Sequential Model Based Optimization with black-box constraints: feasibility determination via Machine Learning

Candelieri Antonio[1, a)] and Archetti Francesco[1,2 b)]
(Use the Microsoft Word template style: *Paper Author*)

[1]University of Milano-Bicocca, Department of Computer Science, Systems and Communications
[2]Consorzio Milano Ricerche

[a)] Corresponding author: antonio.candelieri@unimib.it
[b)] francesco.archetti@unimib.it

Abstract. This paper presents a Sequential Model Based Optimization framework for optimizing black-box expensive objective functions where feasibile search space is unknown a-priori. The framework is organized in two phases, the first uses Machine Learning (a Support Vector Machine classifier) to approximate the boundary of the feasible search space, the second uses standard Bayesian Optimization to perform efficient global optimization. With respect to the first phase, a specific acquisition function, to identify the next promising point to evaluate, has been proposed, dealing with the trade-off between improving the accuracy of the estimated feasible region and the possibility to discover possible disconnections of the actual feasible region. The main difference with standard Bayesian Optimization is that the optimization process is performed on the estimated feasibility region, only. Results on a set of 2D test functions proved that the proposed approach is more effective and efficient than standard Bayesian Optimization using a penalty for infeasibility.

INTRODUCTION

Sequential Model Based Optimization (SMBO), and more precisely Bayesian Optimization (BO), is a global optimization approach proved to be effective and efficient in the case of black-box expensive objective functions [1-7]. Indeed, it is currently the standard approach for solving automatic algorithm configuration and hyper-parameter tuning in the machine learning community [8,9]. While a surrogate model – usually a Gaussian Process (GP) – is used to approximate the objective function, an acquisition function provides the next promising point to evaluate, balancing between exploitation and exploration. Although BO has been used to solve global optimization problems in bounded-box search spaces, more interesting and realistic cases are related to Constrained Global Optimization (CGO) [10-12].

The most challenging setting arises when constraints are not known a priori and the – consequently unknown – feasible search space cannot be analytically expressed as bounds on the search variables. Relevant prior works on BO with unknown constraints proposed new acquisition functions, such as Integrated Expected Conditioned Improvement (IECI) [13] and a modification of the Predictive Entropy Search (PES) [14]. An interesting approach [15] uses GPs to model both the objective function and the different unknown constraints. However, it is based, in our opinion, on too strict assumptions: the number of constraints is known, and the objective and all constraints are independent, so that they can be modelled through independent GPs. Since violation of an unknown constraint in a real-world optimization problem (e.g. optimal control of a manufacturing system) could reflect in some disruptions of a physical system, another important work to consider is Safe BO [16], where optimization is performed while search space is expanded, "safely", by using information about the Lipschitz constant of the objective function. However, safe BO does not account for unknown constraints: the safe condition is only based on not exceeding a predefined threshold on the value of the objective function. Contribution of this work is to use Support Vector Machine (SVM) to estimate and model the overall unknown feasible search space (Ω), so that no assumptions on number of constraints and their independence are strictly required. A similar method has been proposed in [17], but our approach is organized in two phases: the first is aimed to provide a first estimation of Ω and the second is standard BO performed only on Ω. Preliminary results on a small set of test functions proved that the proposed approach is effective and efficient, even when a very limited number of function evaluations.

TWO-PHASE BAYESIAN OPTIMIZATION UNDER UNKNOWN CONSTRAINTS

Problem formulation

We start with the definition of the problem, which is:

$$\min_{x \in \Omega} f(x)$$

Where $f(x)$ is a black-box and expensive objective function and $\Omega \subset \mathrm{X} \subset \mathbb{R}^d$ is the feasible search space, even black-box, within the bounded-box search domain X. We introduce some notation that will be used in the following:

- $D_\Omega^n = \{(x^i, y^i)\}_{i=1,\ldots,n}$ is the **feasibility determination** dataset;
- $D_f^l = \{(x^i, f(x^i))\}_{i=1,\ldots,l}$ is the **function evaluations** dataset, with $l \leq n$ and where l is the number of feasible points out of the n evaluated so far;

where x^i is the *i*-th evaluated point and $y^i = \{+1, -1\}$ defines if x^i is feasible or infeasible, respectively.

Phase 1 – Feasibility determination

The first phase of the approach is aimed to find an estimation $\tilde{\Omega}$ of the actual feasible region Ω in *M* function evaluations ($\tilde{\Omega}^M = \tilde{\Omega}$). The sequence of function evaluations is determined according to an SMBO process where the surrogate model provides the currently estimated feasible region $\tilde{\Omega}^n$. As surrogate model we use an SVM classifier, trained on D_Ω^n. The SVM classifier uses an RBF kernel to model feasible regions with non-linear boundaries. Let denote with $h^n(x)$ the argument of the SVM-based classification function:

$$h^n(x) = \sum_{i=1}^{n_{SV}} \alpha_i y_i k(x_i, x) + b$$

where α_i and y_i are the Lagrangian coefficient and the "feasibility label" of the i-th support vector respectively, $k(.,.)$ is the kernel function (i.e., an RBF kernel, in this study), b is the offset and n_{SV} is the number of support vectors.

