Multi-fidelity constrained Bayesian optimization, application to drone design

Rémy Charayron, Thierry Lefebvre, Nathalie Bartoli, Joseph Morlier

Multi-fidelity constrained Bayesian optimization, application to drone design

Rémy Charayron[1,2], Thierry Lefèbvre[1], Nathalie Bartoli[1], and Joseph Morlier[2,3]

[1] ONERA/DTIS, Université de Toulouse, Toulouse, France
[2] ISAE-SUPAERO, Toulouse, France
[3] ICA, Université de Toulouse, INSA, CNRS, MINES ALBI, UPS, Toulouse, France

Abstract. In aeronautics, the first design stages usually involve to solve a constrained multi-disciplinary optimization problem. The Bayesian optimization strategy is a way to solve such a complex system. This approach requires to evaluate the objective function and the constraints quite a few times. Evaluations are generally performed using numerical models that can be computationally expensive. To alleviate the overall optimization cost variable information sources can be used to make the evaluations. Typically we are dealing with cheap low fidelity models to explore the design space and expensive high fidelity models for exploitation. In the following work, a mono-fidelity Bayesian optimization method and its multi-fidelity counterpart are compared on two analytical test cases and on an aerostructural drone design constrained optimization problem. The multi-fidelity strategy allows to divide the computational cost by 1.3 compared to the mono-fidelity one on these test cases.

Keywords: Drone design · Multi-disciplinary optimization · Constrained Bayesian optimization · Variable fidelity · Surrogate models · Kriging · Gaussian process.

1 Introduction

In the first design steps, drone design optimization relies on multidisciplinary numerical models. These models capture the interactions between the different disciplines (aerodynamic, structure, operations, ...) which play a role in the drone overall performance. It follows that a single evaluation of the model is computationally expensive. Moreover, the complexity of the coupled system does not encourage us to take advantage of the analytical gradients. The high computational cost of a single evaluation implies that finite differences or complex step methods traditionally used in order to approximate the gradient can not be considered, hence classical gradient based optimization methods can not be used and the model is considered as a black-box function for which no information (regularity properties, derivative, ...) are available. Similarly the use of evolutionary optimization algorithms is not allowed due to the large number of function evaluations required. Then the focus is made on gradient-free surrogate-based optimization methods [12]. This involves to replace the initial model to optimize with a cheaper one, called metamodel or surrogate model. To construct this surrogate, gaussian processes (GP) [25] [19] interpolation framework also called kriging [21] is very powerful. Indeed, it allows not only to provide a prediction (which is the mean of the GP) but also the uncertainty of this prediction (the variance of the GP). In fact, the approach takes advantage of the gaussian vectors conditional distribution in order to determine the posterior distribution of the GP knowing some realizations of the initial model called design of experiments (DoE). This gives us a first surrogate model to approximate the initial model. Then it is improved via an iterative process that adds observations to the DoE according to a certain rule that tries to define the most interesting point to evaluate at each iteration doing a trade-off between exploitation and exploration. This rule is called the acquisition function and the whole process defined Bayesian optimization (BO) methods [13] whose first implementation was EGO in [16]. When the cost of evaluating the initial model is so important that even the previous BO method becomes intractable, multi-fidelity BO methods can be useful. These kind of methods use various levels of code, the highest fidelity (HF) code being the initial model and the lowest ones being some cheapest to evaluate approximations of the HF model. The benefit of multi-fidelity BO methods is that, depending on the case, the evaluation at the point to add to the DoE can be done using different codes: with a very precise but very expensive one or with less precise but less expensive codes. In Section 2 the kriging and Bayesian optimization methodology are

introduced. Section 3 focuses on the extension of this methodology to multi-fidelity. Section 4 and Section 5 present respectively some analytical test cases and a drone design case in order to illustrate and validate the proposed approach.

2 State of the Art

Let s the function defined in Eq. (1) that can only be evaluated in order to be optimized

$$\begin{aligned} s : \Omega \subset \mathbb{R}^d &\to \mathbb{R} \\ x &\mapsto y = s(x) \end{aligned} \tag{1}$$

2.1 Gaussian Processes interpolation / kriging method

The function s is considered to be the realization of a gaussian process $Z \sim GP(\mu, k)$ with prior mean $\mu : \Omega \to \mathbb{R}$ and covariance kernel $k : \Omega^2 \to \mathbb{R}$. The covariance kernel has $d+1$ hyperparameters $\theta = \{\sigma^2 = \theta_0, (\theta_i)_{i=1,\dots,d}\}$. The overall prior variance σ^2 is a scaling factor. The θ parameters are the correlation lengths in each direction. Let suppose that l observations of the function s are gathered in a DoE $D = \{x_k, y_k\}_{k=1,\dots,l}$ where $x_k \in \Omega$ and $y_k = s(x_k)$. The GP conditioned by D defines for each point $x \in \Omega$ a random variable y_x^D which follows a gaussian distribution

$$y_x^D \sim \mathbb{P}(y|(D,x)) = GP((\mu,k)|(D,x)) = \mathcal{N}(\hat{\mu}(x), (\hat{\sigma}(x))^2)$$

