Mixed global optimization by algorithms composition: an empirical study with a focus on Bayesian approaches

Marie-Liesse Cauwet¹, Rodolphe Le Riche², Olivier Roustant²

¹ ESIEE, France ² CNRS LIMOS at Mines Saint-Etienne, France

> 23-26 June 2019 EURO 2019 conference UCD, Dublin, Ireland

ML Cauwet et al. (CNRS EMSE ESIEE)

Mixed global optimization

1/34

June 2019 1 / 34



Problem formulation and related work

- 3 Algorithms
 - Composition of algorithms
 - Benchmark on analytical functions
- 4 Bayesian optimization of mixed problems
- 5 Conclusions
- Bibliography

< ∃ >

Mixed optimization problems occur often

 Test case from OQUAIDO research chair [Roustant et al., 2018]: identify the mass of
 ²³⁹Pu in nuclear waste containers by gamma spectrometry. Parameters

Input	Domain
distance	[0,1] continuous
density	[0, 1] continuous
width	[0, 1] continuous
surface	[0, 1] continuous
energy	$\{1,2,3,4,5,6\}$ discrete ordinal
shape	sphere, cylinder, parallelepiped discrete nominal
chemical element	$\{1, 2, \dots, 94\}$ discrete nominal

- Design of systems with discrete material & architecture and continuous dimensions : aircraft wings, thermal insulation systems.
- Machine Learning (neural networks): ordinal number of neurons and layers, continuous weights.

• . . .



Problem formulation and related work

- 3 Algorithms
 - Composition of algorithms
 - Benchmark on analytical functions
- 4 Bayesian optimization of mixed problems
- 5 Conclusions
- Bibliography

Problem formulation

$$\begin{array}{l} \min_{x \in \mathbb{R}^{dr}, u \in \mathbb{D}^{do+dn}} f(x, u) \\ dr \ , \ do \ \text{and} \ dn \ \text{fixed} \\ \mathbb{D} \ \text{ordinal or nominal with levels} \ u_i \in \{1, \dots, L_i\} \ , \ i = 1, (do+dn) \\ dd \ = \ do+dn \end{array}$$

In scope: algorithms that work for a given number of nominal and ordinal and continuous variables, fewer than 100 levels per discrete variable.

Out of scope: specialized algorithms (e.g., for graphs, for convex problems,...), very large number of levels L_i (Bayesian methods).

ヘロト 人間ト 人目下 人口下

Mixed optimization algorithms (1/2)

- Relaxation of integer into continuous variables: branch & bound and MINLP solvers [Belotti et al., 2013], penalization for not being an integer [Shin et al., 1989], softmax reformulation in terms of probabilities with a stochastic (sampling) optimizer.
- Evolutionary optimization: integers as rounded off continuous variables [Lin et al., 2004, Hansen, 2011], composition of operators [Cao et al., 2000], EDA with mixed trees [Ocenasek and Schwarz, 2002].
- Alternating mixed programming with a user definition of the neighborhood of the discrete variables [Audet and Dennis Jr, 2001, Lucidi et al., 2005]

くほと くほと くほと

Mixed optimization algorithms (2/2)

- Model based optimization:
 - model of the function [Bartz-Beielstein and Zaefferer, 2017] [Holmström et al., 2008, Müller et al., 2013, Bajer and Holeňa, 2013];
 - model of the points distribution [Emmerich et al., 2008], [Sadowski et al., 2018].
- Mixed Bayesian optimization: for machine learning [Hutter et al., 2011, Wang et al., 2016], look at kernels in [Pelamatti et al., 2018], [Munoz Zuniga and Sinoquet, 2019].

This talk summarizes intermediate results that are part of an on-going effort to address mixed optimization with Gaussian Processes. May be, on the way, obtain results that apply to other mixed optimization problems.

