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

Kaifeng Yang, Koen van der Blom, Thomas Bäck and Michael Emmerich

Universiteit Leiden
The Netherlands

Outline:

- Background
- Algorithm
- Test problems
- Experiments
- Conclusions

Background:

- Multi-objective optimization problems (MOPs):

$$"min" \ (\boldsymbol{y}(\boldsymbol{x}^1), \boldsymbol{y}(\boldsymbol{x}^2), \dots, \boldsymbol{y}(\boldsymbol{x}^n))$$

$$\boldsymbol{y} = (y_1, y_2, \dots, y_d); \ \boldsymbol{x} \in \boldsymbol{\mathcal{X}}$$

- Bayesian Global Optimization (BGO) is efficient to optimize problems with expensive evaluations.
- The optimization problems are restricted to continuous problems.
  - In normal MOBGO, $\boldsymbol{\mathcal{X}} = \mathbb{R}^m$
  - what if $\boldsymbol{x} = (\boldsymbol{x}_r, \boldsymbol{x}_z, \boldsymbol{x}_d)$? More common in real applications/machine learning
- How to solve discrete (even mixed integer) optimization problems using BGO?

Algorithm – Bayesian Global Optimization

- BGO was proposed by Prof. Mockus and Prof. Žilinskas [1]
- Suppose:
  - Decision vectors: $\boldsymbol{X} = (\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(n)})^\top$
  - Objective values: $\boldsymbol{Y} = \left(\boldsymbol{y}(\boldsymbol{x}^{(1)}), \boldsymbol{y}(\boldsymbol{x}^{(2)}), \ldots, \boldsymbol{y}(\boldsymbol{x}^{(n)})\right)^\top$
- GP/Kriging assumes:
  - $Y(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \epsilon(\boldsymbol{x})$
  - Where $\mu(\boldsymbol{x})$ is the estimated mean value over all given sampled points
  - And $\epsilon(\boldsymbol{x}) \sim \mathcal{N}(0, \sigma^2)$
  - $Corr[\epsilon(\boldsymbol{x}), \epsilon(\boldsymbol{x}')] = R(\boldsymbol{x}, \boldsymbol{x}') = exp^{-\theta d(\boldsymbol{x}, \boldsymbol{x}')^2}$ R(.,.) is the correlation function
- The distance function $(d(.,.))$ utilizes ***Euclidean metric***:
  - $d(\boldsymbol{x}, \boldsymbol{x}') = \sqrt{\sum_{i=1}^{i=m} (x_i - x_i')^2}$ m dimensional search space

[1] Mockus,J.,Tiešis,V., Žilinskas.:The application of Bayesian methods for seeking the extremum. In: L. Dixon, G. Szego (eds.) Towards Global Optimization, vol. 2, pp. 117–131. North-Holland, Amsterdam (1978)

Algorithm – Mixed Integer MOBGO (1)

- ***Euclidean metric***
  - Assumes the continuity of an objective function
  - Only suitable for the continuous space (isotropic in every dimension of the search space)
  - Not suitable for a nominal discrete or an integer space
- Basic idea:
  - Use heterogeneous metric to calculate the distance function $d(.,.)$
- ***Heterogeneous metric***
  - Combine the distance of different type variables

$$d_h(\boldsymbol{x}, \boldsymbol{x}') = \sqrt{\sum_{i=1}^{i=n_r} \left(r_i - r_i{}'\right)^2 + \sum_{i=1}^{i=n_z} |z_i - z_i{}'| + \sum_{i=1}^{i=n_d} I\left(d_i \neq d_i'\right)}$$

where $I = 1$ if the statement $(d_i \neq d_i')$ is true; otherwise, $I = 0$.

Algorithm – Mixed Integer MOBGO (2)

- Mixed Integer MOBGO follows the structure of MOBGO
- Difference is how to build up the Kriging models:
  - Euclidean metric to calculate the distance function in normal MOBGO:
  - Heterogeneous metric to calculate the distance function in mixed integer MOBGO:
- Use EHVI as the infill criterion

Algorithm 1: MOBGO Algorithm

Input: Objective functions $\mathbf{y}$, initialization size μ, termination criterion T_c

