On the use of upper trust bounds in constrained Bayesian optimization infill criterion

R. Priem *
ONERA, DTIS, Université de Toulouse, Toulouse, France
ISAE-SUPAERO, Université de Toulouse, Toulouse, 31055 Cedex 4, France

N. Bartoli †
ONERA, DTIS, Université de Toulouse, Toulouse, France

Y. Diouane ‡
ISAE-SUPAERO, Université de Toulouse, Toulouse, 31055 Cedex 4, France

In order to handle constrained optimization problems with a large number of design variables, a new approach has been proposed to address constraints in a surrogate-based optimization framework. This approach focuses on sequential enrichment using adaptive surrogate models based on Bayesian optimization approach, and Gaussian process models. A constraints criterion using the uncertainty estimation of the Gaussian process models is introduced. Different evolutions of the algorithm, based on the accuracy of the constraints surrogate models, are used for selecting the infill sample points. The resulting algorithm has been tested on the well known modified Branin optimization problem.

Nomenclature

d	=	the number of design variables
m	=	the number of constraints
$\mathcal{X}^d$	=	the domain of the design variables, i.e., $\mathcal{X}^d \subset \mathbb{R}^d$
f	=	the objective function, i.e., $f : \mathbb{R}^d \mapsto \mathbb{R}$
c	=	the constraints function, i.e., $c : \mathbb{R}^d \mapsto \mathbb{R}^m$
D_l^h	=	the design of experiments of a given function $h : \mathbb{R}^d \mapsto \mathbb{R}$ using l points, i.e., $D_l^h = \{x_i, h(x_i)\}_{i=1,\cdots,l}$
μ^h	=	the mean of the Gaussian process of a given function $h : \mathbb{R}^d \mapsto \mathbb{R}$
σ^h	=	the standard deviation of the Gaussian process of a given function $h : \mathbb{R}^d \mapsto \mathbb{R}$
$\Phi(\cdot)$	=	the concentrated distribution function of the standard normal distribution
$\phi(\cdot)$	=	the probability distribution function of the standard normal distribution
τ^i	=	the constraint learning rate associated to the constraint c_i

I. Introduction

Aircraft design aims to create the best aircraft subject to constraints defined by customers, manufacturers or safety requirements. In general, the constraints are impacted by the involved disciplines in the design process, typically, structure, aerodynamic or electronic disciplines. Using analytic approaches, that takes into account all the disciplines involved, aircraft designers are able to conceive the best solution with respect to all the specified requirements. A similar process can be also applied on several kinds of aircraft designs, i.e., airplanes, helicopters, gliders or drones [1].

In the last decade, there has been a significant interest in improving the efficiency of aircraft design processes through the development of tools and techniques in the field of multidisciplinary optimization (MDO) [2]. Each of the connections between disciplines represents the influence of one on the other and makes the problem highly interdependent. These connections may create additional consistency constraints to manage, depending on the MDO

---

*PhD Candidate, Information Processing and Systems Department & Complexes Systems Engineering Department, remy.priem@onera.fr.
†Senior researcher, Information Processing and Systems Department, nathalie.bartoli@onera.fr, AIAA Member.
‡Associate Professor, Complexes Systems Engineering Department, youssef.diouane@isae-supaero.fr.

formulation chosen by the designer. In this way, MDO arises as a powerful tool that can perform this trade-off automatically.

Structure, propulsion, aerodynamic, electronic, automatic and safety are some of the numerous disciplines needed in aircraft design. Each of them has specific local variables and constraints which can be linked to other disciplines or managed by the user. The latter ones are called design variables and their number increases with the number of disciplines. In this way, the resulting MDO problem is a large-scale constrained optimization problem

$$
\begin{cases}
\min_{x \in \mathcal{X}^d} f(x) \\
\text{s.t. } c(x) \geq 0
\end{cases}
\tag{1}
$$

with $f : \mathbb{R}^d \mapsto \mathbb{R}$ the objective function, $c : \mathbb{R}^d \mapsto \mathbb{R}^m$ the m disciplinary and consistency constraints defined on $\mathcal{X}^d \subset \mathbb{R}^d$ the restricted domain of the possible aircraft configuration of d design variables.

The models used to represent the different disciplines [3] are often very accurate, meaning that a large number of design variables has to be managed during the optimization process. Moreover, the models can be very time consuming to evaluate and do not always provide the gradient of the objective and constraints functions. Thus, the classical optimization methods cannot be used as they are derivative based or need lot of model evaluations.

