

# Towards Single- and Multiobjective Bayesian Global Optimization for Mixed Integer Problems

Kaifeng Yang<sup>a)</sup>, Koen van der Blom<sup>b)</sup>, Thomas Bäck<sup>c)</sup> and Michael Emmerich<sup>d)</sup>

LIACS, Leiden University, Niels Bohrweg 1, 2333CA Leiden, Netherlands

<sup>a)</sup>Corresponding author: k.yang@liacs.leidenuniv.nl

<sup>b)</sup>k.van.der.blom@liacs.leidenuniv.nl

<sup>c)</sup>t.h.w.baeck@liacs.leidenuniv.nl

<sup>d)</sup>m.t.m.emmerich@liacs.leidenuniv.nl

**Abstract.** Bayesian Global Optimization (BGO) is a very efficient technique to optimize expensive evaluation problems. However, the application domain is limited to continuous search spaces when using a BGO algorithm. To solve mixed integer problems with a BGO algorithm, this paper adapts the heterogeneous distance function to construct the Kriging models and applies these new Kriging models in Multi-objective Bayesian Global Optimization (MOBGO). The proposed mixed integer MOBGO algorithm and the traditional MOBGO algorithm are compared on three mixed integer multi-objective optimization problems (MOP), w.r.t. the mean value of the *hypervolume* (HV) and the related standard deviation.

## ALGORITHM

*Bayesian Global Optimization* was proposed by the Lithuanian research group of Jonas Mockus and Antanas Žilinskas [1, 2]. The basic idea of BGO, also known as *Efficient Global Optimization*, is to build a statistical model to reflect the relationship between decision vectors  $\mathbf{X} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)})^\top$  in an  $m$  dimensional search space, and their corresponding objective values  $Y(\mathbf{X}) = (y(\mathbf{x}^{(1)}), y(\mathbf{x}^{(2)}), \dots, y(\mathbf{x}^{(n)}))^\top$ . Then an optimizer in a BGO algorithm searches for an optimal solution by using the predictions of the surrogate model(s), instead of evaluating the real objective function.

To construct a surrogate model, it is assumed that the objective function is the realization of a Gaussian random field, which is also called Gaussian processes (GP) or Kriging in BGO. Specifically, Kriging assumes  $y$  to be a realization of a random process  $Y$  and to be of the form [3]:

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + \epsilon(\mathbf{x}) \quad (1)$$

where  $\mu(\mathbf{x})$  is the estimated mean value over all given sampled points, and  $\epsilon(\mathbf{x})$  is a realization of a normally distributed Gaussian random process with a zero mean and a variance of  $\sigma^2$ . The correlation between the deviations at two points ( $\mathbf{x}$  and  $\mathbf{x}'$ ) using an isotropic Gaussian Kernel is defined as:

$$\text{Corr}[\epsilon(\mathbf{x}), \epsilon(\mathbf{x}')] = R(\mathbf{x}, \mathbf{x}') = \exp^{-\theta d(\mathbf{x}, \mathbf{x}')^2} \quad (\theta \geq 0) \quad (2)$$

where  $R(\cdot, \cdot)$  is the correlation function, and  $d(\cdot, \cdot)$  is the Euclidean distance function.

In Kriging, the distance value between two points is usually calculated with the Euclidean metric, that is  $d(\mathbf{x}, \mathbf{x}') = \sqrt{\sum_{i=1}^m (x_i - x'_i)^2}$ . This is useful and straightforward in a continuous search space, as the Euclidean metric assumes the continuity of an objective function. However, this assumption will not be suitable when the search space is a nominal discrete or an integer space.