If no information on the GP mean μ is available, it is assumed to be unknown, then μ is supposed to be the zero constant function: we talk about ordinary kriging. Else, if there is a known trend in the data, it can be modeled using a deterministic basis of functions. The prior mean of the GP at a point x can be written as:

$$\mu(x) = \sum_{k=1,\dots,p} \beta_k f_k(x) \tag{2}$$

with f_k the k-th basis function and β_k the coefficient associated to the k-th basis function. In this case we talk about universal kriging. Lets denote $\boldsymbol{\mu} = \big(\mu(x_0) \dots \mu(x_l)\big)^T$, $\boldsymbol{k}(x) = \big(k(x_0,x) \dots k(x_l,x)\big)^T$, $\boldsymbol{Y} = \big(y_0 \dots y_l\big)^T$ and $\boldsymbol{K} = \Big(k(x_i,x_j)\Big)^T_{i,j=1,\dots,l} = \sigma R$ the covariance matrix on all the sampling points ($\boldsymbol{K}$ depends on θ) where R is the correlation matrix. Then using the gaussian vector conditional rule, it follows the subsequent expressions: $\begin{cases} \hat{\mu}(x) = \mu(x) + \boldsymbol{k}^T \boldsymbol{K}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) \\ \hat{\sigma}(x) = (k(x,x) - \boldsymbol{k}^T \boldsymbol{K}^{-1} \boldsymbol{k})^{\frac{1}{2}} \end{cases}$

The posterior mean $\hat{\mu}$ represents the surrogate model that approximates our function s. An indication on this surrogate model accuracy is given by the posterior standard deviation $\hat{\sigma}$. Note that the kriging kind of surrogate model has been selected specifically because it allows to have the variance expression on all the design space. This information is crucial using a Bayesian optimization strategy described in Section 2.2. To know the variance information using other kinds of surrogate models (radial basis function, neural networks, polynomial approximation, ...) would have required to perform a bootstrap method [11]. Therefore several metamodels should have been constructed in parallel to approximate the variance. Associated computation effort would have then led to an almost intractable method. GPs are parameterized by the kernel hyperparameters that need to be estimated. The maximum likelihood estimation method with a likelihood concentration process is used in this goal [23]. Dealing with high dimensional problems when using a kriging method raises additional difficulties. The number of hyperparameters increases with the dimension and their estimation is harder to optimize. One way to tackle these difficulties is the use of Partial Least Squares (PLS) [6] [22]. The PLS method finds a linear relationship between input variables and the output variable by projecting input variables onto a new space of lower dimension. The latent variables are linear combinations of the initial ones. KPLS defines a new covariance kernel which uses a lower number of hyperparameters [6] [7] and will be used in the Bayesian strategy described in the following..

2.2 Bayesian optimization

Bayesian optimization [13] is a global optimization strategy usually applied to optimize expensive to evaluate black-box functions. It consists in building a surrogate model of the objective function and then iteratively enriching this surrogate model with objective function evaluations to explore the design space and ensuring that the surrogate is precise enough in the optimal area.

Unconstrained BO Let the following unconstrained optimization problem:

$$x^* = \arg\min_{x \in \Omega} s(x) \quad \text{with } \Omega \subset \mathbb{R}^d \tag{3}$$

where $s : \Omega \to \mathbb{R}$ is the objective function introduced in Eq. (1). The efficient global optimization [16] called EGO constructs a GP of the objective function s using an initial DoE. Then the optimal solution is found by enriching iteratively the DoE and the GP. This enrichment is based on a trade-off, the exploration of the design space Ω and the exploitation of the GP model to find the minimum. The strategy involves the resolution of an optimization sub-problem to determine the next point to evaluate. This sub-problem is defined via an acquisition function $\alpha : \mathbb{R}^d \to \mathbb{R}$ to maximize

$$x_{next} = \arg\max_{x \in \Omega} \ \alpha(x) \tag{4}$$

There exists an extensive literature on acquisition functions [24] [13]. Some well known criteria are recalled in the following.

- Expected Improvement (EI): The EI computes the expected improvement of the current minimum value in the DoE by adding an evaluation.

  $$\alpha_{EI}(x) = EI(x) = \mathbb{E}(I(x)) = \mathbb{E}(\max(0, f_{\min} - y_x^D) \tag{5}$$

  where $f_{\min} = \min_{x \in X_D}(s(x))$ with X_D is the input set of the DoE D. In the case where y_x^D follows a gaussian law, $y_x^D \sim \mathcal{N}(\hat{\mu}(x), (\hat{\sigma}(x))^2)$, the $EI(x)$ criterion is analytical:

  $$EI(x) = \alpha_{EI}(x) = \begin{cases} 0 \ \text{ if } \ \hat{\sigma}(x) = 0 \\ (f_{\min} - \hat{\mu}(x))\Phi\left(\frac{f_{\min} - \hat{\mu}(x)}{\hat{\sigma}(x)}\right) + \hat{\sigma}(x)\phi\left(\frac{f_{\min} - \hat{\mu}(x)}{\hat{\sigma}(x)}\right) \ \text{ else} \end{cases} \tag{6}$$

  where Φ and ϕ represent respectively the $\mathcal{N}(0,1)$ cumulative distribution function and the probability density function. The first term of the expression is the exploitation term, it increases when $\hat{\mu}(x)$ decreases while the second term is the exploration term, it increases when the GP is not precise, ie when $\hat{\sigma}(x)$ is large.

- Watson and Barnes 2 ($WB2$): $WB2$ criterion [28] tries to regularize the EI by adding the mean $\hat{\mu}(x)$:

  $$\alpha_{WB2}(x) = \alpha_{EI}(x) + \hat{\mu}(x) \tag{7}$$

Appendix C shows the six first iterations of the BO process on a one dimensional test case using the EGO algorithm.