Gaussian processes for mixed variables

- Use a(n infinite number of) surrogate(s) to the objective function, a Gaussian Process (GP), to reduce the number of calls to the true f().
- A GP depends on a kernel (covariance function) that decides on the type of surrogates we handle.
- GPs (kernels) exist for mixed variables. Here, use product of continuous and discrete kernels.
- GP model for discrete variables:
 - warping of ordinal variables using a normal function



• compound symmetry matrix for the nominal variables: simplest model.

June 2019 8 / 34

Bayesian optimization

BO algorithm template

Input a function f to minimize, a budget an optimization algorithm algo a design of experiment \mathcal{D} an associated Gaussian Process, GPcurrent_fbest $\leftarrow \min(f(\mathcal{D}))$ for k = 1 to budget do Get $(x', u') \in \arg \max_{x,u} EI(GP(x, u))$ using algo Calculate f(x', u') and add [(x', u'), f(x', u')] to \mathcal{D} current_fbest $\leftarrow \min(current_fbest, f(x', u'))$ Update the Gaussian Process end for

no call to f but still a mixed opt problem

< 回 > < 三 > < 三 >

- El(x, u) is a multi-modal mixed function.
 What is a good algorithm to maximize a mixed El?

 studied by composition of global optimizers.
- Is it useful to optimize *El* well? It is just an internal optimization problem



Problem formulation and related work

3 Algorithms

- Composition of algorithms
- Benchmark on analytical functions

4 Bayesian optimization of mixed problems

5 Conclusions

Bibliography

ML Cauwet et al. (CNRS EMSE ESIEE)

June 2019 11 / 34

Algorithms composition: principle

Generalizes operator composition in evolutionary computation.

 $\min_{x\in\mathbb{R}^{dr},u\in\mathbb{D}^{do+dn}}f(x,u)$

Input a function *f* to minimize an continuous optimization algorithm algoC a discrete optimization algorithm *algoD* while not done do $x' \leftarrow algoC$ $u' \leftarrow algoD$ Calculate f(x', u')Update algoC with $[x', f(x', u')] \triangleright$ noisy evaluation Update algoD with $[u', f(x', u')] \triangleright$ noisy evaluation end while

▲□▶ ▲圖▶ ▲ 臣▶ ▲ 臣▶ 三臣 - のへ⊙

June 2019

12 / 34

Component algorithms: Evolution Strategy (continuous)

Initialize state variables m, σ , C and others while termination criteria not satisfied **do** Sample λ individuals, following $\mathcal{N}(m, \sigma^2 C)$ Update the state variables according to the μ best sampled points end while

CMAES

[Hansen and Ostermeier, 2001]: Covariance matrix adaptation evolution strategy.

- State-of-the-art algo. in the noise-free case;
- the matrix might degenerate in case of noise.

pcCMSAES

[Hellwig and Beyer, 2016]: population control covariance self-adaptive evolution strategy

- adapted to the noisy setting;
- if noise is detected: the covariance matrix is fixed and the population λ increases.

Noise \approx "miscommunication" between the continuous (ES) and discrete optimizers.

Component algorithm: UMDA (discrete)

Estimation of Distribution Algorithms (EDAs): sampling using densities of independent categorical variables [Larrañaga and Lozano, 2001]

Univariate Marginal Distribution Algorithm (UMDA).

Input: α , ϵ Initialize probabilities of each variable having value j, p_i^j , i = 1, dd, $j = 1, L_i$ while termination criteria not satisfied **do** Sample λ individuals $u^{(k)} = (u_1^{(k)}, \ldots, u_{dd}^{(k)})$, with $u_i^{(k)} = j$ with proba. p_i^j Select the $\mu < \lambda$ best individals Update the marginal probability:

$$p_i^{j'} = \frac{1}{\mu} \sum_{k=1}^{\mu} \mathbb{1}(u_i^{k:\lambda} = j)$$

 $u_i^{k:\lambda}$ is the *i*th component of the k^{th} best individual.

$$p_i^j = \alpha p_i^j + (1 - \alpha) p_i^{j'} \in [\epsilon, 1 - \epsilon].$$

end while

Component algorithm: EA (discrete)