To solve this problem, we applied the heterogeneous metric [4, 5] as the distance function, which is defined as:

$$d_h(\mathbf{x}, \mathbf{x}') = \sqrt{\sum_{i=1}^{i=n_r} (r_i - r'_i)^2 + \sum_{i=1}^{i=n_z} |z_i - z'_i| + \sum_{i=1}^{i=n_d} I(d_i \neq d'_i)} \quad (3)$$

where  $I(true) = 1$  and  $I(false) = 0$ . The idea of the heterogeneous metric is to combine different metrics by taking the square root of the sum of each distance. As shown in Eq. (3), the distance between any two points in a nominal discrete space and an integer space is measured by the overlap metric and the Manhattan metric, respectively. A detailed motivation of the heterogeneous metric in the context of statistical modeling is provided in [5].

## TEST PROBLEMS

Three mixed integer MOPs are used in this paper, specifically: the double sphere function in Eq. (4) [5], the double barrier function in Eq. (5) [6], and a multi-layer optical filter optimization problem [7].

$$f_{sphere_1}(\mathbf{r}, \mathbf{z}, \mathbf{d}) = \sum_{i=1}^{n_r} r_i^2 + \sum_{i=1}^{n_z} z_i^2 + \sum_{i=1}^{n_d} d_i^2 \quad f_{sphere_2}(\mathbf{r}, \mathbf{z}, \mathbf{d}) = \sum_{i=1}^{n_r} (r_i - 2)^2 + \sum_{i=1}^{n_z} (z_i - 2)^2 + \sum_{i=1}^{n_d} (d_i - 2)^2 \quad (4)$$

$$f_{barrier_1}(\mathbf{r}, \mathbf{z}, \mathbf{d}) = \sum_{i=1}^{n_r} (r_i^2 + \alpha \sin(r_i)^2) + \sum_{i=1}^{n_z} A[z_i]^2 + \sum_{i=1}^{n_d} B_i[d_i]^2 \quad (5)$$

$$f_{barrier_2}(\mathbf{r}, \mathbf{z}, \mathbf{d}) = \sum_{i=1}^{n_r} ((r_i - 2)^2 + \alpha \sin(r_i - 2)^2) + \sum_{i=1}^{n_z} (A[z_i] - 2)^2 + \sum_{i=1}^{n_d} (B_i[d_i] - 2)^2$$

In Eq. (4) and Eq. (5),  $\mathbf{r}$ ,  $\mathbf{z}$ , and  $\mathbf{d}$  represent a real variable vector, an integer variable vector, and a nominal discrete variable vector, respectively.

For the barrier function  $\alpha = 1$ , and  $A$  is generated by Algorithm 6 from Li et al (2013) [6] with the parameter  $C = 75$ , and  $B_{i \in 1, \dots, n_d}$  is a set of  $n_d$  random permutations of the sequence  $0, \dots, 4$ . Both  $A$  and  $B$  remain fixed throughout the experiments. Unlike in [6], here smooth wave-like barriers are used in the real part, rather than staircase-like barriers. For the multi-objective case both of these functions can be adjusted with an offset for each term, such that real, integer, and nominal discrete optima are different in the second objective.

For the optical filter problem [7], the pairs of continuous and (binary) nominal discrete variables are considered. When the binary variable is active the corresponding continuous variable is used in the objective functions, otherwise it is ignored. If all bits are inactive a penalty of (250, 1250) is returned. In addition to the original objective we consider  $f_{optfilt_2}(\mathbf{r}, \mathbf{d}) = \sum_{i=1}^{n_r} r_i d_i$  as a second objective. The parameters of each test problem are given in Table 1.

## RESULTS

All the experiments are performed on the same computer: Intel(R) i7-4800mq CPU @ 2.70GHz, RAM 32GB. The operating system is Ubuntu 16.04 LTS (64 bit), and the platform is MATLAB 8.4.0.150421 (R2014b), 64 bit.

Figure 1 shows the heatmap comparison of the predictions by using different distance metrics for the barrier function in Eq. (5). For the visualization, only one variable is considered for each variable type. The parameter  $\theta$  in Eq. (2) is chosen as  $\theta = [0.01, 0.01, 0.01]$ . Each variable is normalized to  $[0, 1]$  in Figure 1. The number of sampling points is 15 and there are 200 random points for the test points. Note that the pictures in the third column don't have sampling points and are plotted with these 200 test points, which are evaluated on the real objective functions. These 200 test points are used to visualize the landscape of the barrier function. In these figures, the models using the heterogeneous metric are more accurate than those using the Euclidean metric, especially in the  $f_{barrier_1}$  function.