Constrained BO (CBO) Let the following constrained optimization problem:

$$x^* = \arg\min_{x \in \Omega} \ s(x) \quad \text{such that } g(x) \geq 0 \quad \text{and} \ \ h(x) = 0 \tag{8}$$

where the constraints are defined by

– $g : \mathbb{R}^d \to \mathbb{R}^m$ (m inequality constraints)
– $h : \mathbb{R}^d \to \mathbb{R}^p$ (p equality constraints)

The CBO algorithm is quite similar as the one of the unconstrained BO approach except that the optimization sub-problem solved to enrich the DoE takes into account the constraints. The associated sub-problem can take two forms: it can be unconstrained and tries to optimize an adapted function which gathers the constraints and the classical criterion [14]; or it can be constrained and optimizes one of the previous acquisition functions with some feasibility criteria associated to the constraints g and h [4]. Here the focus is made on constrained optimization sub-problem methods. The optimization sub-problem is of the form

$$x_{next} = \arg\max_{x \in \Omega} \alpha(x) \quad \text{with} \quad x \in \Omega_h \cap \Omega_g \tag{9}$$

where Ω_h and Ω_g are respectively the feasible domains defined by the two feasibility criteria: $\alpha_h : \mathbb{R}^d \to \mathbb{R}^p$ and $\alpha_g : \mathbb{R}^d \to \mathbb{R}^m$. To construct the feasibility criteria, the approaches named Super Efficient Global Optimization (SEGO) [27] and the Super Efficient Global Optimization coupled with Mixture Of Experts (SEGOMOE) [3] [2] [4] use the posterior means of the GPs that modelize the constraints as feasibility criterion: $\alpha_h = \hat{\mu}_h$ and $\alpha_g = \hat{\mu}_g$. The feasible domains are $\Omega_h = \{x, \alpha_h(x) = 0\}$ and $\Omega_g = \{x, \alpha_g(x) \geq 0\}$.

3 MFSEGO methodology

Using various information sources can be useful to alleviate the computation cost to build an accurate surrogate model or to perform an optimization. The SEGO type approaches are now extended to multi-fidelity and denoted in the following by MFSEGO.

3.1 Multi-fidelity kriging

Making assumptions in order to link the different fidelity levels is a way to simplify multi-fidelity problems. A discrepancy function δ that captures the difference between the high fidelity (HF) and low fidelity (LF) levels and a scaling factor ρ are considered in [17]

$$f_{HF}(x) = \rho f_{LF}(x) + \delta(x) \quad \text{such that} \quad f_{LF} \perp \delta \tag{10}$$

Le Gratiet [20] proposed to add the LF function to the basis function set $(h_i)_{i=1\ldots p}$ used in the universal kriging regression term (see Eq. (2)) to get:

$$\mu(x) = \sum_{i=1,\ldots,p} \Big(\beta_i h_i(x)\Big) + \beta_\rho f_{LF}(x) \tag{11}$$

β_ρ is an estimation of ρ done at the hyperparameters estimation step (see Section 2.1). Using a nested DoE structure: $D_{HF} \subseteq D_{LF}$, the independence between the high and low fidelity of the surrogate model is assumed. Then the HF surrogate model mean and variance can be expressed as:

$$\begin{cases} \hat{\mu}_{HF} = \rho \hat{\mu}_{LF} + \hat{\mu}_\delta \\ \hat{\sigma}^2_{HF} = \rho^2 \hat{\sigma}^2_{LF} + \hat{\sigma}^2_\delta \end{cases} \tag{12}$$

Le Gratiet's approach can then be extended to $L+1$ fidelity levels. Let us denote the fidelity levels $f_0, ..., f_L$ sorted from the lowest to the highest (we still consider a nested DoE structure: $D_L \subseteq D_{L-1} \subseteq \ldots \subseteq D_0$). The following recursive formulation can be written $\forall k = 1, ..., L$:

$$\begin{cases} \hat{\mu}_k = \rho_{k-1} \hat{\mu}_{k-1} + \hat{\mu}_{\delta_k} \\ \hat{\sigma}^2_k = \rho^2_{k-1} \hat{\sigma}^2_{k-1} + \hat{\sigma}^2_{\delta_k} \end{cases} \tag{13}$$

In this case, ρ is considered as a constant but it can depends on x. Then we have $\rho : x \mapsto \rho(x)$. This have been implemented in the toolbox SMT [8].

To learn the multi-fidelity model, the lowest fidelity level is learnt first, then the relationships (scaling factor ρ and discrepancy function δ) between every successive fidelity level are consecutively learnt. Since the variances can be expressed in closed form, the contribution of each fidelity level to the total variance of the multi-fidelity model can be deduced too. Denoting $\sigma^2_{cont}(k,x)$ the variance contribution of the k^{th} fidelity level at the point x, with the notation $\sigma^2_{\delta_0} = \sigma^2_0$ and assuming that $\prod_{L=k}^{L-1} \rho_j^2 = 1$, we have:

$$\sigma^2_{cont}(k,x) = \sigma^2_{\delta_k}(x) \prod_{j=k}^{L-1} \rho_j^2 \tag{14}$$

3.2 Multi-fidelity Bayesian optimization

With a multi-fidelity Bayesian optimization process, when a point is added to the DoE, not only the most promising point has to be decided, but also the fidelity level to which evaluate it. Splitting the problem of finding the point and the fidelity level in two successive steps has been proposed in [20]. First the point is found using a classical acquisition function as in the mono-fidelity Bayesian optimization (see Section 2.2). Then the variance contribution knowledge at each fidelity level gives some information to smartly decide the fidelity level to choose. The principal advantage of the multi-fidelity Kriging formulation presented in Section 3.1 lies in the fact that the variance contribution of each fidelity level can be known analytically. On the other side the main drawback is that it requires a nested DoE structure.
Let $c_0, ..., c_L$ be respectively the querying costs of all the fidelity levels $f_0, ..., f_L$. Let us denote $\sigma^2_{red}(k, x^*)$ the variance reduction of the high fidelity model when the point x^* is evaluated with all the fidelity levels $\leq k$

$$\sigma^2_{red}(k,x^*) = \sum_{i=0}^{k} \sigma^2_{\delta_i}(x^*) \prod_{j=i}^{L-1} \rho_j^2 \tag{15}$$

A criterion to choose the level of enrichment can be written as

$$t = \underset{k \in 0,...,L}{\arg\max} \ \frac{\sigma^2_{red}(k,x^*)}{(\sum_{i=0}^{k} c_i)^2} \tag{16}$$

This two step approach MFSEGO combining Eq. (9) and Eq. (16) is described in a pseudo-code in Appendix E and in Figure 5 of the Appendix F. It is now applied on different test cases and compared to SEGO.