Output: Pareto-front approximation $\mathcal{P}$

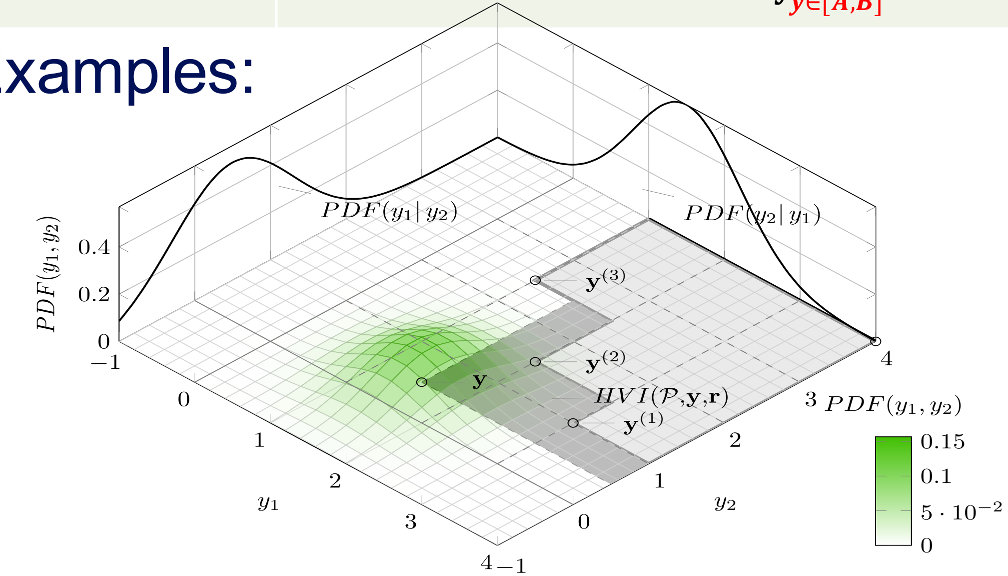
```
Initialize μ points (x^(1), ··· , x^(μ));
Evaluate the initial set of μ points: (y^(1) = y(x^(1)), ..., y^(μ) = y(x^(μ)));
Store (x^(1), ··· , x^(μ)) and (y^(1) = y(x^(1)), ..., y^(μ) = y(x^(μ))) in D:
  D = ((x^(1), y^(1)), ..., (x^(μ), y^(μ)));
Compute the non-dominated subset of D and store it in P ;
g = μ;
while g <= T_c do
    Train Kriging models M_1, ··· , M_d based on D ;
    Use an optimizer (opt) to find the promising point x* based on surrogate
      models M, with the infill criterion C;
    Update D: D = D ∪ (x*, y(x*));
    Update P as the non-dominated subset of D;
    g = g + 1;
Return P
```

Algorithm – Mixed Integer MOBGO (3)

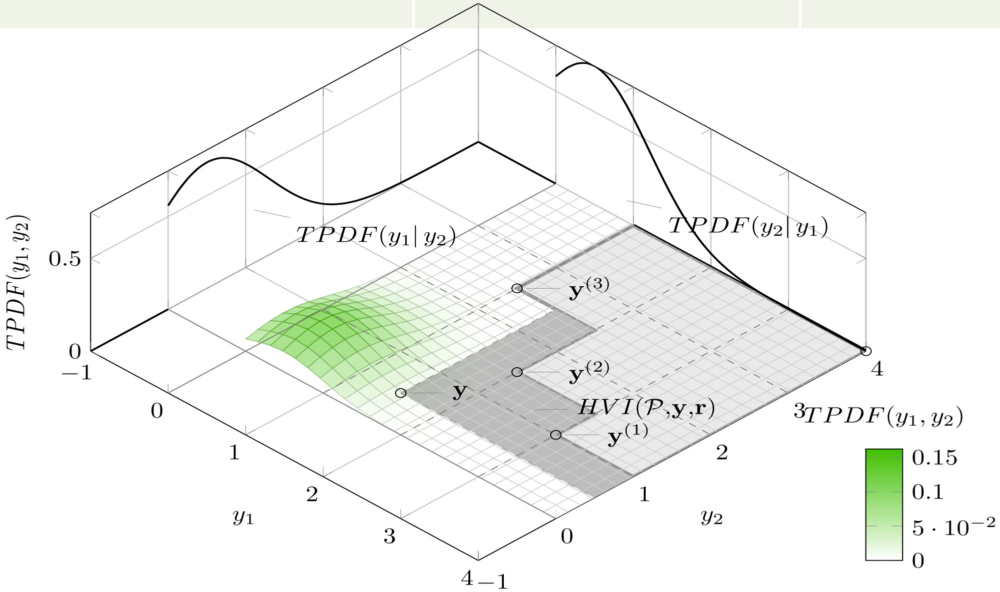
- Definition and computational complexity