Evolutionary Algorithms (EA), actually here a random local hill-climber

 $\begin{array}{c|c} (1+1)\text{-EA.} & \blacktriangleright \text{ expected nb of mutations} \\ \hline \textbf{Input } \kappa & \blacktriangleright \text{ expected nb of mutations} \\ \textbf{Initialize Sample } u \text{ uniformly where each } u_i \text{ is in } \{1, \ldots, L_i\}, i = 1, dd \\ \textbf{while termination criteria not satisfied do} \\ u' \leftarrow u \\ \text{Shift each component of } u' \text{ with probability } \frac{\kappa}{n} \text{ to the previous or next level (circular).} \\ \textbf{if } f(u') \leq f(u) \text{ then} \\ u \leftarrow u' \\ \textbf{end if} \\ \textbf{end while} \end{array}$

イロン 不聞と 不同と 不同と

Algorithms composition: instances

- One-shot sampling:
 - Random Search;
- Hybrid: Evolution Strategy in the continuous part and Evolutionary Algorithms on the discrete part:
 - CMAES + EA;
 - CMAES + UMDA;
 - pcCMSAES + EA;
- Separate the impact of the random search on the continuous and the discrete part:

A B M A B M

16 / 34

- RS + EA;
- CMAES + RS.



Problem formulation and related work

3 Algorithms

- Composition of algorithms
- Benchmark on analytical functions

4 Bayesian optimization of mixed problems

5 Conclusions

Bibliography

ML Cauwet et al. (CNRS EMSE ESIEE)

Making mixed test functions

- Using functions of the BBOB/COCO testbed;
- Like in [Liao et al., 2014],
 - the ordinal variables are restrictions of the continuous variables to *I* given levels,
 - the nominal variables are a shuffling of the ordinal variables.



Ellipsoid in dim. 2. Left: 2 continuous variables. Center: one continuous variable, one categorial variable. Right: one continuous variable, one categorial variable.

ML Cauwet et al. (CNRS EMSE ESIEE)

June 2019 18 / 34

Test functions

• Functions: Sphere, Ackley, Ellipsoid, Griewank, Rastrigin;



Test functions

- Functions: Sphere, Ackley, Ellipsoid, Griewank, Rastrigin;
- L = 3 levels per discrete variable;
- Dimensions

dr	3	3	3	2	2	8	8	4	6	2	2	2	16	16	16
do	2	1	0	8	0	2	0	3	7	9	2	16	2	4	0
dn	0	1	2	0	8	0	2	3	7	9	16	2	2	0	4

- Number of function evaluations: 10^5 ;
- Median over 11 runs.

June 2019 19 / 34

★ ∃ →

Effect on Ackley

Simple Regret= $f(current_best) - f(real_opt.)$



Similar observations have been made on the other test functions

(日) (同) (三) (三)

Observations on analytic test functions

Observations also made on the other test functions:

- The CMAES variants perform best when there is a majority of continuous variables, in particular CMAES-EDA.
- As the number of discrete variables increases, pcCMSAES-EA becomes competitive, especially with nominal variables.
- CMAES-EA better for ordinal variables than CMAES-EDA and vice versa with nominal variables. CMAES-RS is a particularly low performer with discrete variables.
- Explanations: noise tolerance. EA more sensitive to ordering than UMDA.
- \Rightarrow does it apply to Bayesian optimization?