4 Analytical cases

To start, the SEGO and MFSEGO methods have been confronted on two analytical test cases. The Branin and Sasena test cases are two different problems with a 2D objective function ($\Omega_{Branin} = [0;1]^2$) and $\Omega_{Sasena} = [0;5]^2$ respectively) with a single constraint. The cost ratio between high and low fidelity is arbitrarily fixed to $\frac{cost_{HF}}{cost_{LF}} = 5$. The tolerance on the constraints and the reference solutions for these two analytical test cases are detailed in Table 1. The HF and LF expressions of the objective function and constraint for the Branin and Sasena cases are given in Appendices A and B.

Table 1: Tolerance on the constraints and the reference solutions for the Branin and Sasena cases

	optimal objective value	ref_{sol}	ϵ	tol constraint
Branin	24.863	$(0.498, 0.401)$	0.5%	$1e^{-4}$
Sasena	-1.172	$(2.745, 2.352)$	0.5%	$1e^{-4}$

For all the test cases, 10 SEGO runs and 10 MFSEGO runs were made. Mono-fidelity and multi-fidelity runs share the same initial HF DoE. For the multi-fidelity runs, a LF DoE twice the size of the HF DoE is added. For all the optimization runs the squared exponential kernel with a constant trend is chosen to build

the kriging model. The size of the initial DoE and the budget are summed up in Table 3 for each test case. Given the reference solution, a convergence criterion for the mean (over the 10 runs) of the objective value at the best valid point $\bar{y}_{best}^{valid}$ is defined by:

$$\frac{\bar{y}_{best}^{valid} - f_{HF}(\text{ref}_{sol})}{f_{HF}(\text{ref}_{sol})} \leq \epsilon \tag{17}$$

For both mono and multi-fidelity strategies, the number of HF iterations and the total cost required to reach the convergence criterion (see Eq. (17)) are compared for each test case. Results are summed up in Tables 4 and 5 and convergence illustrations are given in Figure 3. Note that for the Branin case it is possible to go slightly below the reference solution because a tolerance on the constraint is considered. For Sasena test case, in the mono-fidelity case, the budget allocated to the simulation is not enough to reach the convergence criterion from Eq. (17). This issue can be explained by the fact that a run did not converge to the global minimum. On Figure 3(b) the median of the best valid value at each HF iteration is considered. Unlike the mean, the median converges even when the optimization is performed with mono-fidelity. Even if the convergence rate of the median is quite the same for mono and multi-fidelity, the multi-fidelity approach seems more robust. In the end, to reach the convergence criterion of the best valid objective mean value for the Branin and the Sasena test case, MFSEGO methodology allows to divide the SEGO total cost by 1.25 and 1.40 respectively.

5 Drone design test case

With the goal of designing a fixed-wing drone, a drone design test case is introduced. It relies on the K75 of Elecnor Deimos, a drone from specific class of about 80 kg MTOW and 35 kg payload mass available at ONERA and illustrated on Figure 1. The K75 delivered to ONERA in April 2018 is the first in the series, it is commercialized by Elecnor Deimos under the name D80-Titan [1, 5].

Fig. 1: Illustration of a K75 drone.

The focus is made on the aerostructure part which involves two disciplines, aerodynamic and structure. Two aerostructural models were developed for the K75 using OpenAeroStruct (OAS) [15] [10] [9], each one associated to a different discretization of the wing and tail meshes (see Table 2 and Figure 2). In our models, two flight points were considered: cruise flight (load factor = 1) and maneuvering (load factor = 9). The HF and the LF models have been evaluated on 200 points chosen randomly in the design space in order to make an estimation of the cost ratio: $\frac{cost_{HF}}{cost_{LF}} = 4.27$.

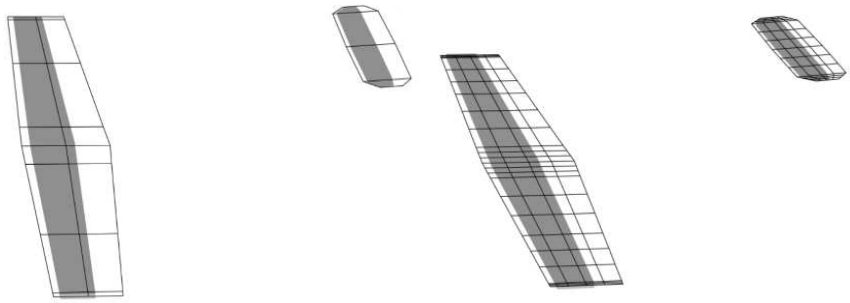
Table 2: HF and LF mesh dimensions

	HF wing mesh	HF tail mesh	LF wing mesh	LF tail mesh
Chordwise dim	5	5	3	2
Spanwise dim	25	13	9	5