	Definition	2-D cc.	3-D cc.
EHVI	$EHVI(\boldsymbol{\mu},\boldsymbol{\sigma},\mathcal{P},\boldsymbol{r}) = \int_{\boldsymbol{y}\in R^d} HVI(\mathcal{P},\boldsymbol{y})\times PDF_{\boldsymbol{\mu},\boldsymbol{\sigma}}(\boldsymbol{y})d\boldsymbol{y}$	$O(n\log n)$ [1]	$O(n\log n)$ [2]
TEHVI	$TEHVI(\boldsymbol{\mu},\boldsymbol{\sigma},\mathcal{P},\boldsymbol{r},\boldsymbol{A},\boldsymbol{B}) = \int_{\boldsymbol{y}\in[\boldsymbol{A},\boldsymbol{B}]} HVI(\mathcal{P},\boldsymbol{y})\times TPDF_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{y})d\boldsymbol{y}$	$O(n\log n)$ [3]	$O(n\log n)$

- Examples:



EHVI in 2-D



TEHVI in 2-D

[2] **M. Emmerich**, **K. Yang**, A. Deutz, H. Wang, C. M. Fonseca, A multicriteria generalization of bayesian global optimization, in: P. M. Pardalos, A. Zhigljavsky, J. Žilinskas (Eds.), Advances in Stochastic and Deterministic Global Optimization, Springer, Berlin, Heidelberg, 2016, pp. 229–243.

[3] **Yang, K.**, **Emmerich, M.**, Deutz, A., & Fonseca, C. M. (2017, March). Computing 3-D expected hypervolume improvement and related integrals in asymptotically optimal time. In *International Conference on Evolutionary Multi-Criterion Optimization* (pp. 685-700). Springer, Cham.

[4] **Yang, K.**, Deutz, A., Yang, Z., **Bäck**, **T.**, & **Emmerich, M**. (2016, July). Truncated expected hypervolume improvement: Exact computation and application. In *Evolutionary Computation (CEC), 2016 IEEE Congress on* (pp. 4350-4357). IEEE.

Test problems

- Three MOPs are tested:
- Sphere functions:
  - $f_{sphere_1}(\boldsymbol{r},\boldsymbol{z},\boldsymbol{d}) = \sum_{i=1}^{n_r} r_i^2 + \sum_{i=1}^{n_z} z_i^2 + \sum_{i=1}^{n_d} d_i^2$
  - $f_{sphere_2}(\boldsymbol{r},\boldsymbol{z},\boldsymbol{d}) = \sum_{i=1}^{n_r}(r_i - 2)^2 + \sum_{i=1}^{n_z}(z_i - 2)^2 + \sum_{i=1}^{n_d}(d_i - 2)^2$
- Barrier functions:
  - $f_{barrier_1}(\boldsymbol{r},\boldsymbol{z},\boldsymbol{d}) = \sum_{i=1}^{n_r}(r_i^2 + \alpha \sin(\mathrm{r_i})^2) + \sum_{i=1}^{n_z} A[z_i]^2 + \sum_{i=1}^{n_d} B_i[d_i]^2$
  - $f_{barrier_2}(\boldsymbol{r},\boldsymbol{z},\boldsymbol{d}) = \sum_{i=1}^{n_r}((r_i - 2)^2 + \alpha \sin(\mathrm{r_i} - 2)^2) + \sum_{i=1}^{n_z}(A[z_i] - 2)^2 + \sum_{i=1}^{n_d}(B_i[d_i] - 2)^2$
  - $\alpha = 1, A$ is generated by Algorithm 6 in Li et al. (2013) with $C = 75$
  - $B_i \in 1, \ldots, n_d$ is a set of n_d random permutations of the sequence 0,...,4
- Optical filter functions:
  - $f_{optfilt_1}$: a continuous variable is used in $f_{optfilt_1}$ if its corresponding binary variable is active; otherwise ignore it
  - $f_{optfilt_2}(\boldsymbol{r},\boldsymbol{d}) = \sum_{i=1}^{n_r} r_i d_i$
  - Penalty: If all bits are inactive, (250, 1250) is returned as the penalty.

