobjective optimization: Noninteractive approaches*, in Multiobjective Optimization: Interactive and Evolutionary Approaches, J. Branke, K. Deb, K. Miettinen, and R. Slowinski, eds., Springer-Verlag Berlin Heidelberg, 2008, pp. 1–26.
- [40] K. MIETTINEN, F. RUIZ, AND A. P. WIERZBICKI, *Introduction to multiobjective optimization: Interactive approaches*, in Multiobjective Optimization: Interactive and Evolutionary Approaches, J. Branke, K. Deb, K. Miettinen, and R. Slowinski, eds., Springer-Verlag Berlin Heidelberg, 2008, pp. 27–57.
- [41] M. D. MORRIS, L. M. MOORE, AND M. D. MCKAY, *Using orthogonal arrays in the sensitivity analysis of computer models*, Technometrics, 50 (2008), pp. 205–215.
- [42] C. H. PAPADIMITRIOU AND M. YANNAKAKIS, *On the approximability of trade-offs and optimal access of web sources*, in Proceedings 41st Annual Symposium on Foundations of Computer Science, 2000, pp. 86–92.
- [43] S. PARASHAR AND C. L. BLOEBAUM, *Multi-objective genetic algorithm concurrent subspace optimization (MOGACSSO) for multidisciplinary design*, in Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2006.
- [44] V. PARETO, *Cours d'économie politique/Professè à l'université de Lausanne*, Rouge, Lausanne, 1896–1897.
- [45] S. POLES AND A. LOVISON, *A polynomial chaos approach to robust multiobjective optimization*, in Hybrid and Robust Approaches to Multiobjective Optimization, K. Deb, S. Greco, K. Miettinen, and E. Zitzler, eds., no. 09041 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2009, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.
- [46] N. POPOVICI, *Pareto reducible multicriteria optimization problems*, Optimization, 54 (2005), pp. 253–263.
- [47] A. SALTELLI, K. CHAN, AND M. SCOTT, eds., *Sensitivity Analysis*, Wiley, New York, 2000.
- [48] T. J. SANTNER, B. J. WILLIAMS, AND W. I. NOTZ, *The design and analysis of computer experiments*, Springer-Verlag, New York, 2003.
- [49] Y. D. SERGEYEV AND D. E. KVASOV, *Global search based on efficient diagonal partitions and a set of Lipschitz constants*, SIAM Journal on Optimization, 16 (2006), pp. 910–937.
- [50] S. SHAN AND G. G. WANG, *Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions*, Structural and Multidisciplinary Optimization, 41 (2010), pp. 219–241.
- [51] C. S. SHIAU AND J. J. MICHALEK, *A game-theoretic approach to finding market equilibria for automotive design under environmental regulation*, in Proceedings of the ASME 2007 international design engineering technical conferences, 2007, pp. DETC2007/DAC–34884.
- [52] J. SOBIESZCZANSKI-SOBIESKI, *Optimization by decomposition: a step from hierarchic to non-hierarchic systems*, in Proceedings of the 2nd NASA/Air Force Symposium on Recent Advances in Multidisciplinary Analysis and Optimization, 1988, pp. 51–78.
- [53] J. SOBIESZCZANSKI-SOBIESKI, J. S. AGTE, AND R. R. SANDUSKY, *Bi-level integrated system synthesis*, in Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, 1998.

- [54] I. M. SOBOL, *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*, Mathematics and Computers in Simulation, 55 (2001), pp. 271–280.
- [55] M. TABATABAEI, J. HAKANEN, M. HARTIKAINEN, K. MIETTINEN, AND K. SINDHYA, *A survey on handling computationally expensive multiobjective optimization problems using surrogates: non-nature inspired methods*, Structural and Multidisciplinary Optimization, 52 (2015), pp. 1–25.
- [56] S. TAKAI, *A game-theoretic model of collaboration in engineering design*, Journal of Mechanical Design, 132 (2010), pp. 0510051–05100510.
- [57] M. TAN, *Sequential Bayesian polynomial chaos model selection for estimation of sensitivity indices*, SIAM/ASA Journal on Uncertainty Quantification, 3 (2015), pp. 146–168.
- [58] R. TAPPETA, J. RENAUD, AND J. RODRÍGUEZ, *An interactive multiobjective optimization design strategy for decision based multidisciplinary design*, Engineering Optimization, 34 (2002), pp. 523–544.
- [59] R. V. TAPPETA AND J. E. RENAUD, *Multiobjective collaborative optimization*, Journal of Mechanical Design, 119 (1997), pp. 403–411.
- [60] P. WANG, H. ZHU, M. WILAMOWSKA-KORSAK, Z. BI, AND L. LI, *Determination of weights for multiobjective decision making or machine learning*, IEEE Systems Journal, 8 (2014), pp. 63–72.
- [61] J. WARD, *Structure of efficient sets for convex objectives*, Mathematics of Operations Research, 14 (1989), pp. 249–257.
- [62] D. J. WHITE, *Epsilon efficiency*, Journal of Optimization Theory and Applications, 49 (1986), pp. 319–337.
- [63] N. WIENER, *The homogeneous chaos*, American Journal of Mathematics, 60 (1938), pp. 897–936.
- [64] A. P. WIERZBICKI, *On the completeness and constructiveness of parametric characterizations to vector optimization problems*, OR Spectrum, 8 (1986), pp. 73–87.
- [65] S. WILD AND C. SHOEMAKER, *Global convergence of radial basis function trust region derivative-free algorithms*, SIAM Journal on Optimization, 21 (2011), pp. 761–781.
- [66] C. F. J. WU AND M. S. HAMADA, *Experiments: Planning, Analysis, and Parameter Design Optimization*, Wiley, New York, 2 ed., 2009.
- [67] A. XIAO, S. ZENG, J. K. ALLEN, D. W. ROSEN, AND F. MISTREE, *Collaborative multidisciplinary decision making using game theory and design capability indices*, Research in Engineering Design, 16 (2005), pp. 57–72.
- [68] M. XIAO, X. SHAO, L. GAO, AND Z. LUO, *A new methodology for multi-objective multidisciplinary design optimization problems based on game theory*, Expert Systems with Applications, 42 (2015), pp. 1602 – 1612.
- [69] I. YEH, Y. FU, B. KACHNOWSKI, AND T. SUBBLAN, *Optimization of a vehicle restraint system using a genetic algorithm*, in SAE Technical Paper, SAE International, 04 2005.
- [70] D. ZHANG, S. TANG, AND J. CHE, *Concurrent subspace design optimization and analysis of hypersonic vehicles based on response surface models*, Aerospace Science and Technology, 42 (2015), pp. 39 – 49.
- [71] K. S. ZHANG, Z. H. HAN, W. J. LI, AND W. P. SONG, *Bilevel adaptive weighted sum method for multidisciplinary multi-objective optimization*, AIAA Journal, 46 (2008), pp. 2611–2622.
- [72] Q. ZHANG, W. LIU, E. TSANG, AND B. VIRGINAS, *Expensive multiobjective optimization by MOEA/D with Gaussian process model*, IEEE Transactions on Evolutionary Computation, 14 (2010), pp. 456–474.