The K75 problem to solve is the following:

$$\min_{x \in \Omega} \ \text{wing}_{\text{mass}}(x) \qquad x \in \mathbb{R}^{15}$$

$$\text{such that} \begin{cases} \text{CL}_{\text{cruise}} = 0.5 & \text{CL}_{\text{maneuver}} = 0.5 \\ \text{wing}_{\text{failure, cruise}} \leq 0 & \text{wing}_{\text{failure, maneuver}} \leq 0 \\ \text{tail}_{\text{failure, cruise}} \leq 0 & \text{tail}_{\text{failure, maneuver}} \leq 0 \\ \text{wingbox}_{\text{volume}} - \text{fuel}_{\text{volume}} > 0 & \frac{\text{fuelmass} - \text{fuelburn}}{\text{fuelmass}} = 0 \end{cases}$$



(a) K75 LF mesh (b) K75 HF mesh

Fig. 2: LF and HF meshes of the K75

The objective function is the mass of the wing. The 15 design variables and their associated bounds are summarized in Table 6 in Appendix D. Eight constraints are considered: two constraints to ensure a certain lift coefficient value in cruising flight as in maneuvering, four failure constraints to ensure that neither the wing nor the tail will break, whether in cruising flight or during maneuvering. These constraints compare the Von mises stress to the yield stress divided by a fixed coefficient (chosen here equal to 2.5) that acts as a safety margin. To simplify the optimization problem, the individual nodal failure constraints are aggregated using a Kreisselmeier-Steinhauser (KS) function [29]. Next, a constraint that ensures that the wingbox has enough internal volume for the fuel and a constraint that ensures that the fuel burn is equal to the fuel mass are added. The tolerances on these constraints are fixed to 10^{-3} except for the failure constraints for which the tolerances are fixed to 10^{-7}. Due to the high number of design variables for the K75 case, the kriging with PLS method introduced in Section 2.1 and its multi-fidelity version are used with 3 latent variables. For each test case, a gradient based optimization algorithm called Sequential Least SQuares Programming (SLSQP) [18] is used to solve the enrichment optimization sub-problem from Eq. (9) with the WB2 acquisition function (see Eq. (7)). The reference value used in the K75 case, is the optimal value obtained by solving the same optimization problem with the SLSQP algorithm on the HF mesh. It is equals to 12.4245 and $\epsilon = 0.5\%$ in this case too. In the end, to reach the convergence criterion of the best valid objective mean value for the K75 test case, MFSEGO methodology allows to divide the SEGO total cost by 1.32.

K75 results are sum up in Tables 4 and 5 and the convergence is illustrated on Fig 3(d).

Table 3: Maximum budget and initial DoE size.

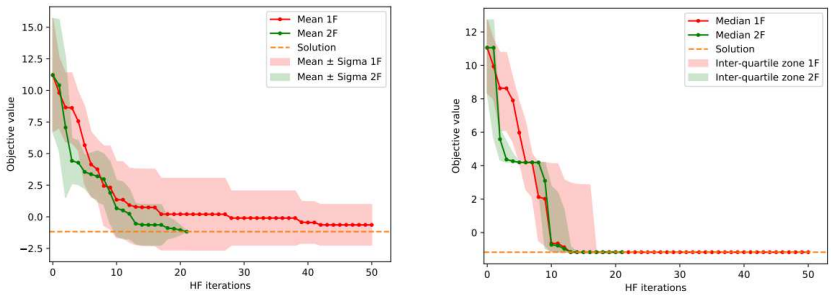
	size initial DoE (mono-fi)	size initial DoE HF (multi-fi)	size initial DoE LF (multi-fi)	budget
Branin	5	5	10	50
Sasena	5	5	10	50
K75	50	50	100	200

Table 4: HF evaluations needed to satisfy the convergence criterion from Eq. (17) and maximum number of LF evaluations over the 10 runs.

	HF evals (mono-fidelity)	HF evals (bi-fidelity)	max LF evals over the 10 runs (bi-fidelity)
Branin case	16	7	29
Sasena case	Don't have enough budget to converge	27	61
K75 case	131	70	144

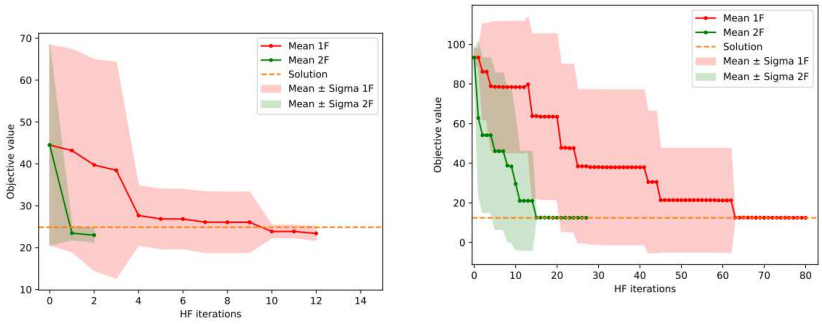
Table 5: Total cost (one computational unit is equivalent to the cost of one HF evaluation) needed to satisfy the convergence criterion for each test case.

	Branin	Sasena	K75
Mono-fidelity	16	greater than 55	131
Multi-fidelity	$7 + \frac{29}{5} = 12.8$	$27 + \frac{61}{5} = 39.2$	$70 + \frac{144}{4.27} = 98.8$



(a) Mean and 1-sigma confidence interval of the objective (Sasena) value at the best valid point at each HF iteration

(b) Median and interquartile range of the objective (Sasena) value at the best valid point at each HF iteration.