Bayesian Optimization (BO) [4] methods are proposed to solve expensive time consuming problems [3, 5–10]. The BO procedure uses only few objective function evaluations to solve the regarded optimization problems by means of Gaussian process (GP) models [11]. Currently, most of the BO methods are not adapted to handle large-scale constrained optimization problems, particularly this is due to the difficulty of estimating accurately the feasible domain. A possible solution was proposed in [3] where one uses the super efficient global optimization (SEGO) framework [12] and handles separately the objective and constraints functions in the optimization procedure using a dedicated criterion. The latter constraints criterion, although is practical, does not take into account the accuracy level of the constraints surrogate models during the estimation of the feasible domain. Inspired by [13], we propose to use an upper trust bound estimation of the GP models of the constraints to generalize existing feasibility criteria [3, 12]. The goal here is to improve the estimation (and hence the exploration) of the feasible domain.

This paper is organized as follows: we first remind the general constrained BO framework as well as the different existing strategies to handle the constraints. Then, we introduce the SEGO framework within the proposed feasibility criterion and its variants. Finally, the proposed criterion is tested on a well known constrained optimization problem.

II. Constrained bayesian optimization

Bayesian optimization (BO) framework is designed to solve expensive black box constrained optimization problems in wide domains of application, such as aircraft design [3, 14], fast neural network design [5], or hybrid electric vehicle design [12].

Starting from an initial design of experiments (DoE) using a first set of sample points chosen in design domain, BO builds cheap surrogate models, of the objective function f and constraints function c, by using GP models [11], also known as Kriging models [15]. The surrogate models are employed in an iterative enrichment process to find the optimum of the regarded constrained optimization problem. The enrichment process is led by a given infill criterion, it encompasses the search strategy of the global optimum (the balance between exploiting the surrogate models and exploring the domain of the design variables). The enrichment process is computationally inexpensive as it uses only the surrogate models information. At the end of the enrichment process, the point provided by maximizing a given infill criterion is evaluated using the true objective and constraints functions and included in the DoE in order to improve the surrogate models. The same process is repeated until a maximum number of iterations is reached.

The main steps of a BO framework are summarized in Algorithm 1. The two next subsections will describe the theory of GP as well as the different existing infill criteria in relation with the BO framework.

A. Gaussian processes

Gaussian processes [11] are one of the most popular surrogate models used in the context of aircraft design. Scalar GP models are fully defined by a mean function μ and a standard deviation function σ. The mean function describes the global behaviour of the GP whereas the standard deviation function depicts the influence of each sample on the entire domain. A detailed description of GP can be as follows. Let $h : \mathbb{R}^d \mapsto \mathbb{R}$ be a scalar function for which one tries to build a GP model using a DoE of l points $D^h = \{x_i, y_i^h\}_{i=1,\cdots,l}$ where $x_i \in \mathcal{X}^d \subset \mathbb{R}^d$ and $y_i^h = h(x_i) \in \mathbb{R}$ (for clarity

Input: Objective and constraints functions, an initial DoE, a maximum number of iterations max_nb_it;

1 for $l = 1$ **to** max_nb_it **do**
2 | Build the surrogate models using GP;
3 | Find x_l a point that maximizes a chosen infill criterion;
4 | Evaluate the objective and constraints functions at x_l;
5 | Update the DoE;
6 end
Output: Return the best point found in the DoE;

Algorithm 1: The Bayesian optimization framework.

reasons, we use the function h, in the context of our optimization problem h can represent the true objective function $h = f$ or a constraint function $h = c_i$ for a given $i \in \{1, \ldots, m\}$).

The GP model related to the function h using l sample points is family of functions defined by a mean function μ_l^h and standard deviation σ_l^h. Namely, at each point x of the bounded domain $\mathcal{X}^d \subset \mathbb{R}^d$, the GP model of h is defined with a multivariate Gaussian distribution $\mathcal{N}(\mu_l^h(x), \sigma_l^h(x))$. In the context of this paper, the mean μ_l^h and the standard deviation σ_l^h are given by

$$
\begin{aligned}
\mu_l^h(x) &= \hat{\mu}_l^h + k_l^h(x)^\top [K_l^h]^{-1} \left(Y_l^h - \hat{\mu}_l^h \mathbb{1}_l \right) \\
\sigma_l^h(x) &= \hat{\sigma}_l^h \sqrt{1 - k_l^h(x)^\top [K_l^h]^{-1} k_l^h(x)},
\end{aligned}
\tag{2}
$$

where $\mathbb{1}_l \in \mathbb{R}^l$ is a vector of only ones, $Y_l^h = [y_1^h, \ldots, y_l^h]^\top \in \mathbb{R}^l$, $k_l^h(x) = \left[k_l^h(x_1, x), \ldots, k_l^h(x_l, x) \right]^\top \in \mathbb{R}^l$, $K_l^h = \left[k_l^h(x_i, x_j) \right]_{i,j=1,\cdots,l} \in \mathbb{R}^{l \times l}$, $\hat{\mu}_l^h = \left(\mathbb{1}_l^\top [K_l^h]^{-1} \mathbb{1} \right)^{-1} \mathbb{1}_l^\top [K_l^h]^{-1} Y_l^h$, and $\hat{\sigma}_l^h = \frac{1}{l} \left(Y_l^h - \hat{\mu}_l^h \mathbb{1}_l \right)^\top [K_l^h]^{-1} \left(Y_l^h - \hat{\mu}_l^h \mathbb{1}_l \right)$. The definition of the kernel function $k_l^h(., .)$ depends on a set of hyper-parameters estimated by maximizing a likelihood function. Unfortunately, such maximization can be computationally challenging for large scale functions or with a large DoE. Practical approaches for estimating the hyper-parameters can be found in [16–18].