[5] Li, R., Emmerich, M.T., Eggermont, J., Bäck, T., Schütz, M., Dijkstra, J. and Reiber, J.H., 2013. Mixed integer evolution strategies for parameter optimization. *Evolutionary computation*, *21*(1), pp.29-64.

Experiments – Parameter settings

- Hardware:
  - Intel(R) i7-4800mq CPU @ 2.70GHz, RAM 32GB
- Software:
  - OS: Ubuntu 16.04 LTS (64 bit)
  - Platform: MATLAB 8.4.0.150421 (R2014b), 64 bit
- MOBGO (mixed integer version and normal version)
  - Number of sampling points for the initialization ($N_{initial}$): 90
  - Number of function evaluations (N_{max}), including $N_{initial}$: 200
  - Optimal $\boldsymbol{\theta}$ strategy: simplex search method of Lagarias et al. (fminsearch) with max function evaluations of 1000
  - Infill criterion (C): Expected Hypervolume Improvement (EHVI)
  - Optimizer (opt): Genetic algorithm (MATLAB build-in function)
- Others
  - Repetitions: 10
  - Performance is evaluated by mean HV and the std. over 10 repetitions

Experiments – Results (1)

- Landscape of the $f_{mbarrier}$
  - For the visualization $n_r, n_z, n_d = 1$
  - Range: [0,4]
- Parameters:
  - Fixed $\boldsymbol{\theta}$ strategy, $\boldsymbol{\theta} = [1\ 1\ 1]\times 0.01$
  - Decision variables are normalized to [0,1] in Fig. 1
  - $N_{inital} = 15$ to build up the Kriging models
  - Plotted by 200 predictions/evaluations
- Results:
  - The landscapes by Heterogeneous metric are more accurate than using the Eucl. metric.
  - The first column is very analogous to the third column, but not the exactly the same

