Noisy Multiobjective Black-Box Optimization using Bayesian Optimization

Hongyan Wang
Tsinghua University
Beijing, China
why17@mails.tsinghua.edu.cn

Hua Xu*
Tsinghua University
Beijing, China
xuhua@tsinghua.edu.cn

Yuan Yuan
Michigan State University
East Lansing, Michigan
yyuan@msu.edu

Junhui Deng
Tsinghua University
Beijing, China
deng@tsinghua.edu.cn

Xiaomin Sun
Tsinghua University
Beijing, China
sxm123@tsinghua.edu.cn

ABSTRACT

Expensive black-box problems are usually optimized by Bayesian Optimization (BO) since it can reduce evaluation costs via cheaper surrogates. The most popular model used in Bayesian Optimization is the Gaussian process (GP) whose posterior is based on a joint GP prior built by initial observations, so the posterior is also a Gaussian process. Observations are often not noise-free, so in most of these cases, a noisy transformation of the objective space is observed. Many single objective optimization algorithms have succeeded in extending efficient global optimization (EGO) to noisy circumstances, while ParEGO fails to consider noise. In order to deal with noisy expensive black-box problems, we extending ParEGO to noisy optimization according to adding a Gaussian noisy error while approximating the surrogate. We call it noisy-ParEGO and results of S-metric indicate that the algorithm works well on optimizing noisy expensive multiobjective black-box problems.

CCS CONCEPTS

• **Mathematics of computing** → **Bayesian computation**; • **Theory of computation** → *Evolutionary algorithms*; • **Computing methodologies** → *Optimization algorithms*; *Gaussian processes*;

KEYWORDS

black-box optimization, expensive multiobjective optimization, Gaussian Process, Gaussian noise, ParEGO

ACM Reference Format:
Hongyan Wang, Hua Xu, Yuan Yuan, Junhui Deng, and Xiaomin Sun. 2019. Noisy Multiobjective Black-Box Optimization using Bayesian Optimization. In *Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion), July 13–17, 2019, Prague, Czech Republic.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3319619.3321898

*Corresponding author.

GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07…$15.00
https://doi.org/10.1145/3319619.3321898

1 INTRODUCTION

Bayesian Optimization (BO)[7] is an effective method to optimize expensive black-box functions whose landscapes are unknown and evaluations are rather expensive. It usually builds a cheaper approximate GP surrogate for original black-box problem. Meanwhile, an acquisition function which balances exploitation and exploration is maximized to obtain new recommendations. Expected improvement (EI) function [5] is one of the most popular ones among these acquisition functions.

Efficient global optimization (EGO) [5] is an effective global objective optimization algorithm using the GP for single objective problems. Pareto EGO (ParEGO) [6] extends EGO to optimize multiobjective black-box problems. These two algorithms both assume noise-free observations which can be surrogated by smoothing, continuous approximation. While in many real problems, initial observations are often noisy. In these cases, a smooth continuous surrogate can no longer approximate true black-box problems correctly. By taking this problem into consideration, the sequential Kriging optimization (SKO) [4] and the sequential parameter optimization (SPO) [1] are proposed to optimize noisy single objective functions. However, as far as multiobjectives are concerned, ParEGO obviously fails to consider the noisy circumstances.

In this paper, we propose an extension of ParEGO to optimize multiobjective black-box problems with noisy observations. We use reinterpolation to build the GP surrogate to eliminate the misleading expected improvement caused by noise [3] in multiobjective black-box optimization problems. Results show that the algorithm become more stable if noise in observations is considered.

2 EXTENDING PAREGO TO NOISY MULTIOBJECTIVE BLACK-BOX PROBLEMS

Our algorithm begins with LHS sampling to generate initial solutions as in ParEGO [6]. At every iteration, only one of the weight vectors $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ is selected to calculate the augmented Tchebycheff function value $f_\lambda(x)$ which is named as scalar cost, and k is the dimension of objective space. The scalar cost is a measurement of solutions and a smaller scalar cost usually indicates a better Pareto optimal solution. Here, ρ is set to 0.05.

The key idea and the only difference between our algorithm and ParEGO is the DACE model building. In both algorithms, the scalar cost is regarded as objective value y used in EI function. However,

Algorithm 1 **Noisy-ParEGO Algorithm**

Latin Hpercube Sample to generate $xpop$
For $t = 1, 2, \ldots, T$ do
    Initialize normalized weight vector $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$
    Evaluate $xpop$: $f_\lambda(x) = \max_{i=1}^{k}(\lambda_i \cdot f_i(x)) + \rho \sum_{i=1}^{k} \lambda_i \cdot f_i(x)$
    Modelling: $y = f_{model}(x) + \epsilon \ where \ \epsilon \sim N(0, \sigma^2_{noise})$
    Sampling via maximizing $EIfunction$
    Update $xpop$ and Model
End for