B. Infill criteria

BO optimization combines iteratively the surrogate model provided by the GP and an enrichment strategy. The latter strategy is driven by an infill criterion which can be formulated as an optimization problem. In fact, for a given iteration l, using the current DoEs (i.e., D_l^f and $D_l^{c_i}$ for $i = 1, \ldots, m$), one builds the GP models of the functions f and c_i for $i = 1, \ldots, m$. Let the associated mean and standard deviation functions be denoted by μ_l^f, $\mu_l^{c_i}$, σ_l^f, and $\sigma_l^{c_i}$ for $i = 1, \ldots, m$.

In an unconstrained optimization context, the BO framework is led by an acquisition function α_l^f modeling how much a given $x \in \mathcal{X}^D$ respects the trade-off between the exploration of unknown areas of the domain and exploitation of the model. By maximizing the acquisition function, most likely, one gets the most promising points, i.e.,

$$
x_l = \arg \max_{x \in \mathcal{X}^d} \alpha_l^f(x).
\tag{3}
$$

The objective function f is then evaluated at the new point x_l and added to the current DoE. The same process is repeated until convergence is detected or a maximum number of iterations max_nb_it is reached.

Several acquisition functions have been proposed in relation with the BO framework.

- **The Upper Confidence Bound (UCB)** acquisition function of [13] uses the variance of the GP to explore the uncertain areas of the domain. This is done by considering the best possible case given a confidence on the GP model. This criterion is explicit and is largely used for large-scale optimization problem [17–19]. It depends unfortunately on the user confidence on the model which may mislead the optimization process.
- **The Expected Improvement (EI)** acquisition function, EI_l^f, introduced first by [4], and used later in BO [20]. This acquisition function gives how much, in expectation, the point x_l improves the minimum of the surrogate model. EI is known to be highly multi-modal with flat areas, hence maximizing the EI can be hard using local optimization solvers.

- **The Watson and Barnes 2nd (WB2)** acquisition function [21] is proposed to improve the EI by shifting it with a the GP mean, i.e., $WB2_l^f = EI_l^f - \mu_l^f$, in order to avoid the flat areas on the landscape of the acquisition function. The maximization of the WB2 turns to be easier than handling the EI. However, this acquisition function may not allow exploration of the design domain.
- **The scaled Watson and Barnes 2nd (WB2S)** acquisition function [3] scaled the WB2 acquisition function by artificially increasing the influence of EI using a scaling factor s_l changing the $WB2S_l^f = s_l\, EI_l^f - \mu_l^f$. The factor s_l is chosen so that the maximum of the EI is higher than the corresponding mean function μ_l^f. This method removes the flat areas and also allows a better exploration of the design domain compared to WB2. All the acquisition functions EI, WB2, and WB2S are computed explicitly.
- **The Step-Wise Uncertainty Reduction (SUR)** acquisition function, proposed by [7], represents the expected volume reduction below the current minimum y_{min}^f in the DoE D_l^f. Such acquisition function has shown very efficient results when handling small size optimization problems. In fact, unlike all the previous acquisition functions (EI, WB2, and WB2S) which are given explicitly, the computation of the SUR acquisition function requires the estimation of a d-dimensional integral. For this reason, the SUR acquisition function is known to be not adapted for large scale optimization problems.
- **The Predictive Entropy Search (PES)** acquisition function [5] evaluates the gain of information on the global minimum location using the expected propagation [22]. Similarly to the SUR acquisition function, the PES is not straightforward to estimate and can be out of reach for large scale optimization problems.
- **The Max-Value Entropy Search (MES)** acquisition function, introduced by [6], is a recent variant of the PES for solving large scale optimization problems. The acquisition function requires only the estimation of a 1-dimensional integral.

To solve constrained optimization problems, the BO framework can be extended in the two following ways. The first way, is by using merit functions approaches α_l^m where ones combines the acquisition function (related to the objective function) and the constraints violation. Thus, the new enrichment point is computed by maximizing α_l^m all over the design domain, i.e.,

$$x_l = \arg\max_{x \in \mathcal{X}^d} \alpha_l^m(x). \tag{4}$$

Many examples of merit functions that extend the unconstrained acquisition functions are proposed, for instance, PESC [5], SUR [7], and EIC [23]. Other merit function approaches, such as ALBO [8], using the augmented Lagrangian approach are combined with the BO framework. All the cited merit function approaches are either implicit, hence they are computationally expensive to estimate, or explicit but not adapted to handle a large number of constraints. The second way of handling constraints, in the context of BO, is by considering the infill criterion as a constrained optimization:

