



HAL
open science

Constrained bayesian optimization over mixed categorical variables, with application to aircraft design

Paul Saves, Nathalie Bartoli, Youssef Diouane, Thierry Lefebvre, Joseph Morlier, Christophe David, Eric Nguyen Van, Sébastien Defoort

► To cite this version:

Paul Saves, Nathalie Bartoli, Youssef Diouane, Thierry Lefebvre, Joseph Morlier, et al.. Constrained bayesian optimization over mixed categorical variables, with application to aircraft design. AeroBest 2021, Jul 2021, Lisbonne, Portugal. hal-03346341

HAL Id: hal-03346341

<https://hal.science/hal-03346341v1>

Submitted on 16 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONSTRAINED BAYESIAN OPTIMIZATION OVER MIXED CATEGORICAL VARIABLES, WITH APPLICATION TO AIRCRAFT DESIGN

Paul Saves^{1,2}, Nathalie Bartoli², Youssef Diouane¹, Thierry Lefebvre², Joseph Morlier³, Christophe David², Eric Nguyen Van² and Sébastien Defoort²

1: ISAE-SUPAERO, Université de Toulouse, Toulouse, France
{paul.saves2,youssef.diouane}@isae-supero.fr

2: ONERA/DTIS, Université de Toulouse Toulouse, France
{nathalie.bartoli,thierry.lefebvre,christophe.david,eric.nguyen_van,sebastien.defoort}@onera.fr

3: ICA, Université de Toulouse, ISAE-SUPAERO, MINES ALBI, UPS, INSA, CNRS,
Toulouse, France
joseph.morlier@isae-supero.fr

Abstract. *Multidisciplinary Design Optimization (MDO) methods aim at adapting numerical optimization techniques to the design of engineering systems involving multiple disciplines or components. Among MDO architectures, various ones are considering the resolution of the Multidisciplinary Design Analysis (MDA). In our study, the system of interest being an aircraft, the resolution of the MDA will be provided by the Future Aircraft Sizing Tool with Overall Aircraft Design (FAST-OAD), a point mass approach that estimates the required fuel and energy consumption for a given set of top-level aircraft requirements. In this context, a large number of mixed continuous, integer and categorical variables that arise from aircraft design has to be tackled by the optimization process.*

Recently, there has been a growing interest in mixed variables constrained Bayesian optimization based on Gaussian process surrogate models. In this setting, most existing approaches severely increase the dimension of the covariance matrix related to the surrogate. In fact, the construction of the Gaussian process model may not be scalable to practical applications involving a large number of mixed variables.

In this paper, we address this issue by constructing a covariance kernel for the surrogate model that depends on only a few hyperparameters. The new kernel is constructed based on the information obtained from the partial least squares method. The obtained numerical results lead to interesting results for the optimization of a baseline aircraft and to reduce the fuel consumption of “DRAGON”, a new hybrid electric propulsion aircraft, with a high number of mixed variables and for a small budget of time-consuming evaluations.

Keywords: Green Aircraft, Future Aircraft, Multidisciplinary Design Optimization, Bayesian Optimization, Surrogate-based Optimization, Gaussian Process

1 INTRODUCTION

New aircraft configurations with a lower footprint on the environment (also known as Eco-aircraft design) have seen a resurgence of interest [1]. In this context, one targets to minimize the footprint on the environment of the aircraft using a *Multidisciplinary Design Analysis* (MDA) [2–4] with dedicated disciplines such as emissions, noise, ... The process of finding the best configuration, known as *Multidisciplinary Design Optimization* (MDO), is formulated as a minimization constrained problem where the objective and the constraints functions are typically expensive-to-evaluate and their derivatives are not available.

In the context of aircraft design, the MDO process generally involves mixed continuous-categorical design variables. For instance, the size of aircraft structural parts can be described using continuous variables; in case of thin-sheet stiffened sizing, they represent panel thicknesses and stiffening cross-sectional areas. The set of discrete variables can encompass design variables such as the number of panels, the list of cross sectional areas or the material choices. Thus, the regarded optimization problem is of the following form:

$$\min_{w=(x,z,c)\in\Omega\times S\times\mathbb{F}^l} \{f(w) \text{ s.t. } g(w) \leq 0 \text{ and } h(w) = 0\} \quad (1)$$

where $\Omega \subset \mathbb{R}^n$ represents the bounded continuous design set for the n continuous variables; $S \subset \mathbb{Z}^m$ represents the bounded integer set where L_1, \dots, L_m are the numbers of levels of the m quantitative integer variables on which we can define an order relation and $\mathbb{F}^l = \{1, \dots, L_1\} \times \{1, \dots, L_2\} \times \dots \times \{1, \dots, L_l\}$ is the design space for l categorical qualitative variables with their respective L_1, \dots, L_l levels.

$f : \mathbb{R}^n \times \mathbb{Z}^m \times \mathbb{F}^l \mapsto \mathbb{R}$ is the objective function, $g : \mathbb{R}^n \times \mathbb{Z}^m \times \mathbb{F}^l \mapsto \mathbb{R}^{n_g}$ gives the inequality constraints, and $h : \mathbb{R}^n \times \mathbb{Z}^m \times \mathbb{F}^l \mapsto \mathbb{R}^{n_h}$ returns the equality constraints. The functions f , g , and h are typically expensive-to-evaluate simulations with no exploitable derivative information.

