



# A Partition Based Bayesian Multi-objective Optimization Algorithm

Antanas Žilinskas<sup>(✉)</sup>  and Linas Litvinas

Institute of Data Science and Digital Technologies,  
Vilnius University, Akademijos Street 4, 08663 Vilnius, Lithuania  
antanas.zilinskas@mii.vu.lt, linas.litvinas@mif.vu.lt

**Abstract.** The research is aimed at coping with the inherent computational intensity of Bayesian multi-objective optimization algorithms. We propose the implementation which is based on the rectangular partition of the feasible region and circumvents much of computational burden typical for the traditional implementations of Bayesian algorithms. The included results of the solution of testing and practical problems illustrate the performance of the proposed algorithm.

**Keywords:** Non-convex optimization · Multi-objective optimization · Bayesian approach

## 1 Introduction

Applied optimization problems, especially those in engineering design, frequently are multi-objective and non-convex. The class of non-convex objective functions is non-homogeneous from the point of view of design of optimization algorithms. We address the black-box problems differently from the problems, where objective functions and constraints are described by mathematical formulas. Methods for the latter problems generalize classical mathematical programming methods [8, 10]. Metaheuristic methods are popular for black-box problems [5]. However, the metaheuristic methods frequently are not appropriate for the expensive black-box problems because of the limited budget of the computations of the objective functions.

The optimization of expensive black-box objective functions can be considered as a sequence of decisions under uncertainty, and the ideas of the theory of rational decision making seem most appropriate for the development of the corresponding algorithms. Gaussian random fields (GRF) normally are used as models representing uncertainty about aimed objective functions. Bayesian algorithms are designed maximizing a criterion of average utility with respect to the random field chosen for a model. The most frequently used criteria are: the maximum average improvement and the maximum improvement probability. The research and applications of the single-objective Bayesian methods is booming during last years. The Bayesian approach has recently extended to multi-objective optimization. The idea of the maximization of the multi-objective improvement

probability is implemented in [17]. A popular alternative method is based on the reduction to single objective optimization where the hyper-volume of a sub-region of the objective space, bounded by the approximation of the Pareto front, is maximized; see e.g. [6, 7]. For the relevant single-objective black-box optimization methods we refer to [11–14].

The inherent computational burden of Bayesian multi-objective algorithms bounds their application area similarly to the single-objective case. In the present paper we propose coping with such a challenge by the implementation based on the rectangular partition of the feasible region. We merge the ideas of [17] (where the multi-objective P-algorithm was proposed), and of [3] (where the partition based single-objective P-algorithm was implemented). The developed algorithm was applied to the optimization of a biochemical process modeling of which is computationally extremely intensive.

## 2 The Proposed Algorithm

A black-box multi-objective minimization problem is considered

$$\min F(x), x \in A \subset \mathbb{R}^d, F(x) = (f_1(x), \dots, f_m(x))^T; \tag{1}$$

we assume that  $A$  is of simple structure, e.g., a hyper-rectangular. We assume that objective functions are computationally expensive.

We start from a brief introduction of the original single-objective ( $m = 1$ ) P-algorithm. For a review of related methods we refer to [15]. A GRF  $\xi(x)$ ,  $x \in A$ , is accepted for a model of an objective function. The results of  $k$  function evaluations  $y_i = f(x_i)$ ,  $i = 1, \dots, k$ , are available and can be taken into account for planning a current evaluation point. The P-algorithm selects for the evaluation of the objective function the point of maximum of conditional improvement probability

$$x_{k+1} = \arg \max_{x \in A} \mathbb{P}\{\xi(x) < y_{ok} \mid \xi(x_i) = y_i, i = 1, \dots, k\}, \tag{2}$$

where  $y_{ok} = \min\{y_1, \dots, y_k\} - \varepsilon_k$ ,  $\varepsilon_k > 0$  is an improvement threshold.

A generalization of the P-algorithm to the multi-objective case was proposed in [17]. A vector valued GRF  $\Xi(x) = (\xi_1(x), \dots, \xi_m(x))^T$ ,  $m > 1$ ,  $x \in A$ , is accepted for a model of objective functions. Let  $Y_i = F(x_i)$ ,  $i = 1, \dots, k$ , denote the vectors of objective function values evaluated in previous iterations, and  $Y_{ok}$  be a reference point. The P-algorithm computes a current vector of the objectives at the point

$$x_{k+1} = \arg \max_{x \in A} \mathbb{P}\{\Xi(x) < Y_{ok} \mid \Xi(x_i) = Y_i, i = 1, \dots, k\}. \tag{3}$$

The main disadvantage of the standard implementation of the multi-objective P-algorithm is its computational burden where the maximization problem (3) is multimodal, and the computation of  $\mathbb{P}(\cdot)$  involves inverting large ill conditioned

matrices. The computational burden in a single-objective case can be substantially reduced by the partition based implementation as shown in [3]. We will generalise that implementation, in the present paper, for the multi-objective case.