Context

Problem formulation and related work

3 Algorithms

- Composition of algorithms
- Benchmark on analytical functions

Bayesian optimization of mixed problems

5 Conclusions

Bibliography

ML Cauwet et al. (CNRS EMSE ESIEE)

Bayesian optimization (reminder)

BO algorithm template

Input a function f to minimize an optimization algorithm algo a design of experiment \mathcal{D} an associated Gaussian Process current_fbest $\leftarrow \min(f(\mathcal{D}))$ for k = 1 to 20 do Get $(x', u') \in \arg \max(EI)$ using algo \blacktriangleright no call to f but still a mixed opt problem Calculate f(x', u') and add (x', u'), f(x', u')) to \mathcal{D} current_fbest $\leftarrow \min(current_fbest, f(x', u'))$ Update the Gaussian Process end for

Monitor EI(x', u') and f(x', u')

イロト 不得下 イヨト イヨト 二日

El test cases

- Functions: Sphere, Ackley, Ellipsoid, Griewank, Rastrigin;
- Dimensions

Nb. continuous var.	Nb. ordinal var.	Nb. nominal var.
2	1	1
2	8	0
2	0	8
3	3	3
5	1	1
5	1	3
5	3	1
5	3	3
10	1	1

3

(日) (同) (三) (三)

• Functions: Sphere, Ackley, Ellipsoid, Griewank, Rastrigin;

•
$$-1 \le x_i \le 1$$
, $i = 1, dr$;

- Number of iterations: 20;
- Median over 21 runs;
- Number of evaluations within the El optimization sub-procedure: 1500.
- No update of the model.

Detailed case: El on the mixed sphere



dr = 10, do = 1, dr = 5, do = 1, dn = 3 dr = 2, do = 8, dn = 0

- Comparison possible only on the 1st iteration, where the El is the same for all the algos.
- RS-EA good when nb. discrete var \gg nb. continuous var.
- pcCMSAES-ES competitive with RS-EA when nb. discrete var ≫ nb. continuous var.
- ES variants better when nb. continuous var \gg nb. discrete var.

Effect on Bayesian optimization (sphere)



ML Cauwet et al. (CNRS EMSE ESIEE)

Mixed global optimization

26/34

June 2019

26 / 34

Effect on Bayesian optimization (Ackley)



- good algo. for max. EI \iff good SR
- Same observation for the other dr/do/dn/function cases

ML Cauwet et al. (CNRS EMSE ESIEE)

Mixed global optimization

27/34

June 2019 27 / 34

Effect on Bayesian optimization (all functions)

Average performance on all functions based on median simple regret after 20 iterations (over 11 runs).

Total of $5 \times 9 \times 11 = 495$ optimizations.



 $dr \gg do + dn$ $do + dn \approx dr$ $do + dn \gg dr$ • Case (i,j) display the % of time where algo. on column j outperform algo. on row i.

• pcCMSAES-EA is the most robust, RS and CMAES-RS the weakest, CMAES-UMDA the best when there are more continuous variables

Conclusions and further work

What is a good algorithm to maximize a mixed *E*/?

- It depends on the ratios dc/(do + dn) and do/dn.
- pcCMSAES-EA is robust: composing algorithms is like having noisy observations.

Is it useful to optimize El well?

• Yes! The quality in optimizing the (mixed) Expected Improvement is important to the performance of the Bayesian optimizer down the road.

Further work: better kernels for mixed variable Gaussian processes; associated acquisition criteria (new *El*'s) easier to optimize.

• • = • • = •

References I



Audet, C. and Dennis Jr, J. E. (2001).

Pattern search algorithms for mixed variable programming. *SIAM Journal on Optimization*, 11(3):573–594.



Bajer, L. and Holeňa, M. (2013).

Surrogate model for mixed-variables evolutionary optimization based on glm and rbf networks.

In International Conference on Current Trends in Theory and Practice of Computer Science, pages 481–490. Springer.



Bartz-Beielstein, T. and Zaefferer, M. (2017). Model-based methods for continuous and discrete global optimization. *Applied Soft Computing*, 55:154–167.



Belotti, P., Kirches, C., Leyffer, S., Linderoth, J., Luedtke, J., and Mahajan, A. (2013). Mixed-integer nonlinear optimization. *Acta Numerica*, 22:1–131.

Cao, Y., Jiang, L., and Wu, Q. (2000).

An evolutionary programming approach to mixed-variable optimization problems. *Applied Mathematical Modelling*, 24(12):931–942.