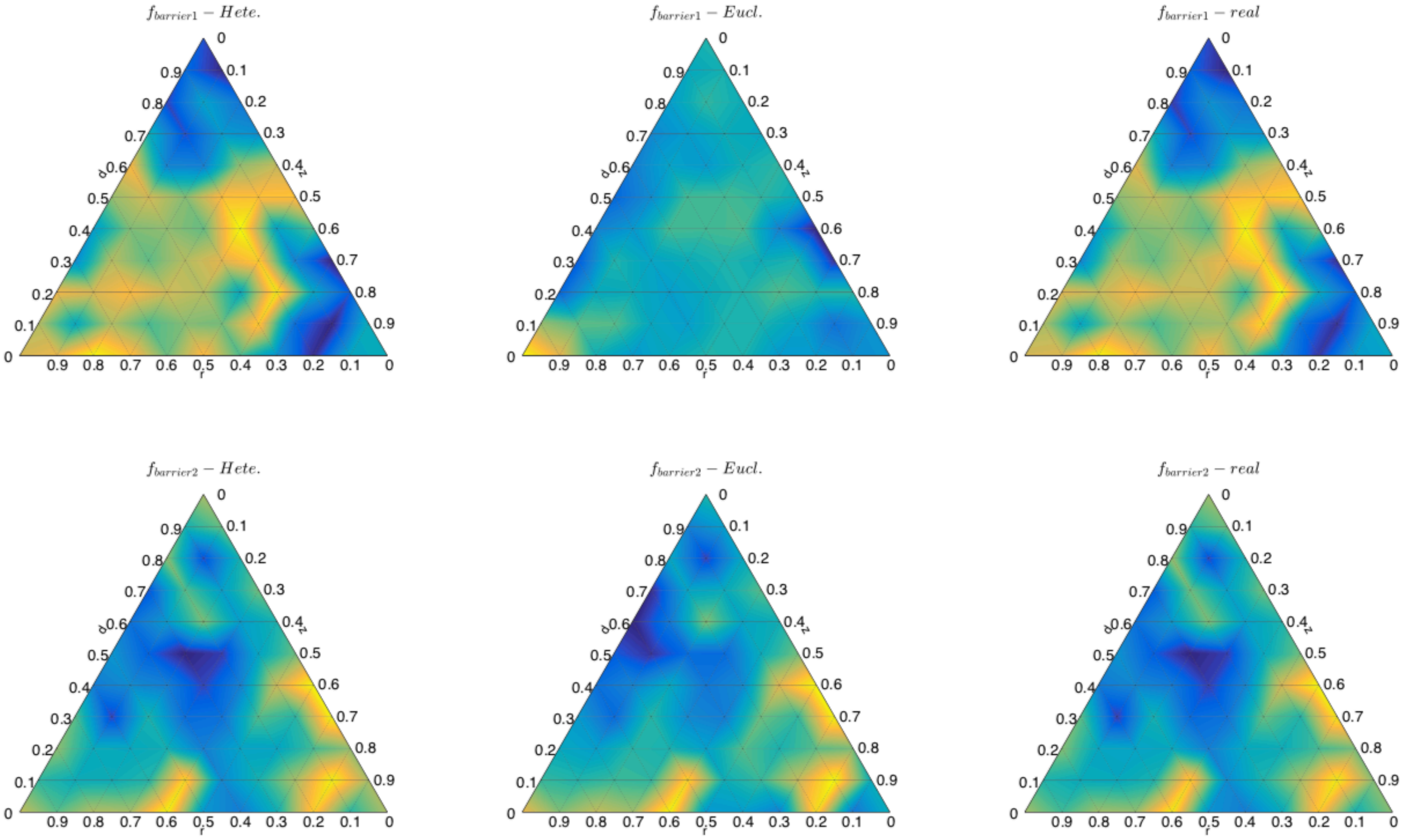


FIGURE 1. Landscape of the $f_{mbarrier}$ function

Experiments – Results (2)

- Comparison of predictions on $f_{msphere}$
- Parameters:
  - $n_r, n_z, n_d = 5$
  - Range: [0,4]
  - Optimal $\boldsymbol{\theta}$ strategy
  - $\hat{\boldsymbol{y}}$ and σ represent the predicted mean and std.
  - $N_{inital} = 90$ to build up the Kriging models
  - Plotted by 200 predictions/evaluations.
- Results:
  - Using the optimal $\boldsymbol{\theta}$ strategy, the Kriging of using the Hete. metric is still slightly better than that of using the Eucl. metric, w.r.t. mean and std.