Let the feasible region  $A$  be a hyper-rectangle. The algorithm is designed as a sequence of subdivisions by means of the bisection. The selected hyper-rectangle is bisected by a hyper-plane orthogonal to its longest edges. The values of  $F(x)$  are computed at  $2^{d-1}$  intersection points. A hyper-rectangle is selected for the subdivision according to a criterion which is an approximation of (3). The criterion of a hyper-rectangle is computed as the conditional improvement probability at its center  $x_c$ . The computational complexity for that probability is defined by the complexity of the computations of the conditional mean  $\mu(x_c | \cdot)$  and conditional variance  $\sigma^2(x_c | \cdot)$  of  $\Xi(x_c)$ . We approximate the conditional probability in (3) by restricting information used in the definition of  $\mu(x_c | \cdot)$  and  $\sigma^2(x_c | \cdot)$  with function values at the vertices of the considered rectangle. Thus the computational complexity of  $\mu(\cdot)$  and  $\sigma(\cdot)$  in the proposed implementation is lower than in the standard implementation thanks to the restriction of the involved information. Further, the expressions of  $\mu(\cdot)$  and  $\sigma(\cdot)$  are replaced with their asymptotic expressions obtained by shrinking the hyper-rectangle to a point [3]:

$$\mu(x_c | \cdot) \sim \frac{1}{|I|} \sum_{i \in I} Y_i, \quad \sigma^2(x_c | \cdot) \sim V, \quad (4)$$

where  $I$  denotes the set of indices of the vertices, and  $V$  denotes the hyper-volume of the considered hyper-rectangle. These simplifications imply the following expression of the selection criterion

$$\frac{V}{\left\| \sum_{i \in I} Y_i - Y_{ok} \right\|}.$$

The high asymptotic convergence of the bi-objective version of that algorithm is proved in [4]. The partition of the feasible region at the initial iterations is quite uniform, thus it is rational with respect to the modest information about the considered objective functions at the initial iterations [16]. The accumulated information guides the selection of hyper-rectangles at later iterations towards the set of efficient decisions. The computational complexity of the proposed algorithm at a current iteration  $t$  can be evaluate similarly to the single-objective optimization algorithm [20] since the same operations are performed to manage the accumulated data. Thus the computational complexity of iteration  $t$  is  $T(n) = O(n \times m \times \log(n \times m))$ , where  $n$  is the number of evaluations of  $F(x)$  made at previous iterations. The complexity of computations at a current  $t$  iteration of the standard implementation of Bayesian algorithms, e.g. of the algorithm of average improvement, is much higher than  $T(n)$  (here  $t = n$ ). The iteration  $t$  includes inverting of, generally speaking,  $m \ n \times \ n$  matrices the time-complexity of which is  $O(m \times n^3)$ . Moreover, the inverting of the considered matrices is challenging since their condition numbers typically are very large. The other serious computational complexity of a standard implementation is the maximization of the average improvement which is a non-convex optimization problem.

### 3 Comparison with the Standard Implementation of the P-Algorithm

The proposed algorithm is a simplified version of its predecessor, i.e., of the standard implementation of the P-algorithm. It is interesting to compare their performance. The standard implementation is described in detail in [17] where several test problems are solved to illustrate its performance. We use the same test problems.

A non-convex problem, proposed in [9], is quite frequently used for testing multi-objective algorithms; see e.g., [5]. The objective functions are

$$\begin{aligned} f_1(\mathbf{x}) &= 1 - e^{-\sum_{i=1}^d (x_i - 1/\sqrt{d})^2}, \\ f_2(\mathbf{x}) &= 1 - e^{-\sum_{i=1}^d (x_i + 1/\sqrt{d})^2}, \end{aligned} \tag{5}$$