Table 1: Mean of the S metric after 100 iterations

Problem	*noisy-ParEGO*	*ParEGO*
MOP2	**7.53519695E-02**	6.09992913E-02
MOP3	**1.48622930E+01**	1.38593883E+00
OKA1	3.78962194E+00	4.47665674E+00
OKA2	2.19196161E+00	3.65442752E+00
DTLZ1A	1.89938845E+04	2.45833250E+04
DTLZ2A	9.30368517E-01	9.54595053E-01

only scalar cost is considered in ParEGO which may cause deviation between the prediction and the true value. While in our paper, we suppose there is a Gaussian noise when predicting objective values, so we add the noise to the predictor. The predictor here is:

$$\hat{y}_{ri} = \mathbf{1}\hat{\mu} + k'(K + \sigma^2_{noise} I)^{-1}(y - \mathbf{1}\hat{\mu})$$

where

$$\hat{\mu} = \frac{\mathbf{1}^T(K + \sigma^2_{noise} I)^{-1} y}{\mathbf{1}^T(K + \sigma^2_{noise} I)^{-1}\mathbf{1}}$$

Meanwhile, the error is redefined to measure the uncertainty and therefore, it eliminates the error due to the noise [2]. We do not consider error in each objective dimension but a total error with regard to the scalar cost. The interpolating error is defined as:

$$\hat{s}^2_{ri} = \hat{\sigma}^2_{ri}\left[1 - k'K^{-1}k + \frac{(1 - \mathbf{1}'K^{-1}k)^2}{\mathbf{1}'K^{-1}\mathbf{1}}\right]$$

where

$$\hat{\sigma}^2_{ri} = \frac{(y - \mathbf{1}\hat{\mu})^T(K + \sigma^2_{noise} I)^{-1} K (K + \sigma^2_{noise} I)^{-1}(y - \mathbf{1}\hat{\mu})}{n}$$

The noise σ^2_{noise} is a regression constant and its value lies in the range [0.001,1] [2]. All parameters in the model including θ and the regression constant are obtained via maximum likelihood estimation. Then, recommendations are generated by maximizing the *EI* function to update the GP model. The whole iteration continues until the termination is met.

3 RESULTS AND ANALYSIS

In order to compare our algorithm with ParEGO, we run both algorithms 21 times on test problem MOP2, MOP3, OKA1, OKA2, DTLZ1A and DTLZ2A which are described in ParEGO [6]. S-metric (also known as Hypervolume, HV) [8] is used as the measurement. We calculate the HV mean and standard deviation (SD) of all runs to make clearer comparisons. Better results of our algorithm are marked in bold font.

Table 2: SD of the S metric after 100 iterations

Problem	*noisy-ParEGO*	*ParEGO*
MOP2	2.65094669E-02	2.22394704E-02
MOP3	2.42874655E+01	1.29445734E+00
OKA1	**2.65124976E+00**	3.15407394E+00
OKA2	**1.20951660E+00**	5.19219873E+00
DTLZ1A	**2.07737749E+04**	6.42568464E+04
DTLZ2A	**2.38271660E-01**	3.47269666E-01

From Table 1, it can be seen that when we add noise to original ParEGO, HV values change a lot. For problem MOP3 and DTLZ2A, two algorithm perform similarly. While for problem MOP3, our algorithm performs much better. As for the remaining ones, ParEGO performs better. Our algorithm achieves smaller SD on most problems which can be found from Table 2. It is concluded that by adding noise to ParEGO, especially when the number of objectives is larger than 3 (including 3), the algorithm performance can be improved a lot. When objective number is 2, ParEGO is more suitable for optimizing them. On the whole, by adding noises while optimizing multiobjective problems, algorithm can be more stable.

4 CONCLUSIONS

In cases of multiobjective black-box optimization, ParEGO fails to consider noise. While in fact, the sampling points are often with noise. An error also exists between the prediction and the true value, so we assume there is an Gaussian error while predicting. According to Gaussian reinterpolation, we approximate the objective space more accurately. Results of S-metric verifies that when considering the noise error, the multiobjective optimization model works better.

ACKNOWLEDGMENTS

This paper is founded by National Natural Science Foundation of China (Grant No: 61673235).

REFERENCES

[1] T. Bartz-Beielstein, C. W. G. Lasarczyk, and M. Preuss. 2005. Sequential parameter optimization. In *2005 IEEE Congress on Evolutionary Computation*, Vol. 1. 773–780 Vol.1. https://doi.org/10.1109/CEC.2005.1554761
[2] Alexander Forrester, Andras Sobester, and Andy Keane. 2008. *Engineering design via surrogate modelling: a practical guide*. Wiley. https://eprints.soton.ac.uk/64699/
[3] A. I. J. Forrester, A. J. Keane, and N. W. Bressloff. 2006. Design and Analysis of "Noisy" Computer Experiments. *AIAA Journal* 44 (Oct 2006), 2331–2339. https://doi.org/10.2514/1.20068
[4] D. Huang, T. T. Allen, W. I. Notz, and N. Zeng. 2006. Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models. *Journal of Global Optimization* 34, 3 (01 Mar 2006), 441–466. https://doi.org/10.1007/s10898-005-2454-3
[5] Donald R. Jones, Matthias Schonlau, and William J. Welch. 1998. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization* 13, 4 (01 Dec 1998), 455–492. https://doi.org/10.1023/A:1008306431147
[6] J. Knowles. 2006. ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation* 10, 1 (Feb 2006), 50–66. https://doi.org/10.1109/TEVC.2005.851274
[7] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas. 2016. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* 104, 1 (Jan 2016), 148–175. https://doi.org/10.1109/JPROC.2015.2494218
[8] Eckart Zitzler. 1999. Evolutionary algorithms for multiobjective optimization: methods and applications.