(c) Mean and 1-sigma confidence interval of the objective (Branin) value at the best valid point at each HF iteration

(d) Mean and 1-sigma confidence interval of the objective (OAS K75 model) value at the best valid point at each HF iteration

Fig. 3: Comparison of SEGO and MFSEGO methods on the Branin, Sasena and K75 test cases.

6 Conclusion

In this work, mono and multi-fidelity Bayesian methods, SEGO and MFSEGO have been confronted. First, the two approaches were compared on analytical models. Next, a more complex test case involving aerostructural models of the K75 drone has been considered. The MFSEGO algorithm reduces the required number of HF evaluations and the total cost needed so that the mean value (over 10 runs) of the objective function at the best valid point has a 0.5% relative accuracy. In the Branin, Sasena and K75 cases, MFSEGO allows to respectively divide the total cost by at least 1.25, 1.40 and 1.32 compared to SEGO. Future works deal with different research axis like using more than two fidelity levels, using other criteria [26] to determine the enrichment point and the enrichment level or using more complicated multi-disciplinary drone models by adding other components like operations or propulsion to the already implemented disciplines (aerodynamic and structure). As our multi-fidelity strategy can be extended to N fidelity levels, we could also study the effects of employing more than two levels.

Acknowledgements

The PhD is funded by the defense innovation agency (AID) and by the Directorate General of Armaments (DGA) as part of the CONCORDE project. This work is also supported by ONERA internal research project dedicated to multidisciplinary and multi-fidelity design optimization, namely MUFIN and is part of the activities of ONERA - ISAE - ENAC joint research group.

References

[1] Page du d80 titan sur le site web d'elecnor deimos, https://elecnor-deimos.com/project/d80-titan/

[2] Bartoli, N., Bouhlel, M.A., Kurek, I., Lafage, R., Lefebvre, T., Morlier, J., Priem, R., Stilz, V., Regis, R.: Improvement of efficient global optimization with application to aircraft wing design. In: 17th AIAA/ISSMO Multidisciplinary analysis and optimization conference. p. 4001 (2016)

[3] Bartoli, N., Lefebvre, T., Dubreuil, S., Olivanti, R., Bons, N., Martins, J.R., Bouhlel, M.A., Morlier, J.: An adaptive optimization strategy based on mixture of experts for wing aerodynamic design optimization. In: 18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. p. 4433 (2017)

[4] Bartoli, N., Lefebvre, T., Dubreuil, S., Olivanti, R., Priem, R., Bons, N., Martins, J.R., Morlier, J.: Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design. Aerospace Science and technology **90**, 85–102 (2019)

[5] Boucher, Y., Amiez, A., Barillot, P., Chatelard, C., Coudrain, C., Déliot, P., Rivière, N., Riviere, T., Roupioz, L.: Terriscope: An optical remote sensing research platform using aircraft and uas for the characterization of continental surfaces. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences (2018)

[6] Bouhlel, M.A., Bartoli, N., Otsmane, A., Morlier, J.: Improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction. Structural and Multidisciplinary Optimization **53**(5), 935–952 (2016)

[7] Bouhlel, M.A., Bartoli, N., Regis, R.G., Otsmane, A., Morlier, J.: Efficient global optimization for high-dimensional constrained problems by using the kriging models combined with the partial least squares method. Engineering Optimization **50**(12), 2038–2053 (2018)

[8] Bouhlel, M.A., Hwang, J.T., Bartoli, N., Lafage, R., Morlier, J., Martins, J.R.R.A.: A python surrogate modeling framework with derivatives. Advances in Engineering Software p. 102662 (2019). https://doi.org/https://doi.org/10.1016/j.advengsoft.2019.03.005

[9] Chaudhuri, A., Jasa, J., Martins, J.R., Willcox, K.E.: Multifidelity optimization under uncertainty for a tailless aircraft. In: 2018 AIAA Non-Deterministic Approaches Conference. p. 1658 (2018)

[10] Chauhan, S.S., Martins, J.R.: Low-fidelity aerostructural optimization of aircraft wings with a simplified wingbox model using openaerostruct. In: International Conference on Engineering Optimization. pp. 418–431. Springer (2018)

[11] Efron, B., LePage, R.: Introduction to bootstrap. Wiley & Sons, New York (1992)

[12] Forrester, A.I., Sóbester, A., Keane, A.J.: Multi-fidelity optimization via surrogate modelling. Proceedings of the royal society a: mathematical, physical and engineering sciences **463**(2088), 3251–3269 (2007)

[13] Frazier, P.I.: A tutorial on bayesian optimization. arXiv preprint arXiv:1807.02811 (2018)

[14] Hernández-Lobato, J.M., Gelbart, M.A., Adams, R.P., Hoffman, M.W., Ghahramani, Z.: A general framework for constrained bayesian optimization using information-based search (2016)

[15] Jasa, J.P., Hwang, J.T., Martins, J.R.: Open-source coupled aerostructural optimization using python. Structural and Multidisciplinary Optimization **57**(4), 1815–1827 (2018)

[16] Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. Journal of Global optimization **13**(4), 455–492 (1998)

[17] Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63**(3), 425–464 (2001)

[18] Kraft, D., et al.: A software package for sequential quadratic programming (1988)

[19] 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 **52**(6), 119–139 (1951)

[20] Le Gratiet, L.: Multi-fidelity Gaussian process regression for computer experiments. Ph.D. thesis, Université Paris-Diderot-Paris VII (2013)

[21] Matheron, G., de Géostatistique Appliquée, T., Tome, I.: Mémoires du bureau de recherche géologiques et minières, n. 14. Ed. Technip, Paris (1962)