$d = 2$ , and the feasible region is  $\mathbf{A} : -4 \leq x_1, x_2 \leq 4$ . The next test problem is composed of two Shekel functions which are frequently used for testing of single-objective global optimization algorithms:

$$\begin{aligned} f_1(\mathbf{x}) &= -\frac{0.1}{(0.1 + (x_1 - 0.1)^2 + 2(x_2 - 0.1)^2)} \\ &\quad - \frac{0.1}{(0.14 + 20((x_1 - 0.45)^2 + (x_2 - 0.55)^2))}, \\ f_2(\mathbf{x}) &= -\frac{0.1}{(0.15 + 40((x_1 - 0.55)^2 + (x_2 - 0.45)^2)} \\ &\quad - \frac{0.1}{(0.1 + (x_1 - 0.3)^2 + (x_2 - 0.95)^2)}. \end{aligned} \tag{6}$$

The visualization of these test problems including graphs of the Pareto fronts and sets of Pareto optimal decisions is presented, e.g. in [5, 10].

Several metrics are used for the quantitative assessment of the precision of a Pareto set approximation. For the comparison of the proposed implementation of the P-algorithm with the standard one we apply the metrics which were used in recent publications related to the standard implementation. The generational distance (GD) is used to estimate the distance between the found approximation and the true Pareto front [5]. GD is computed as the maximum of distances between the found non-dominated solutions and their closest neighbors from the Pareto set. The epsilon indicator (EI) is a metric suggested in [21] which integrates measures of the approximation precision and spread: it is the min max distance between the Pareto front and the set of the found non-dominated solutions

$$EI = \max_{1 \leq i \leq K} \min_{1 \leq j \leq N} \|Z_i - F_j^*\|, \tag{7}$$

where  $F_j^*, j = 1, \dots, N$ , are the non-dominated solutions found by the considered algorithm, and  $\{Z_i, i = 1, \dots, K\}$  is the set of points well representing the Pareto set, i.e.  $Z_i$  are sufficiently densely and uniformly distributed over the Pareto set.

The algorithms were stopped after 100 computations of  $F(x)$ . Although the P-algorithm theoretically is deterministic, its version implemented in [17] is randomised because of a stochastic maximization method used for (3). Therefore, test problem were solved 100 times. We present the mean values and standard deviations of the considered metrics from [10] in two columns of Table 1. The proposed algorithm is deterministic; thus its results occupy single column for each test problem. The results of the proposed algorithm for the test problem (5) are even better than the results of the standard version of the P-algorithm. However, the standard version outperforms the proposed algorithm in solving (6). The set of optimal decisions of the latter problem consists of three disjoint subsets, and the diameter of one of subsets is relatively small. A considerable number of partitions of the feasible region was needed to indicate the latter subset. The experimentally measured solution time was 8.6 ms for the proposed algorithm, and 3.9 s for the algorithm of [17].

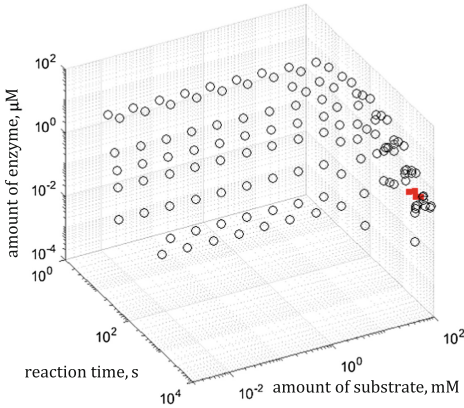
**Table 1.** Performance criteria of the standard and partition based implementations of the P-algorithm for Problems (5) and (6)

Implementation	Standard				Partition based	
Problem	Problem (5)		Problem (6)		Problem (5)	Problem (6)
NN	9.87	1.4	15.7	2.0	27	18
GD	0.015	0.0061	0.070	0.051	0.015	0.21
EI	0.20	0.034	0.13	0.053	0.092	0.25

## 4 Performance Evaluation on a Real World Problem

Quite many optimization problems in biotechnology can be characterized as black-box expensive ones. For example, optimal design of bio-sensors and bio-reactors requires solving optimization problems the objective functions of which are defined by computer models of high complexity [1, 2]. The micro bio-reactors are computationally modeled by a two-compartment model based on reaction-diffusion equations containing a nonlinear term related to the Michaelis-Menten enzyme kinetics; we refer to [1, 2] for the description and substantiation of the model. The computation of one objective function value of such problems take up to 10 min, and for some special cases possibly more. However, we are optimistic about these challenges since the problems in question are low dimensional.