$$x_l = \begin{cases} \arg\max_{x \in \mathcal{X}^d} \alpha_l^f(x) \\ \text{s. t. } \alpha_l^{c_i}(x) \geq 0, \ i = 1, \ldots, m \end{cases} \tag{5}$$

where α_l^f is a given acquisition function related to the objective function f and α_l^c is a feasibility criterion related to the constraints. The two main feasibility criteria are:
- **The Expected Violation (EV)** criterion [10] corresponds to the amount of c that violates, in expectation, the threshold. This criterion thus rules out the areas that are likely unfeasible or highly uncertain even if they can hide a feasible domain.
- **The Mean Constraint (MC)** criterion [12] directly uses the mean of the GP prediction as a feasibility criterion, i.e., $\mu_l^{c_i}$ for $i = 1, \ldots, m$. It has been successfully applied in aircraft design [3] and multi-disciplinary optimization [14]. This approach is known in the literature as Super Efficient Global Optimization (SEGO) framework.

Unlike most approaches which (e.g., PESC, SUR, EIC), although possibly efficient, are either based on estimated acquisition functions and/or not adapted to handle a large number of constraints, the main advantage of the SEGO framework is that the user is free to choose any acquisition function (related with the objective function). In the context of our paper, we are interested by the SEGO framework where WB2S is used as an acquisition function but combined with more efficient feasibility criterion. The next section will present an improvement of the SEGO framework by including uncertainties in the feasibility criterion.

III. An upper trust bound to handle the constraints

In the BO framework, the optimization is performed via GP models where the true objective and constraints functions are given with certain uncertainties depending on the quality of the DoE. Typically, a GP model trained with a large DoE is more accurate than a GP trained with a smaller one. In addition to the approximation of the targeted functions, GP models provide uncertainty estimation functions (given by σ^h in Eq. (2)). Those functions combined with the mean GP approximation allowed to obtain a more robust model approximation up to a confidence level. Namely, for a given function h, using an upper trust bound factor $\tau^h > 0$, the functions $\mu^h \pm \tau^h \sigma^h$ approximate the targeted function h with a level confidence related to τ^h, e.g., when $\tau^h = 3$ the confidence interval is expressed with 99% confidence. Outside of this zone, the value cannot be trusted as a reliable sample of the GP.

In the context of constrained BO, both the constraints and the objective functions are modeled with GPs. Inspired by [13], the inclusion of an upper trust bound in the SEGO framework, while modeling the constraints, can be very useful when one targets the global minimum of the original constrained optimization problem. In fact, the main drawback of the MC criterion (i.e., SEGO) is that the optimization process may be misled during the early stages due to possibly bad approximations of the constraints (as one uses only the mean information without including the uncertainty information given by the standard deviation functions). As example, Fig. 1 shows a constraint approximation during a given iteration of the SEGO framework from a DoE with 4 points. To ensure feasibility, the constraint must be higher than the constraint threshold (horizontal dotted green line) and the associated feasible domain is hatched. The GP prediction (in red color) is inaccurate, for all $x \in [0, 0.5]$, which indicates that no feasible point can be found in $[0, 0.5]$. Hence, the infill criterion Eq. (5) when using only μ^c will only provide enrichment points in the domain $[0.5, 1]$. In this way, the domain $[0, 0.5]$ will be never explored, which may hide the global optimum of the objective function, for instance, if one has $f : x \rightarrow (x - 0.1)^2$.
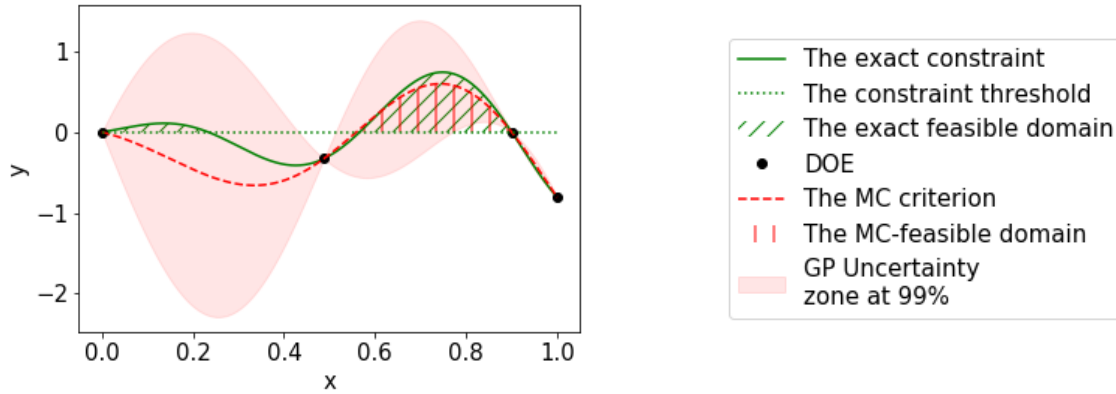

Fig. 1   Constraint approximation during a given iteration of the SEGO framework.