Figure 2 illustrates the comparison of the predictions of the sphere functions using the Euclidean metric and the heterogeneous metric. There are 15 variables in total, 5 variables for each type. The number of sampling points is 90 and the number of test points is 200. The optimal  $\theta$  in the Kriging models are calculated with the simplex search method of Lagarias et al. (*fminsearch*) [8], with the parameter of max function evaluations as 1000. In Figure 2,  $\hat{y}$  and  $\sigma$  represent the predicted mean and the predicted standard deviation, respectively. After applying the optimal  $\theta$  strategy for both the Euclidean metric and the heterogeneous metric, the prediction of using the Euclidean metric is similar to that of using the heterogeneous metric. However, the Kriging models using the heterogeneous metric still outperform the Kriging models using the Euclidean metric, w.r.t. the standard deviation of the predictions.

As the execution time of each algorithm requires more than 3 hours for only one experiment, ten repetitions of each algorithm are performed. Table 1 shows the mean and standard deviation of the final hypervolume values on the test problems. All the parameters of each algorithm are also shown in Table 1, including reference points (ref.). The number of initialized sampling points ( $N_{init.}$ ) is 90 and the max number of function evaluations ( $N_{max}$ ) is 200. For

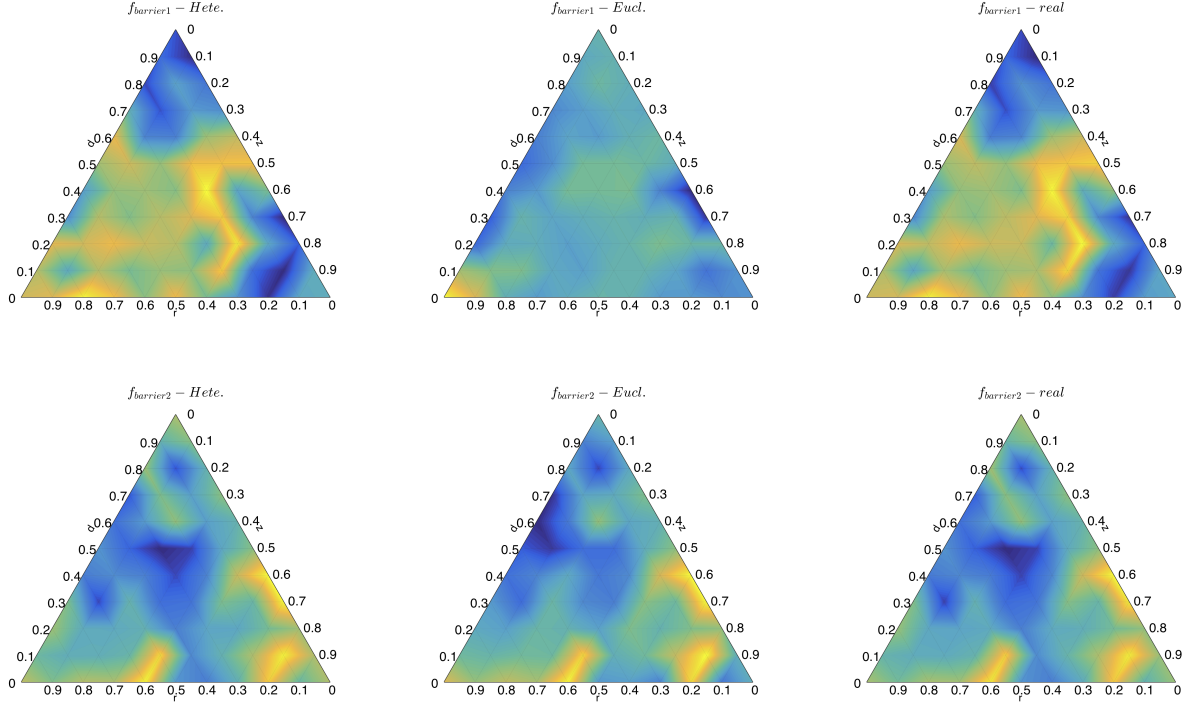


FIGURE 1. Landscape of the  $f_{mbarrier}$  function

both algorithms, the optimal  $\theta$  calculated by  $fminsearch$  [8] are used, the infill criterion is the *Expected Hypervolume Improvement* [9] and the optimizer is a build-in genetic algorithm [10] in MATLAB. The results in Table 1 show that the mixed integer MOBGO using the