We consider an optimization problem related to the optimal design of a micro bio-reactor, which well represents real world problems for which the proposed algorithm is potentially appropriate. Specific technical aspects of the computer model of the micro bio-reactor, and the substantiation of the objective functions are addressed in [18].



**Fig. 1.** Pareto front of the set of objective vectors computed by the proposed and genetic algorithms. (Color figure online)

A single value of the first objective function is 4.32 min (a computer with Intel Xeon X5650 2.66 GHz processor was used). A long reaction time means that the bio-reactor with such parameters is not appropriate; correspondingly, the simulation was interrupted if it exceeded 10 min.

We demonstrate the performance of the proposed algorithm under the conditions typical to a real world applications. The following optimization problem is considered where  $F(x)$  is available as a function in  $C$ :

$$\min_{x \in A} F(x), F(x) = (f_1(x), f_2(x), f_3(x))^T,$$

$$A = \{x : -4 \leq x_1, x_2 \leq -3, -8 \leq x_3 \leq -4, -5 \leq x_4 \leq -1\}. \quad (8)$$

We focus here on black-box optimization without discussing its applied aspects. The problem will be presented from the point of view of applied optimal designed in the next paper co-authored with experts in biotechnology. The proposed algorithm was applied to the solution of (8) with the predefined budget of evaluations of  $F(x)$  equal to 1000; the solution time was 72 h and 14 min. The genetic algorithm (GA) from the MATLAB Optimization Toolbox was also applied to the solution of the considered problem. The following parameters of GA were chosen: the population size was 50, the crossover fraction was 0.8, Pareto fraction was 0.35, and the other parameters were chosen as predefined in the Optimization Toolbox. The termination condition of GA was the same: budget of the objective function evaluations equal to 1000. GA is a randomized algorithm, thus results of a single run are not sufficiently reliable from the point of view of statistical testing. Indeed, the use of real world problems with expensive objective functions for testing randomised algorithms is challenging especially where the experimentation time is limited.

The Pareto front approximation computed by the proposed algorithm consisted of 124 vectors, and the approximation computed by GA consisted of 15 vectors. None vector of the of first approximation was dominated by a vector of

Optimal design of a micro bio-reactor is formulated as a three-objective optimization problem of four variables. The objectives are: the time of the reactions; the amount of the substrate per volume unite converted to the product; the total amount of the enzyme used per volume unit of the reactor; first and third objectives are minimized, and the second one is maximized. The variables are: two constructive parameters of a bio-reactor, and the concentrations of the enzyme and of the substrate.

The computation time depends on the parameters of the bio-reactor; the average time of computation of a single value of the first objective function is 4.32 min (a computer with Intel Xeon X5650 2.66 GHz processor was used). A long reaction time means that the bio-reactor with such parameters is not appropriate; correspondingly, the simulation was interrupted if it exceeded 10 min.

the second approximation. Only four vectors of the second approximation were not dominated by the vectors of the first approximation. The hypervolume indicators of both approximations were equal to 42.3 and 12.4 correspondingly. Thus the proposed algorithm clearly outperforms GA. The non-dominated vectors (in the commonly used physical units) of the union of both approximations are presented in Fig. 1 where circles denote the points of the first approximation, and the red rectangles denote the points of the second one. This figure shows only a general shape of the Pareto front. User oriented interfaces and visualization methods are available to experts in biotechnology to aid selecting an appropriate Pareto optimal decision; the advantages of visualising not only Pareto optimal solutions but also Pareto optimal decisions are argued in [18].

The performance of the proposed method is appropriate for low dimensional black-box expensive problems. Extensions to higher dimensionality can be challenging because of large number of computations of the objective function values ( $2^{d-1}$ ) at a current iteration. We plan the investigation of the simplicial partition based Bayesian algorithms to cope with that challenge; note that such a partition proved efficient in Lipschitzian optimization [11]. The hybridization of the proposed global search algorithm with a local one seems promising. The good performance of hybrids of the single-objective Bayesian algorithms with local ones [19,20] gives hope that similar synergy of the global and local search strategies will be achieved also in the multi-objective case.

## 5 Conclusions

A rectangular partition based implementation of a Bayesian multi-objective method is proposed where the typical for Bayesian algorithms inherent computational burden is avoided. The proposed algorithm is appropriate for solving practical problems characterized as expensive black-box problems of modest dimensionality. The prospective directions of further research are: increasing dimensionality of efficiently solvable problems, and studying efficiency of other partition strategies.

**Acknowledgments.** This work was supported by the Research Council of Lithuania under Grant No. P-MIP-17-61. We thank the reviewers for their valuable remarks.