For a given iteration l of the SEGO framework, the introduced upper trust bound (UTB) will represent the highest trusted value of the constraint c_i (related to a given confidence level which is governed by τ_l^i), namely, for all $i \in \{1, \ldots, m\}$,

$$UTB_l^i(x) = \mu_l^{c_i}(x) + \tau_l^i \sigma_l^{c_i}(x). \tag{6}$$

Unlike the MC criterion, the use of the UTB criterion instead enlarges the regarded feasible domain during the first stages of the optimization process (when the DoE is small and the uncertainties are large) which allows a better exploration of the design domain and finds a potential global optimum. Figure 2 shows the trusted zone at 99% that can be found with a constraint learning rate τ equals to 3. The corresponding UTB is drawn with the blue line. The MC criterion (red line) cannot be trusted as the best constraint representation. The left part of the domain is MC-unfeasible whereas the true constraint is feasible. On the contrary, this domain is UTB-feasible with a constraint learning rate of 3. In what comes next, we note the obtained SEGO framework when using the UTB criterion by SEGO-UTB.

As far as the DoE is getting larger, the GP mean approximation gets more and more accurate. It is thus natural to trust more the GP prediction, by reducing the constraint learning rates τ^i, when the number of points of the DoE increases. The constraint learning rate is then decreased from its initial value to reach zero at the end of the optimization
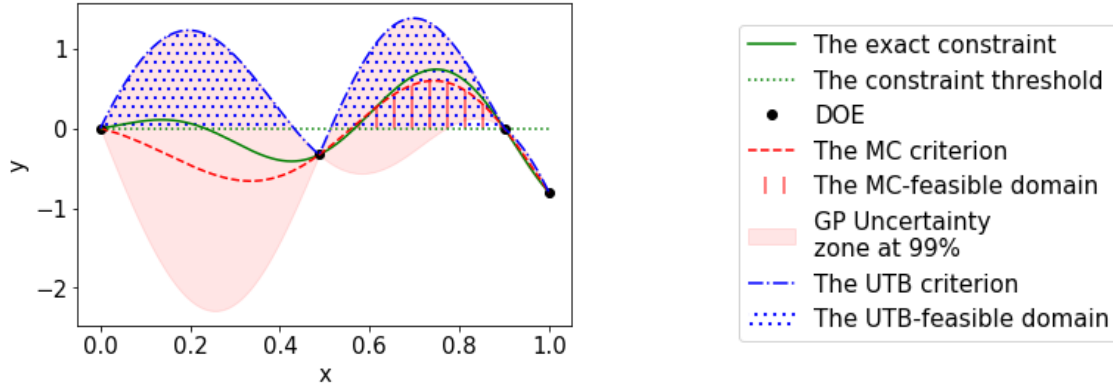
Fig. 2 MC versus UTB ($\tau = 3$) criterion.

process. In our case, five different evolution schemes are considered during the application of the algorithm, see Fig. 3: arc-tangent (Arc), constant, linear (Lin), logarithmic (Log), and exponential (Exp) evolution.
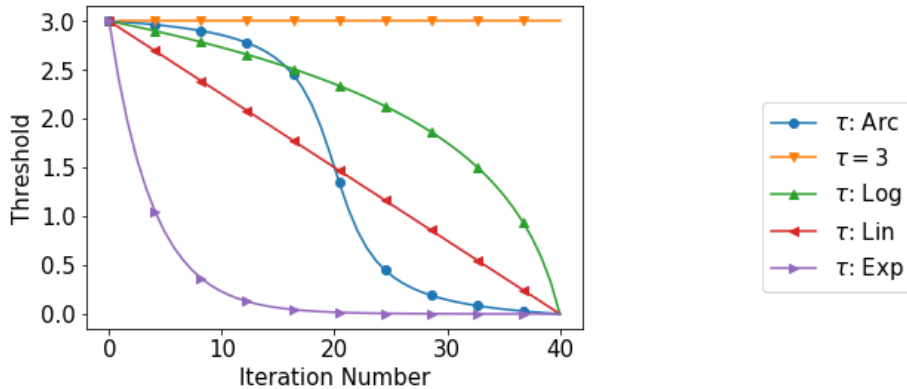

Fig. 3 Evolution of τ over the iterations during the application of SEGO-UTB.

IV. Results on the modified Branin test case

In this section, we perform the test on the 2-dimensional test problem, known as the modified Branin problem [24] with one inequality nonlinear constraint and $\mathcal{X}^d = [-5, 10] \times [0, 15]$, see Appendix.