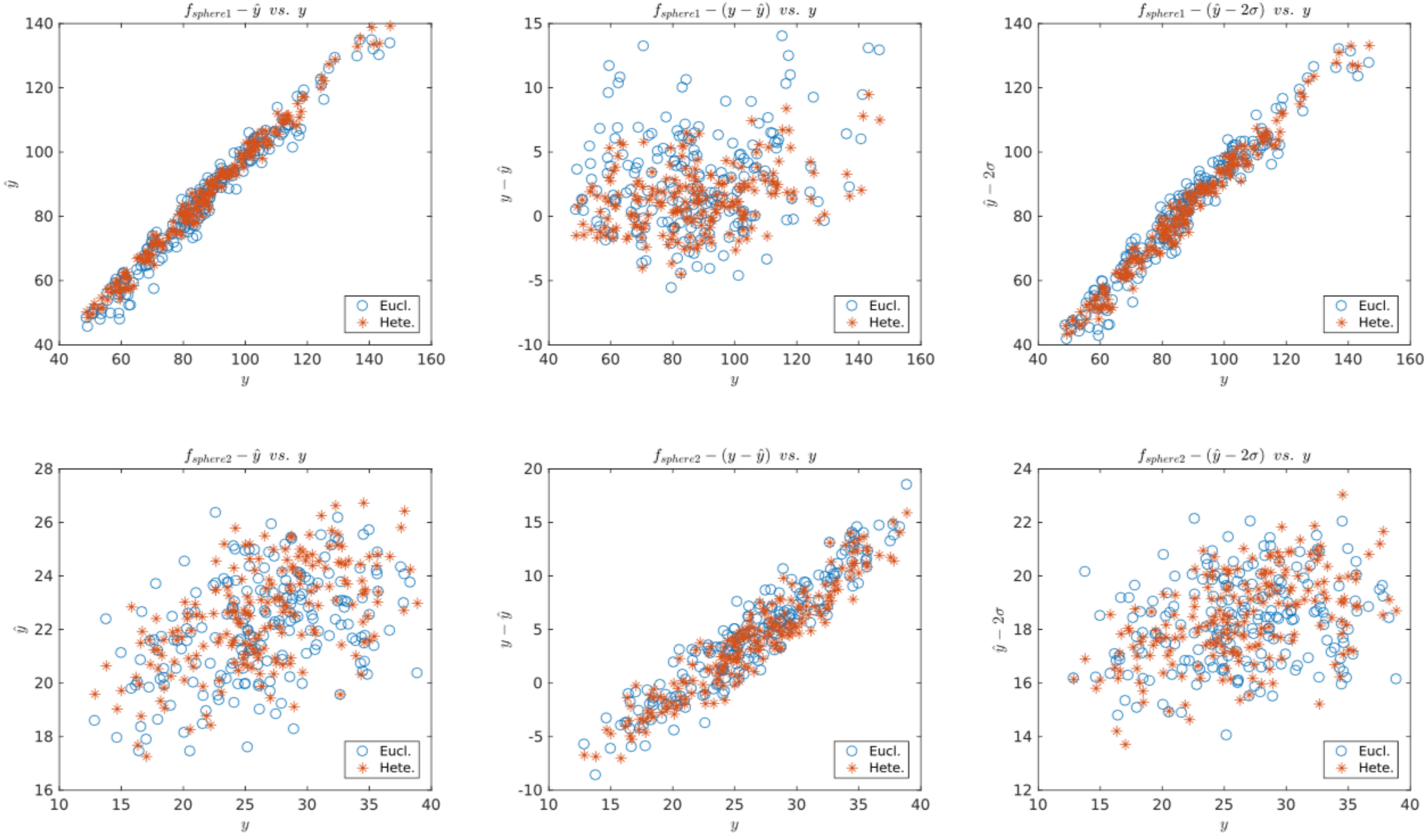


FIGURE 2. Comparison of predictions on $f_{msphere}$ functions

Experiments – Results (3)

TABLE 1. Parameter settings and empirical experimental results

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

- Sphere and Barrier problems:
  - The proposed algorithm outperforms the MOBGO algorithm using the Euclidean metric, w.r.t mean HV.
  - The proposed algorithm is more robust, w.r.t. the std.
- Optical filter problems:
  - The proposed algorithm is more robust, w.r.t. the std.
  - The MOBGO algorithm using the Eucl. metric is slightly better than the algorithm in this paper.

  MAYBE: $N_{max} = 200$ is too small for this problem, as the 14 variables are in this problem.

Conclusions and future work

- Conclusions:
  - Proposes a mixed integer MOBGO to solve the mixed integer multi-objective optimization problems
  - Achieved by calculating the distance function using the heterogeneous metric, instead of the Euclidean metric.
  - On sphere and barrier problems: outperforms the traditional MOBGO algorithm w.r.t mean HV and std.
  - On optical filter problem: more robust, but the traditional MOBGO performs slightly better, w.r.t. mean HV.

    Maybe can increase the N_{max}, as 14 variables in the search space
- Future work:
  - Compare the proposed algorithm with another integer-based BGO, using the one-hot encoding strategy.

References

- [1] Mockus,J.,Tiešis,V., Žilinskas.:The application of Bayesian methods for seeking the extremum. In: L. Dixon, G. Szego (eds.) Towards Global Optimization, vol. 2, pp. 117–131. North-Holland,Amsterdam (1978)
- [2] **M. Emmerich**, **K. Yang**, A. Deutz, H. Wang, C. M. Fonseca, A multicriteria generalization of bayesian global optimization, in: P. M. Pardalos, A. Zhigljavsky, J. Žilinskas(Eds.), Advances in Stochastic and Deterministic Global Optimization, Springer, Berlin, Heidelberg, 2016, pp. 229–243.
- [3] **Yang, K**., **Emmerich, M**., Deutz, A., & Fonseca, C. M. (2017, March). Computing 3-D expected hypervolume improvement and related integrals in asymptotically optimal time. In *International Conference on Evolutionary Multi-Criterion Optimization* (pp. 685-700). Springer, Cham.
- [4] **Yang, K.**, Deutz, A., Yang, Z., **Bäck**, **T**., & **Emmerich, M**. (2016, July). Truncated expected hypervolume improvement: Exact computation and application. In *Evolutionary Computation (CEC), 2016 IEEE Congress on* (pp. 4350-4357). IEEE.
- [5] Li, R., **Emmerich, M.T.**, Eggermont, J., **Bäck, T.**, Schütz, M., Dijkstra, J. and Reiber, J.H., 2013. Mixed integer evolution strategies for parameter optimization. *Evolutionary computation*, *21*(1), pp.29-64.

Thanks for your attention!
Questions and suggestions?