heterogeneous metric outperforms the traditional MOBGO on both the barrier and sphere test problems, w.r.t. mean hypervolume (HV) values and the standard deviations of the HV over 10 repetitions. For the optical filter problem, the traditional MOBGO can generate a better Pareto front set, but the mixed integer MOBGO is more robust than the traditional MOBGO, w.r.t. standard deviation.

TABLE 1. Parameter settings and empirical experimental results

	HV - Eucl.		HV - Hete.		ref.	$n_r$	Para. setting				
	mean	std.	mean	std.			range	$n_z$	range	$n_d$	range
$f_{sphere}$	8.4644e+03	98.5064	<b>8.5446e+03</b>	<b>18.3754</b>	[100, 100]	5	[0, 4]	5	[0, 4]	5	[0, 4]
$f_{barrier}$	6.4769e+03	175.5024	<b>6.6927e+03</b>	<b>55.9169</b>	[100, 100]	5	[0, 4]	5	[0, 4]	5	[0, 4]
$f_{optfilt}$	<b>1.5469e+04</b>	27.4487	1.5456e+04	<b>6.3562</b>	[50, 800]	7	[0, 1]	N/A	N/A	7	[0, 1]

## CONCLUSIONS AND FUTURE WORK

This paper proposes a mixed integer MOBGO by using the heterogeneous metric, instead of the Euclidean metric. Both of these two algorithms are compared in terms of the prediction errors and performance of the final Pareto front sets on three mixed integer multi-objective test problems. The results show that the proposed method surpasses the traditional MOBGO on both sphere and barrier test functions, with regard to the prediction errors, mean HV and standard deviation of the HV. On the optical filter problem, the traditional MOBGO performs slightly better than the proposed algorithm. The reason for this could be that  $N_{max}$  is too small, as this problem contains 14 variables.

In future work, we will compare the performance of the proposed algorithm with another recent integer-based BGO algorithm using the one-hot encoding strategy, proposed by Garrido-Merchán and Hernández-Lobato [11].