As seen in Fig. 4, this problem is challenging because it has three disjoint feasible zones. The global optimum of this problem is $f^* = 12.005$. To perform the optimization procedure, GP with a constant mean function and a squared exponential covariance function is used. The hyper-parameters are updated via an optimization process after each enrichment of the DoE. An initial DoE of 5 points is generated for an optimization with a maximum budget of 80 evaluations. The problem is considered solved by a given algorithm if it reaches the global optimum value with a relative tolerance of $\epsilon_f = 10^{-3}$ and with a constraints violation of no more than $\epsilon_c = 10^{-4}$ (i.e., $|f(x^*) - f^*| < \epsilon_f |f^*|$ and $c(x^*) > -\epsilon_c$ where x^* is the obtained solution by the regarded algorithm). The SEGO framework is used with the WB2S acquisition function associated either to the MC criterion or to the five different evolutions (see Fig. 3) of the UTB constraints criterion. This optimization procedure is repeated 100 times with 100 different DoEs for each problem. For the comparisons, the DoE for each run is kept the same for the different infill criteria.

Table 1 displays the obtained results when using six different infill criteria combined withing the SEGO framework (listed in the first column). The second column of Table 1 corresponds to the number of runs that converge to an
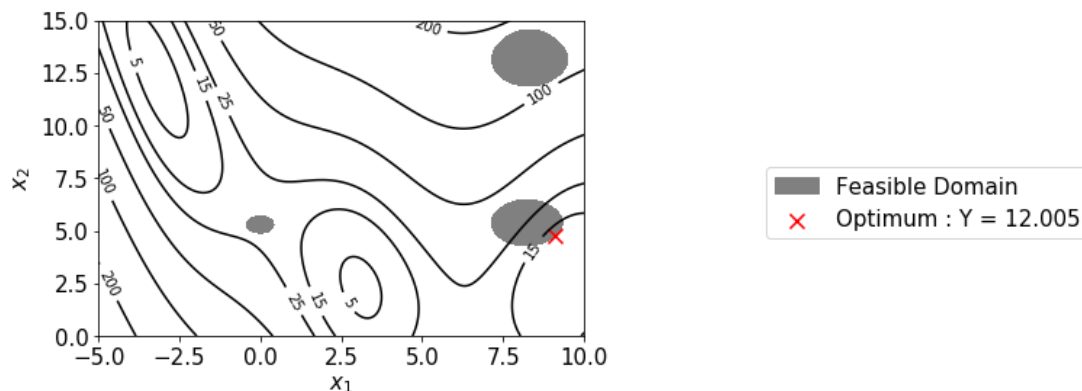
Fig. 4   Modified Branin test problem.

asymptotically feasible point (i.e., $c(x^*) > -\epsilon_c$) while the third column shows to the number of runs where the problem is solved (i.e., $|f(x^*) - f^*| < \epsilon_f (|f^*| + 1)$ and $c(x^*) > -\epsilon_c$). The last column reports the number of iterations required for a solver to reach the global optimum regarding only the solved runs (in average and with the associated standard deviation).

Table 1   Results of the modified Branin problem using different feasibility criteria (100 runs).

Solver	Nb. of feasible runs (out of 100)	Nb. of solved runs (out of 100)	Required iterations number	
			Average	Standard Deviation
SEGO	86	47	19	13
SEGO-UTB ($\tau = 3$)	100	100	30	4
SEGO-UTB (τ: Arc)	100	100	29	3
SEGO-UTB (τ: Log)	100	100	29	4
SEGO-UTB (τ: Lin)	100	98	27	4
SEGO-UTB (τ: Exp)	95	57	22	10

The MC criterion shows the worst convergence accuracy among the other five UTB criteria. In fact, the MC criterion is able to find the feasible domain for 86 runs out of 100, but only 47 run converges to the global minimum. Concerning the different evolutions for the UTB criterion, the exponential one, i.e., SEGO-UTB (τ: Exp), is the worst among all the UTB tested schemes (but still better than SEGO). This is due to the fast decreasing rate from the early stages of the optimization process. In this case, the exploration of the feasible domain is less encouraged, hence the exponential evolution tends rapidly to behave exactly as the MC criterion. The other UTB criteria exhibit a better efficiency compared to the MC criterion. In fact most runs (if not all) converge to the global optimum of the problem.

Figure 5 draws the average of the best valid objective value found at the lth iteration for the 100 runs. If the point is not feasible at the lth iteration, the best valid objective value is replaced by 200. Clearly, the two most efficient UTB criteria are SEGO-UTB ($\tau = 3$) and SEGO-UTB (τ: Arc), they both converged fast to the global optimum value (horizontal dotted red line). The MC and UTB (τ: Exp) feasibility criteria converged to a local optima of the modified Branin problem.