Emmerich, M. T., Li, R., Zhang, A., Flesch, I., and Lucas, P. (2008). Mixed-integer bayesian optimization utilizing a-priori knowledge on parameter dependences.

イロト イヨト イヨト イヨト

References II



Hansen, N. (2011). A CMA-ES for Mixed-Integer Nonlinear Optimization. Research Report RR-7751, INRIA.



Hansen, N. and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195.



Hellwig, M. and Beyer, H.-G. (2016).

Evolution Under Strong Noise: A Self-Adaptive Evolution Strategy Can Reach the Lower Performance Bound - The pcCMSA-ES.

In Handl, J., Hart, E., Lewis, P. R., López-Ibáñez, M., Ochoa, G., and Paechter, B., editors, *Parallel Problem Solving from Nature – PPSN XIV*, pages 26–36, Cham. Springer International Publishing.



Holmström, K., Quttineh, N.-H., and Edvall, M. M. (2008). An adaptive radial basis algorithm (ARBF) for expensive black-box mixed-integer constrained global optimization.

Optimization and Engineering, 9(4):311–339.

June 2019 31 / 34

References III

	ī

Hutter, F., Hoos, H. H., and Leyton-Brown, K. (2011). Sequential model-based optimization for general algorithm configuration. In *International Conference on Learning and Intelligent Optimization*, pages 507–523. Springer.



Larrañaga, P. and Lozano, J. A. (2001). *Estimation of distribution algorithms: A new tool for evolutionary computation*, volume 2. Springer Science & Business Media.



Liao, T., Socha, K., de Oca, M. A. M., Stützle, T., and Dorigo, M. (2014). Ant colony optimization for mixed-variable optimization problems. *IEEE Transactions on Evolutionary Computation*, 18(4):503–518.



Lin, Y.-C., Hwang, K.-S., and Wang, F.-S. (2004). A mixed-coding scheme of evolutionary algorithms to solve mixed-integer nonlinear programming problems.

Computers & Mathematics with Applications, 47(8-9):1295–1307.

Lucidi, S., Piccialli, V., and Sciandrone, M. (2005). An algorithm model for mixed variable programming. *SIAM Journal on Optimization*, 15(4):1057–1084.

June 2019 32 / 34

References IV



Müller, J., Shoemaker, C. A., and Piché, R. (2013).

So-mi: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems.

Computers & Operations Research, 40(5):1383–1400.

Munoz Zuniga, M. and Sinoquet, D. (2019).

Derivative free optimization with mixed discrete and continuous variables : motivations and limitations of the current methods.

EURO2019, 30th European Conference on Operational Research, Dublin, Ireland, 23-26 June 2019.

Ocenasek, J. and Schwarz, J. (2002).

Estimation of distribution algorithm for mixed continuous-discrete optimization problems. In 2nd Euro-International Symposium on Computational Intelligence, pages 227–232. IOS Press Kosice, Slovakia.



Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y. (2018). Efficient global optimization of constrained mixed variable problems. *Journal of Global Optimization*, pages 1–31.

Roustant, O., Padonou, E., Deville, Y., Clément, A., Perrin, G., Giorla, J., and Wynn, H. (2018).

Group kernels for gaussian process metamodels with categorical inputs. arXiv preprint arXiv:1802.02368.

イロト イポト イヨト イヨト

References V



Sadowski, K. L., Thierens, D., and Bosman, P. A. (2018).

Gambit: a parameterless model-based evolutionary algorithm for mixed-integer problems. *Evolutionary computation*, 26(1):117–143.



Shin, D. K., Gurdal, Z., and Griffin, O. (1989).
A penalty approach for nonlinear optimization with discrete design variables.
In *Discretization Methods and Structural Optimization—Procedures and Applications*, pages 326–334. Springer.



Wang, Z., Hutter, F., Zoghi, M., Matheson, D., and de Feitas, N. (2016). Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*, 55:361–387.

June 2019 34 / 34

- 4 同 6 4 日 6 4 日 6