The boundaries of the estimated feasible region $\tilde{\Omega}^n$ are given by $h^n(x) = 0$ (i.e. non-linear separation hyperplane). The SVM-based classification function provides the estimated feasibility for any $x \in \mathrm{X}$:

$$\tilde{y} = \operatorname{sign}\left(h^n(x)\right) = \begin{cases} +1 \; if \; x \in \tilde{\Omega}^n \\ -1 \; if \; x \notin \tilde{\Omega}^n \end{cases}$$

With respect to the aim of the first phase, we propose an ***acquisition function*** aimed at identifying the next promising point according to two different goals:

- Improving accuracy of the boundaries of the estimated feasible region
- Discovering possible disconnections in the feasible region

To deal with the first goal, we use the distance from the boundaries of the currently estimated feasible region $\tilde{\Omega}^n$, using the following formula from the SVM classification theory:

$$d^n(h^n(x), x) = |h^n(x)| = \left| \sum_{i=1}^{n_{SV}} \alpha_i y_i k(x_i, x) + b \right|$$

To deal with the second goal, we introduce the concept of "coverage of the search space", defined by:

$$c^n(x) = \sum_{i=1}^{n} e^{-\frac{\|x^i - x\|^2}{2\sigma^2}}$$

So, $c^n(x)$ is a sum of *n* RBF functions centred on the points evaluated so far. The max of coverage is at evaluated points and decreases far from each of them, according to a Gaussian function parametrized on σ.

Finally, the phase 1 acquisition function is the sum of $d^n(h^n(x), x)$ and $c^n(x)$, and the next promising points is:

$$x^{n+1} = \underset{x \in X}{\operatorname{argmin}} \{d^n(h^n(x), x) + c^n(x)\}$$

Thus, we want to select the point associated to minimal coverage (i.e., max uncertainty) and minimal distance from the boundaries of the current estimated feasible region. This allows us to balance between improving accuracy of the boundaries and discovering possible disconnections of the actual feasible region (in less explored areas of the search space). It is important to highlight that, in phase 1, the optimization is on the overall bounded-box domain X.

After the function evaluation of the new point x^{n+1}, the following information is available:

$$y^{n+1} = \begin{cases} +1 \; if \; x \in \Omega \, ; \; with \; f(x^{n+1}) \\ \quad -1 \; if \; x \notin \Omega \end{cases}$$

and the following updates are performed:

- Feasibility determination dataset and estimated feasible region $\widetilde{\Omega}^{n+1}$

$$D_{\Omega}^{n+1} = D_{\Omega}^{n} \cup \{(x^{n+1}, y^{n+1})\}$$
$$h^{n+1}(x) \,|D_{\Omega}^{n+1}$$
$$n \leftarrow n + 1$$

- Only if $x \in \Omega$, function evaluations dataset

$$D_f^{l+1} = D_f^{l} \cup \{(x^{n+1}, f(x^{n+1}))\}$$
$$l \leftarrow l + 1$$

The SMBO process for phase 1 is repeated until $n = M$.

Phase 2 – Bayesian Optimization in the estimated feasible region

In this phase a standard BO process is performed, but with the following relevant differences:

- the search space is not a bounded-box but the estimated feasible region $\widetilde{\Omega}^n$ identified in phase 1
- the surrogate model – a GP – is fitted only using the feasible solutions observed so far, D_f^l
- the acquisition function – Lower Confidence Bound (LCB), in this study – is defined only on $\widetilde{\Omega}^n$

Thus, the next point to evaluate is given by:

$$x^{n+1} = \underset{x \in \widetilde{\Omega}^n}{\operatorname{argmin}} \{LCB^n(x) = \mu^n(x) - \gamma^n \, \sigma^n(x)\}$$

where $\mu^n(x)$ and $\sigma^n(x)$ are the mean and the standard deviation of the current GP-based surrogate model and $\gamma^n$is the inflate parameter to deal with the trade-off between exploration and exploitation for this phase. It can be set up adaptively or let constant (in this study it was set to 1). It is important to highlight that, contrary to phase 1, the acquisition function is here minimized on $\widetilde{\Omega}^n$, only, instead of the entire bounded-box search domain X.

The point x^{n+1} is just expected to be feasible, according to $\widetilde{\Omega}^n$, but the information on its actual feasibility is known only after evaluating $f(x^{n+1})$. Subsequently, the feasibility determination dataset is updated as follows:

$$D_{\Omega}^{n+1} = D_{\Omega}^{n} \cup \{(x^{n+1}, y^{n+1})\}$$

and according to the two alternative cases:

- x^{n+1} is actually feasible: $x^{n+1} \in \Omega$, $y^{n+1} = +1$;
  the function evaluations dataset is updated as follows: $D_f^{l+1} = D_f^{l} \cup \{(x^{l+1}, f(x^{l+1}))\}$, with $l \leq n$ is the number of the feasible solutions with respect to all the points observed so far. The current estimated feasible region $\widetilde{\Omega}^n$ can be considered accurate and retraining of the SVM classifier can be avoided: $\widetilde{\Omega}^{n+1} = \widetilde{\Omega}^n$
- x^{n+1} is actually infeasible: $x^{n+1} \notin \Omega$, $y^{n+1} = -1$;
  the estimated feasible region **must** be updated to reduce the risk for further infeasible evaluations

$$h^{n+1}(x) \mid D_f^{l+1} \implies \widetilde{\Omega}^{n+1}$$

The phase 2 continues until the overall available budget $n = N$ is reached.