V. Conclusion

Aircraft design involved many disciplines, as structure or aerodynamic, that are inter-dependant and expensive to compute. The resulting multi-disciplinary optimization framework is thus defined with a large number of design variables and needs to respect lot of constraints. To handle this optimization problem, we chose an iterative enrichment strategy: the Bayesian optimization framework. This procedure builds cheap to evaluate Gaussian process models of the entire multi-disciplinary optimization workflow and uses especially the model uncertainty estimations in the enrichment
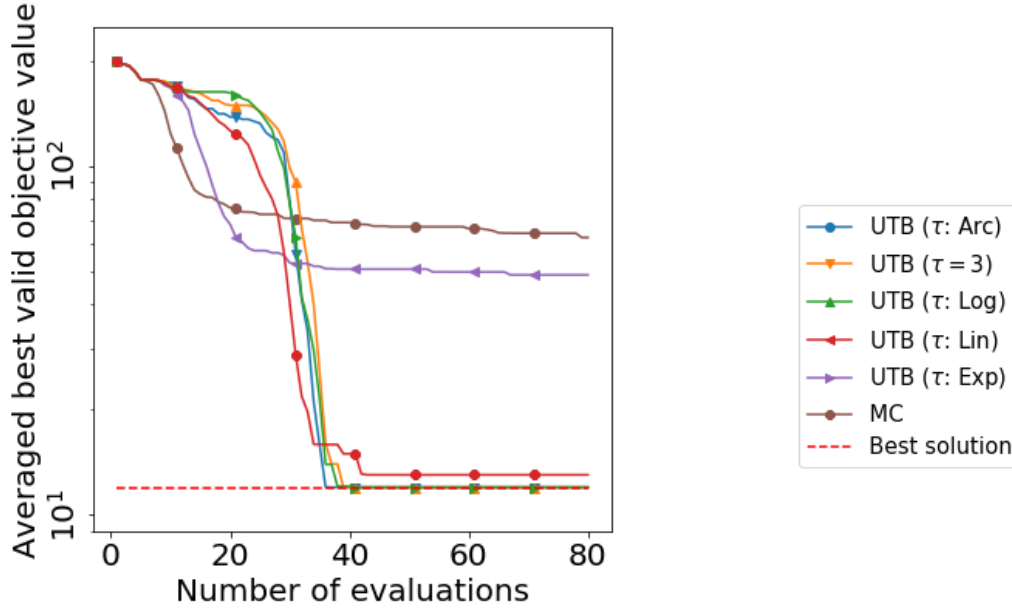
Fig. 5 Modified Branin convergence curve (100 runs).

process. An inner optimization problem is then defined to lead the enrichment process thanks to a trade-off between exploration and exploitation of the domain. We worked especially on the super efficient global optimization procedure that segregates the objective and constraints functions and takes into account explicit infill criteria that are adapted for large-scale constrained optimization problems.

The original super efficient global optimization algorithm uses only the prediction of the constraints model to detect the feasible domain at each enrichment step. The accuracy of the Gaussian process model is thus crucial for a good convergence rate of the algorithm. In this scope, we defined an explicit feasibility criterion named the upper trust bound criterion, it takes into account the accuracy of the surrogate model in the enrichment process by using the uncertainty estimation provided by the GP as well as a scaling constraint learning rate. The constraint learning rate reflects the confidence level on the surrogate model to represent the exact feasible domain. The upper trust bound feasibility criterion allows a better exploration of the feasible domain as it includes highly uncertain zones in the explored feasible domain. The proposed method has been initially tested on the modified Branin problem, the obtained results are very promising.

Acknowledgments

Appendix

Modified Branin function

The modified Branin problem [24] is given by:

$$f(x_1^*, x_2^*) = \begin{cases} \min\limits_{[x_1, x_2] \in \mathcal{X}^d} & f(x_1, x_2) \\ \\ \text{s. t.} & c\left(\dfrac{x_1 - 2.5}{7.5}, \dfrac{x_2 - 7.5}{7.5}\right) \geq 0 \end{cases},$$

where $\mathcal{X}^d = [-5, 10] \times [0, 15]$,

$$f(x_1, x_2) = \left(x_2 - 5.1\frac{x_1^2}{4\pi^2} + 5\frac{x_1}{\pi} - 6\right)^2 + 10\left[\left(1. - \frac{1}{8\pi}\right)\cos(x_1) + 1\right] + \frac{5x_1 + 25}{15}$$

and

$$c(x_{g1}, x_{g2}) = \left(4 - 2.1x_{g1}^2 + \frac{x_{g1}^4}{3}\right)x_{g1}^2 + x_{g1}x_{g2} + 4\left(x_{g2}^2 - 1\right)x_{g2}^2 + 3\sin\left(6\left(1 - x_{g1}\right)\right) + 3\sin\left(6\left(1 - x_{g2}\right)\right) - 6.$$

The solution of the modified Branin problem, with a constraint violation tolerance of $\epsilon_c = 10^{-4}$, is given by:

$$f(x_1^*, x_2^*) = 12.005, \quad x_1^* = 9.107, \quad x_2^* = 4.754$$

.