[22] Ng, K.S.: A simple explanation of partial least squares. The Australian National University, Canberra (2013)
[23] Pavlyuk, D.: Computing the maximum likelihood estimates: concentrated likelihood, em-algorithm
[24] Priem, R.: Optimisation bayésienne sous contraintes et en grande dimension appliquée à la conception avion avant projet. Ph.D. thesis, ISAE-SUPAERO (2020)
[25] Rasmussen, C.E., Williams, C.: Gaussian processes for machine learning, vol. 1 (2006)
[26] Sacher, M., Le Maitre, O., Duvigneau, R., Hauville, F., Durand, M., Lothodé, C.: A non-nested infilling strategy for multifidelity based efficient global optimization. International Journal for Uncertainty Quantification **11**(1) (2021)
[27] Sasena, M.J.: Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations. Ph.D. thesis, Citeseer (2002)
[28] Watson, A.G., Barnes, R.J.: Infill sampling criteria to locate extremes. Mathematical Geology **27**(5), 589–608 (1995)
[29] Wrenn, G.A.: An indirect method for numerical optimization using the Kreisselmeier-Steinhauser function, vol. 4220. National Aeronautics and Space Administration, Office of Management ... (1989)

7 Appendices

A Branin case definition

The HF and LF functions and constraints of the Branin problem are:

$$\begin{aligned} f_{Branin,HF}(x_0,x_1) &= (15x_1 - \frac{5.1}{4\pi^2} * (15x_0-5)^2 + \frac{5}{\pi}(15x_0-5) - 6)^2 \\ &\quad + 10((1-\frac{1}{8\pi})\cos(15x_0-5)+1) + 5(15x_0-5) \end{aligned} \tag{18}$$

$$f_{Branin,LF}(x_0,x_1) = f_{Branin,HF}(x_0,x_1) - \cos(0.5x_0) - x_1^3 \tag{19}$$

$$g_{Branin,HF}(x_0,x_1) = -x_0x_1 + 0.2 \leq 0 \tag{20}$$

$$g_{Branin,LF}(x_0,x_1) = -x_0x_1 - 0.7x_1 + 0.3x_0 \leq 0 \tag{21}$$

B Sasena case definition

The HF and LF functions and constraints of the Sasena problem are:

$$\begin{aligned} f_{Sasena,HF}(x_0,x_1) &= 2 + 0.01(x_1 - x_0^2)^2 + (1-x_0)^2 + 2(2-x_1)^2 \\ &\quad + 7\sin(0.5x_0)\sin(0.7x_0x_1) \end{aligned} \tag{22}$$

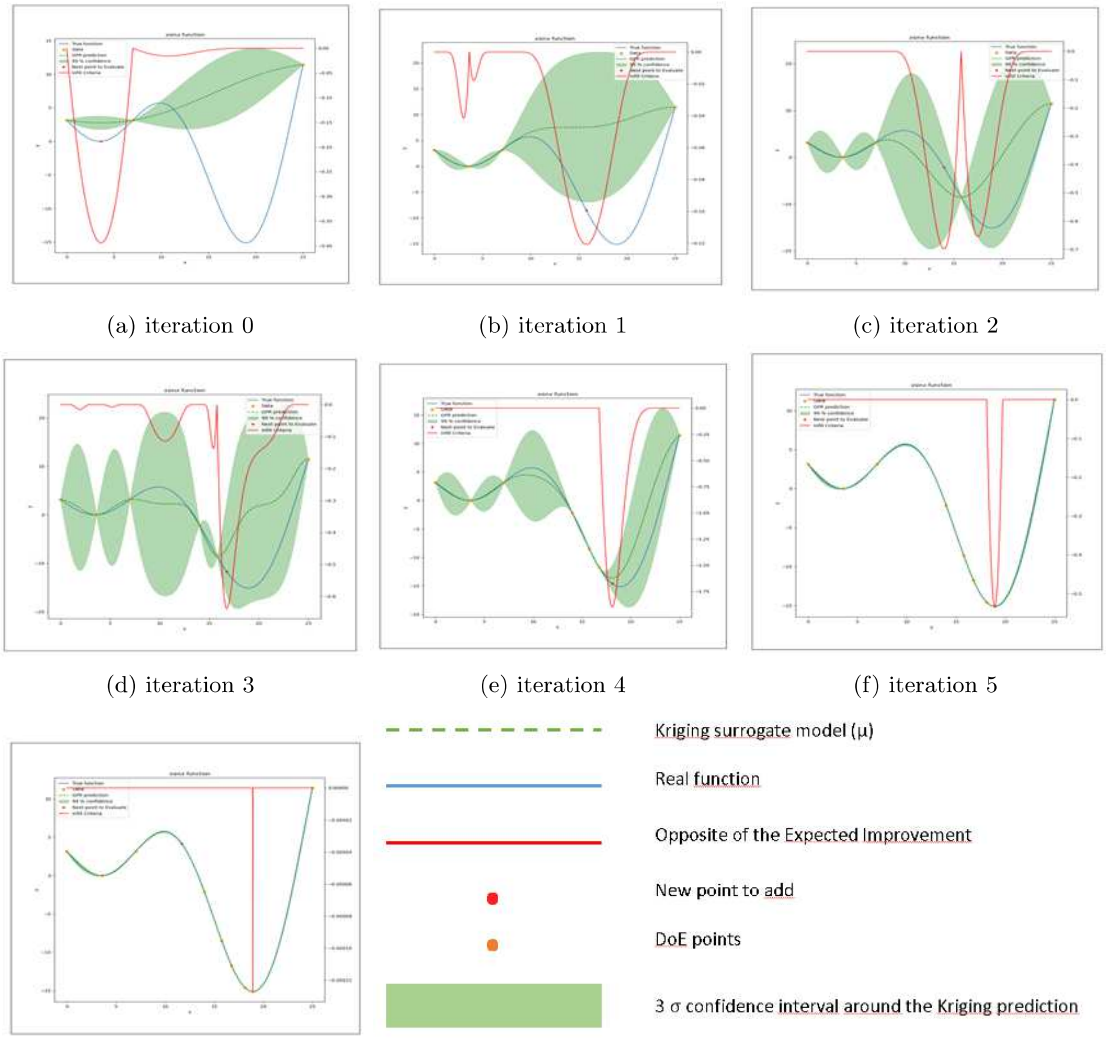
$$f_{Sasena,LF}(x_0,x_1) = f_{Sasena,LF}(x_0,x_1) + \exp(x_0) - x_1^2 \tag{23}$$