## References

1. Baronas, R., Ivanauskas, F., Kulys, J.: Mathematical modeling of biosensors based on an array of enzyme microreactors. *Sensors* **6**(4), 453–465 (2006)
2. Baronas, R., Kulys, J., Petkevičius, L.: Computational modeling of batch stirred tank reactor based on spherical catalyst particles. *J. Math. Chem.* **57**(1), 327–342 (2019)
3. Calvin, J., Gimbutienė, G., Phillips, W., Žilinskas, A.: On convergence rate of a rectangular partition based global optimization algorithm. *J. Global Optim.* **71**, 165–191 (2018)

4. Calvin, J., Žilinskas, A.: On efficiency of a single variable bi-objective optimization algorithm. *Optim. Lett.* **14**(1), 259–267 (2020). <https://doi.org/10.1007/s11590-019-01471-4>
5. Deb, K.: *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley, Chichester (2009)
6. Emmerich, M., Deutz, A.H., Yevseyeva, I.: On reference point free weighted hypervolume indicators based on desirability functions and their probabilistic interpretation. *Proc. Technol.* **16**, 532–541 (2014)
7. Feliot P., Bect J., Vazquez E.: User preferences in Bayesian multi-objective optimization: the expected weighted hypervolume improvement criterion. [arXiv:1809.05450v1](https://arxiv.org/abs/1809.05450v1) (2018)
8. Floudas, C.: *Deterministic Global Optimization: Theory, Algorithms and Applications*. Kluwer, Dordrecht (2000)
9. Fonseca, C.M., Fleming, P.J.: On the performance assessment and comparison of stochastic multiobjective optimizers. In: Voigt, H.-M., Ebeling, W., Rechenberg, I., Schwefel, H.-P. (eds.) PPSN 1996. LNCS, vol. 1141, pp. 584–593. Springer, Heidelberg (1996). <https://doi.org/10.1007/3-540-61723-X-1022>
10. Pardalos, P., Žilinskas, A., Žilinskas, J.: *Non-Convex Multi-Objective Optimization*. Springer, Cham (2017). <https://doi.org/10.1007/978-3-319-61007-8>
11. Paulavičius, R., Žilinskas, J.: *Simplicial Global Optimization*. Springer, New York (2014). <https://doi.org/10.1007/978-1-4614-9093-7>
12. Paulavičius, R., Sergeyev, Y., Kvasov, D., Žilinskas, J.: Globally-biased DISIMPL algorithm for expensive global optimization. *J. Global Optim.* **59**, 545–567 (2014)
13. Sergeyev, Y., Kvasov, D., Mukhametzhanov, M.: On the efficiency of nature-inspired metaheuristics in expensive global optimization with limited budget. *Sci. Rep.* **453**, 1–8 (2018)
14. Sergeyev, Y., Kvasov, D.: *Deterministic Global Optimization: An Introduction to the Diagonal Approach*. Springer, New York (2017). <https://doi.org/10.1007/978-1-4939-7199-2>
15. Žilinskas, A., Zhigljavsky, A.: Stochastic global optimization: a review on the occasion of 25 years of informatica. *Informatica* **27**(2), 229–256 (2016)
16. Žilinskas, A.: On the worst-case optimal multi-objective global optimization. *Optim. Lett.* **7**(8), 1921–1928 (2013)
17. Žilinskas, A.: A statistical model-based algorithm for black-box multi-objective optimisation. *Int. J. Syst. Sci.* **45**(1), 82–92 (2014)
18. Žilinskas, A., Baronas, R., Litvinas, L., Petkevičius, L.: Multi-objective optimization and decision visualization of batch stirred tank reactor based on spherical catalyst particles. *Nonlinear Anal. Model. Control* **24**(6), 1019–1033 (2019)
19. Žilinskas, A., Žilinskas, J.: A hybrid global optimization algorithm for non-linear least squares regression. *J. Global Optim.* **56**(2), 265–277 (2013)
20. Žilinskas, A., Gimbutienė, G.: A hybrid of Bayesian approach based global search with clustering aided local refinement. *Commun. Nonlinear Sci. Numer. Simul.* **78**, 104785 (2019)
21. Zitzler, E., Thiele, L., Lauman, M., Fonseca, C., Fonseca, V.: Performance measurement of multi-objective optimizers: an analysis and review. *IEEE Trans. Evol. Comput.* **7**(2), 117–132 (2003)