References

[1] Raymer, D. P., *Aircraft Design: A Conceptual Approach, Sixth Edition*, Virginia: American Institute of Aeronautics & Astronautics, 2018.

[2] Cramer, E., Dennis, J., J., Frank, P., Lewis, R., and Shubin, G., "Problem Formulation for Multidisciplinary Optimization," *SIAM Journal on Optimization*, Vol. 4, 1994, pp. 754–776.

[3] Bartoli, N., Lefebvre, T., Dubrueil, S., Olivanti, R., Priem, R., Bons, N., Martins, J. R. R. A., and Morlier, J., "Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design," *Aerospace Science and technology*, 2019. In press.

[4] Močkus, J., "On Bayesian Methods for Seeking the Extremum," *Optimization Techniques IFIP Technical Conference*, 1975, pp. 400–404.

[5] Hernández-Lobato, J. M., Gelbart, M. A., Adams, R. P., Hoffman, M. W., and Ghahramani, Z., "A General Framework for Constrained Bayesian Optimization Using Information-Based Search," *Journal of Machine Learning Research*, 2016, pp. 5549–5601.

[6] Wang, Z., and Jegelka, S., "Max-Value Entropy Search for Efficient Bayesian Optimization," *34th International Conference on Machine Learning*, Vol. 70, 2017, pp. 3627–3635.

[7] Picheny, V., "A Stepwise Uncertainty Reduction Approach to Constrained Global Optimization," *17th on Artificial Intelligence and Statistics*, Vol. 33, 2014, pp. 787–795.

[8] Gramacy, R. B., Gray, G. A., Le Digabel, S., Lee, H. K., Ranjan, P., Wells, G., and Wild, S. M., "Modeling an Augmented Lagrangian for Blackbox Constrained Optimization," *Technometrics*, Vol. 58, 2016, pp. 1–11.

[9] Keane, A. J., "Statistical Improvement Criteria for Use in Multiobjective Design Optimization," *AIAA journal*, Vol. 44, 2006, pp. 879–891.

[10] Audet, C., Dennis, J., Moore, D., Booker, A., and Frank, P., "A Surrogate-Model-Based Method for Constrained Optimization," *8th Symposium on Multidisciplinary Analysis and Optimization*, 2000, p. 4891.

[11] Rasmussen, C. E., and Williams, C. K. I., *Gaussian Processes for Machine Learning*, The MIT Press, 2006.

[12] Sasena, M. J., Papalambros, P., and Goovaerts, P., "Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization," *Engineering optimization*, Vol. 34, 2002, pp. 263–278.

[13] Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M., "Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design," *27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 1015–1022.

[14] Lefebvre, T., Bartoli, N., Dubreuil, S., Panzeri, M., Lombardi, R., D'Ippolito, R., Della Vecchia, P., Nicolosi, F., and Ciampa, P. D., "Methodological enhancements in MDO process investigated in the AGILE European project," *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2017, p. 4140.

[15] Krige, D. G., "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand," *Journal of the Southern African Institute of Mining and Metallurgy*, Vol. 52, 1951, pp. 119–139.

[16] Bouhlel, M. A., Bartoli, N., Otsmane, A., and Morlier, J., "Improving Kriging Surrogates of High-Dimensional Design Models by Partial Least Squares Dimension Reduction," *Structural and Multidisciplinary Optimization*, Vol. 53, 2016, pp. 935–952.

[17] Kandasamy, K., Schneider, J., and Póczos, B., "High Dimensional Bayesian Optimisation and Bandits via Additive Models," *32nd International Conference on Machine Learning*, 2015, pp. 295–304.

[18] Wang, Z., Hutter, F., Zoghi, M., Matheson, D., and de Feitas, N., "Bayesian Optimization in a Billion Dimensions via Random Embeddings," *Journal of Artificial Intelligence Research*, Vol. 55, 2016, pp. 361–387.

[19] Wang, Z., Li, C., Jegelka, S., and Kohli, P., "Batched High-Dimensional Bayesian Optimization via Structural Kernel Learning," *21th International Conference on Artificial Intelligence and Statistics*, Machine Learning Research, Vol. 84, 2018, pp. 745–754.

[20] Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global optimization*, Vol. 13, 1998, pp. 455–492.

[21] Watson, A. G., and Barnes, R. J., "Infill Sampling Criteria to Locate Extremes," *Mathematical Geology*, Vol. 27, 1995, pp. 589–608.

[22] Minka, T. P., "Expectation Propagation for Approximate Bayesian Inference," *7th Conference on Uncertainty in Artificial Intelligence*, 2001, pp. 362–369.

[23] Schonlau, M., Welch, W. J., and Jones, D. R., "Global versus Local Search in Constrained Optimization of Computer Models," *Lecture Notes-Monograph Series*, Vol. 34, 1998, pp. 11–25.

[24] Forrester, A., Sobester, A., and Keane, A., *Engineering Design via Surrogate Modelling: A Practical Guide*, John Wiley & Sons, 2008.