$$g_{Sasena,HF}(x_0,x_1) = -\sin(x_0 - x_1 - \frac{\pi}{8}) \leq 0 \tag{24}$$

$$g_{Sasena,LF}(x_0,x_1) = g_{Sasena,HF}(x_0,x_1) + 0.2x_1 - 0.7x_0 + x_0x_1 \leq 0 \tag{25}$$

C Illustration of the BO process

The six first iterations of the EGO algorithm on a one dimensional unconstrained objective function are presented in Figure 4.



(a) iteration 0 (b) iteration 1 (c) iteration 2

(d) iteration 3 (e) iteration 4 (f) iteration 5

(g) iteration 6 (h) legend

Fig. 4: Illustration of the Bayesian optimization process on a one dimensional unconstrained case: $s(x) = x\sin(3\pi(x+0.1))$.

D K75 design variables table

Table 6: Design space and unit of the design variables

	design space	unit
3 wingbox spar thickness control points along wing span	$[0.001, 0.01]^3$	m
3 wingbox skin thickness control points along wing span	$[0.001, 0.01]^3$	m
3 wingbox spar thickness control points along tail span	$[0.001, 0.01]^3$	m
3 wingbox skin thickness control points along tail span	$[0.001, 0.01]^3$	m
angle of attack cruise flight	$[0, 15]$	deg
angle of attack maneuver	$[-15, 15]$	deg
fuelmass	$[0, 50]$	kg

E MFSEGO methodology pseudo-code

Algorithm 1 MFSEGO algorithm

Compute initial DoE using LHS
while (maximum budget is not reached) and ($y_{best}^{valid} > f_{HF}(ref_{sol}) + tol$) **do**
  Learn LF Kriging surrogate model ($\hat{\mu}_0$ and $\hat{\sigma}_0^2$)
  for $k = 1...L$ **do**
    Learn ρ_{k-1} and $\hat{\mu}_{\delta_k}$
    Deduce $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ and so the k-th fidelity level Kriging surrogate model
  end for
  Choose x_{next} that optimizes acquisition function (Eq 9)
  Select the level of enrichment t (Eq 16)
  for $l = 0...t$ **do**
    Add $(x_{next}; f_l(x_{next}))$ to the DoE
  end for
  $y_{best}^{valid} = \min\ Y^{valid}$
end while
return y_{best}^{valid}

F MFSEGO methodology diagram

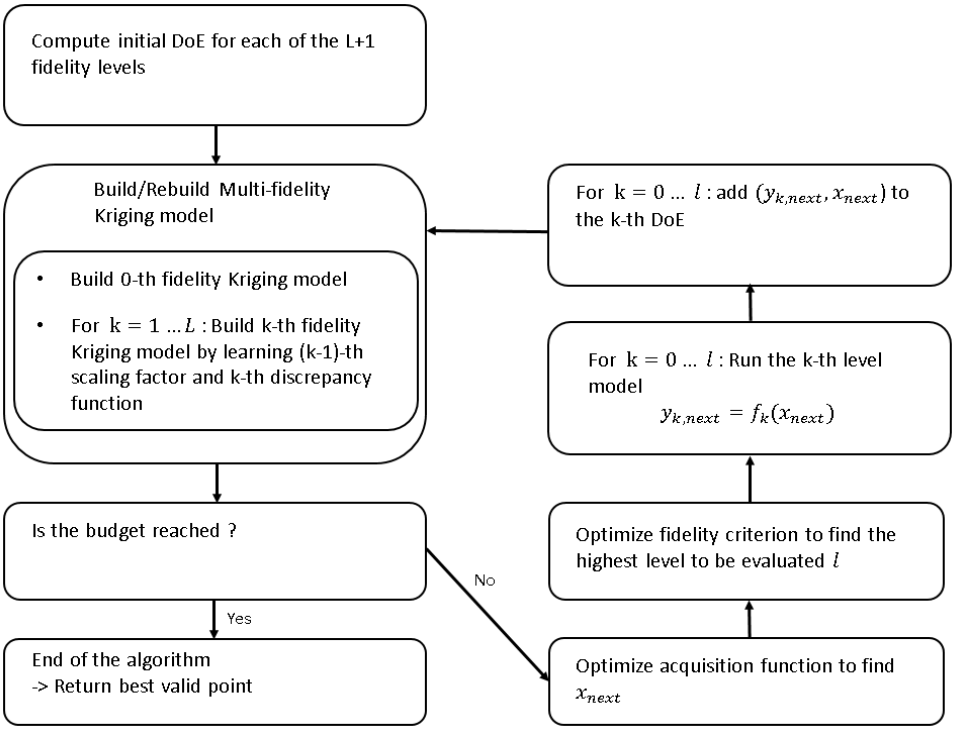


Fig. 5: MFSEGO methodology diagram