EXPERIMENTAL SETTING

The approach has been validated on two well-known 2D test functions for CGO (i.e., Rosenbrock constrained to a disk and Mishra's Bird constrained). Since test functions are usually defined on a connected feasible region, we have also defined a supplementary test function: a Branin rescaled whose feasibility region is the area within two disconnected ellipses. The overall budget was fixed to 30 function evaluations: 10 for initialization (Latin Hypercube Sampling); 10 for feasibility estimation (phase 1) and 10 for BO constrained to estimated feasible region (phase 2). Although the budget is clearly not sufficient for solving CGO problems, our goal was to compare, on a very limited budget, our approach with traditional BO using penalties (in this case budget is divided in: 10 LHS evaluations for initialization and 20 for BO). For statistical significance, 30 different runs have been performed for every test function.

RESULTS AND CONCLUSIONS

The approach was able to identify a good estimation of the feasible search space, even for complex feasible search spaces consisting of disconnected regions. Seldom, the BO process, in phase 2, suggested infeasible points but, thanks to the updating of the estimated feasible region, it was able to quickly land near the optimum, anyway.

This preliminary study provides an effective approach for solving global optimization problems with unknown constraints. With respect to prior works it is not based on strict assumptions about the constraints and provides a good estimation of the feasible search space along with an optimal solution at the end of the optimization process even with a very limited budget for function evaluations. Since the approach tries to sample in the possibly infeasible region, it is well suited for expensive simulation-optimization tasks rather than system control/optimization, where safe (Bayesian) optimization is probably the better solution. Future works are focused on expanding the experimental setting and combining safe-exploration to make this approach suitable also for system control.

REFERENCES

1. D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions", *J Global Optim*, **13**(4), 455-492 (1998).
2. B. Shahriari, K. Swersky, Z. Wang, R. P. Adams and N. De Freitas, "Taking the human out of the loop: A review of Bayesian Optimization", in *Proceedings of the IEEE*, **104**(1), 148-175 (2016).
3. A. Žilinskas and J. Žilinskas, "Global optimization based on a statistical model and simplicial partitioning", *Comput Math Appl*, **44**(7), 957-967 (2002).
4. Y. D. Sergeyev, D. E. Kvasov and M. S. Mukhametzhanov, "On the efficiency of nature-inspired metaheuristics in expensive global optimization with limited budget". *Sci Rep-UK*, **8**(1), 453 (2018).
5. Y. D. Sergeyev, R. G. Strongin and D. Lera, *Introduction to global optimization exploiting space-filling curves*, (Springer Science & Business Media, 2013).
6. Y. D. Sergeyev and D. E. Kvasov, *Deterministic global optimization: An introduction to the diagonal approach* (Springer, 2017).
7. A. Zhigljavsky and A. Zilinskas, *Stochastic global optimization* **9** (Springer Science & Business Media, 2007).
8. C. Thornton, F. Hutter, H. H. Hoos, K. Leyton-Brown, "Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms", in *Proceedings of ACM-SIGKDD*, 847–855 (2013).
9. M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum and F. Hutter, "Efficient and robust automated machine learning", in *Adv Neur In*, 2962-2970 (2015).
10. Y. D. Sergeyev, *et al.*, "Index information algorithm with local tuning for solving multidimensional global optimization problems with multiextremal constraints", *Math Program*, **96**(3), 489-512 (2003).
11. R. Paulavičius and J. Žilinskas, "Advantages of simplicial partitioning for Lipschitz optimization problems with linear constraints", *Optim Lett*, **10**(2), 237-246 (2016).
12. R.G. Strongin, and Y. D. Sergeyev, *Global optimization with non-convex constraints: Sequential and parallel algorithms* **45** (Springer Science & Business Media, 2013).
13. J. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M Smith and M. West, "Optimization under unknown constraints", *Bayesian Statistics*, **9**(9), 229 (2011).
14. J. M. Hernández-Lobato, M. A. Gelbart, M. W. Hoffman, R. P. Adams and Z. Ghahramani, "Predictive entropy search for Bayesian Optimization with unknown constraints", in *Proceedings of the 32nd International Conference on Machine Learning*, **37** (2015).
15. M. A. Gelbart, J. Snoek and R. P. Adams, "Bayesian Optimization with unknown constraints", arXiv preprint arXiv:1403.5607 (2014).
16. Y. Sui, A. Gotovos, J. Burdick and A. Krause, "Safe exploration for optimization with Gaussian processes" in *International Conference on Machine Learning*, 997-1005 (2015).
17. A. Basudhar, C. Dribusch, S. Lacaze and S. Missoum, "Constrained efficient global optimization with support vector machines", *Struct Multidiscip O*, **46**(2), 201-221 (2012).