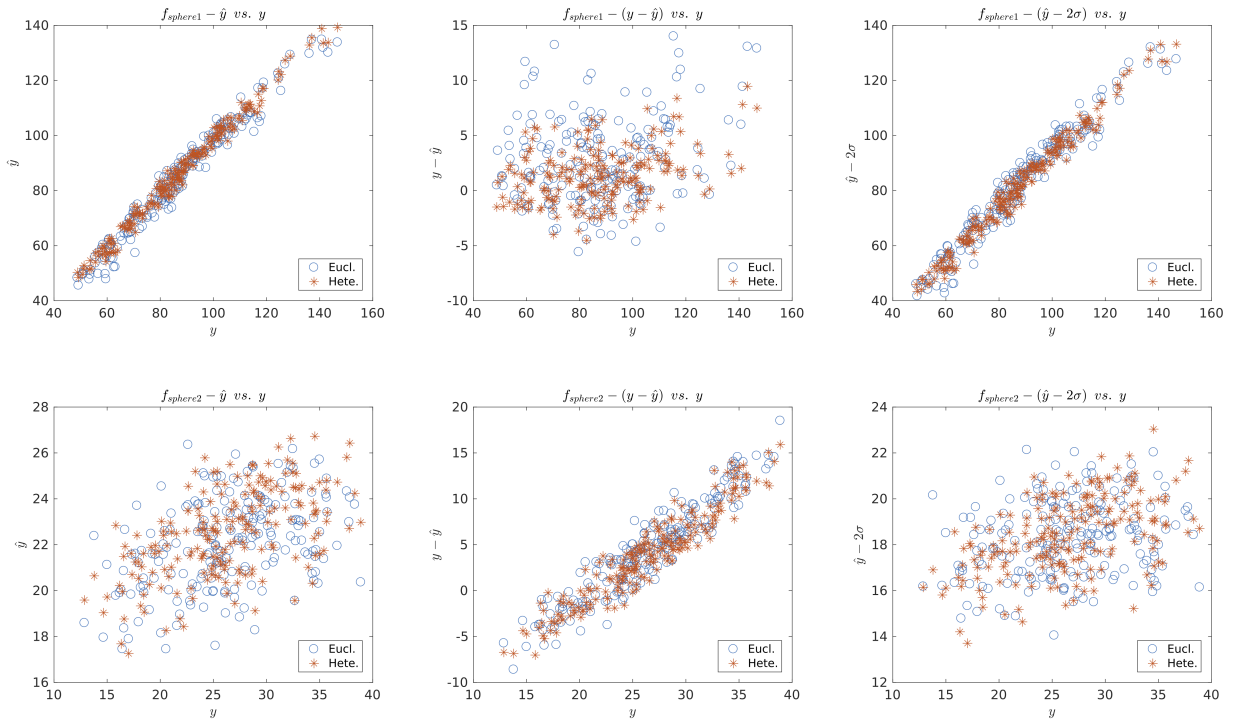


FIGURE 2. Comparison of predictions on  $f_{msphere}$  functions

## REFERENCES

- [1] A. Žilinskas and J. Mockus, *Avtomatica i Vychislitel'naya Teknika* **4**, 42–44 (1972).
- [2] J. Močkus, “On bayesian methods for seeking the extremum,” in *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974*, edited by G. I. Marchuk (Springer Berlin Heidelberg, Berlin, Heidelberg, 1975), pp. 400–404.
- [3] D. R. Jones, M. Schonlau, and W. J. Welch, *Journal of Global optimization* **13**, 455–492 (1998).
- [4] D. R. Wilson and T. R. Martinez, *Journal of artificial intelligence research* **6**, 1–34 (1997).
- [5] R. Li, M. T. M. Emmerich, J. Eggermont, E. G. P. Bovenkamp, T. Back, J. Dijkstra, and J. H. C. Reiber, “Metamodel-assisted mixed integer evolution strategies and their application to intravascular ultrasound image analysis,” in *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)* (2008), pp. 2764–2771.
- [6] R. Li, M. T. M. Emmerich, J. Eggermont, T. Bäck, M. Schütz, J. Dijkstra, and J. H. C. Reiber, *Evolutionary computation* **21**, 29–64 (2013).
- [7] J.-M. Yang and C.-Y. Kao, “An evolutionary algorithm for synthesizing optical thin-film designs,” in *Parallel Problem Solving from Nature — PPSN V*, edited by A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel (Springer Berlin Heidelberg, Berlin, Heidelberg, 1998), pp. 947–956.
- [8] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, *SIAM Journal on Optimization* **9**, 112–147 (1998), <https://doi.org/10.1137/S1052623496303470>.
- [9] M. Emmerich, K. Yang, A. Deutz, H. Wang, and C. M. Fonseca, in *Advances in Stochastic and Deterministic Global Optimization*, edited by P. M. Pardalos, A. Zhigljavsky, and J. Žilinskas (Springer, Berlin, Heidelberg, 2016), pp. 229–243.
- [10] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed. (Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989).
- [11] E. C. Garrido-Merchán and D. Hernández-Lobato, arXiv preprint [arXiv:1805.03463](https://arxiv.org/abs/1805.03463) (2018).