When only continuous design variables are optimized (i.e., $S \times \mathbb{F}^l$ is reduced to a single choice), *Bayesian optimization* (BO) is shown to be a powerful strategy for solving problem (1) [5]. BO uses *Gaussian processes* (GPs) [6–10] to define response surfaces, the sequential enrichment is performed by maximizing a given acquisition function [6]. The latter is meant to model a compromise between exploration of new zones in the design space and exploitation (i.e. minimization) of the GPs. For general mixed integer problems, several modeling strategies to build GPs have been proposed [11, 12]. Based on these GPs models, a trade-off acquisition function was adapted for optimization [13]. Some other modelling strategies consist in computing a continuous model for each category [14], either by continuously relaxing the design variables [15], by using a multi-armed bandit strategy to handle the categorical choices [16] or by considering a Gower distance to model simultaneously the proximity over categorical and continuous variables [17]. Recently, a continuous relaxation BO based method [18] to tackle mixed integer variables has been shown to solve efficiently expensive-to-evaluate optimization problems. In fact, using continuous relaxation within BO leads to better results. However, the relaxation of the categorical design variables increases the number of the hyperparameters needed (to be tuned) associated with the GP model. This in particular constrained the method in [18] to be used only for small dimensional optimization problems. Since, the construction of the GP model may not be scalable to practical applications involving a large number of mixed variables.

In this work, we show how to reduce the computational cost related to the construction of the mixed categorical GP model as proposed in [15]. Our proposed approach allows in particular to solve higher dimensional mixed integer MDO problems. The method relies on combining continuous relaxation and the use of principle components analysis to reduce the number of the hyperparameters (known as KPLS [19, 20], Kriging model with Partial Least Squares). The good potential of the proposed approach is showed over a set of analytical test cases.

The performance of the proposed approach is also confirmed on two MDO applications from the FAST-OAD framework [21]: “CERAS” and “DRAGON”. FAST-OAD is an open-source Python framework that provides a flexible way to build and solve the Overall Aircraft Design problems by assembling discipline models from various sources: FAST-OAD currently comes with some bundled, quick and simple, models dedicated to commercial aircraft. In this paper, FAST-OAD will resolve an MDA problem that mainly: (a) sizes the geometry of main aircraft components, (b) computes mass and centers of gravity of aircraft parts, (c) estimates the aerodynamics and propulsion along the computed mission, and (d) return the fuel consumption related to the mission. These estimated quantities will be used to define the objective and the constraints of our two optimization problems.

The outline of the paper is as follows. In Section 2, a detailed review of the Bayesian optimization framework is given. The continuous relaxation as well as the use of the KPLS technique are detailed in Section 3. Section 4 presents our academical tests as well as the obtained results on the two regarded MDO test cases. Conclusions and perspectives are finally drawn in Section 5.

2 CONTINUOUS CONSTRAINED BAYESIAN OPTIMIZATION

In this section, we will only consider that all the design variables are continuous in problem (1). Namely, the design space will be restricted to $\Omega \subset \mathbb{R}^n$; hence $w = x$ in the optimization problem (1). The Efficient Global Optimization (EGO) framework [6] suggests to use the GP model to find the global minimum of an expensive-to-evaluate black-box objective function (based on the preliminary results of Moćkus [7]). In general, a GP is used to fit a response surface model from an initial set of points known as the Design of Experiments (DoE) [8, 9, 22]. The GP provides a mean response hypersurface as well as a pointwise estimation of its variance. Thereafter, we will consider that our unknown black-box objective function f is a realization of an underlying distribution of mean \hat{f} and of standard deviation s^f such that $f(\cdot) \sim GP(\hat{f}(\cdot), [s^f]^2(\cdot))$.

Let n_t be the number of already evaluated points in \mathbb{R}^n of the deterministic function f and $\forall i \in \{1, \dots, n_t\}$, let $x^i = (x_1^i, \dots, x_n^i) \in \mathbb{R}^n$ be the i^{th} point with its respective n continuous variables values, we can define the known data as (x, y^f) . The stochastic model [23] writes as: $f(x^i) = \mu(x^i) + \epsilon^i \in \mathbb{R}$ with ϵ^i the error term between f and the model approximation $\mu(x)$. The errors terms are considered *iid* of variance σ^2 .

Let R be the error correlation matrix between the inputs points $R_{ij} = Corr(\epsilon(x^i), \epsilon(x^j))$. The correlation function $Corr$ is computed using a kernel function k that relies on n hyperparameters θ estimated typically using maximum likelihood estimator (MLE) [24]. Let $r_i(x^*) = Corr(\epsilon(x^*), \epsilon(x^i))$ for a given x^* , then

$$\hat{f}(x^*) = \hat{\mu}^f + r(x^*)^T R^{-1}(y^f - \mathbf{1}\hat{\mu}^f), \quad (2)$$

and

$$[s^f]^2(x^*) = [\hat{\sigma}^f]^2 \left[1 - r(x^*)^T R^{-1} r(x^*) + \frac{(1 - \mathbf{1}^T R^{-1} r(x^*))^2}{\mathbf{1}^T R^{-1} \mathbf{1}} \right], \quad (3)$$

where $\hat{\mu}^f$ and $\hat{\sigma}^f$, respectively, are the maximum likelihood estimator of μ and σ given the data set (x, y^f) .

Within EGO, at a given iteration t , a GP surrogate, referred as Kriging, model of the black-box f is computed from the current DoE. Henceforth, one wants to estimate the best new point to evaluate, as it is costly, by taking into account all the available information to converge as fast as possible to the real optimum of the black-box. The point that we will evaluate next is the one that gives the best improvement *a priori* according to an acquisition function like the Expected Improvement (*EI*). The objective value as this new point will then be evaluated and used to enrich the surrogate database. The hyperparameters that characterize and define the model are updated at each iteration until convergence. The Bayesian optimization process is thus made from these GPs in an iterative manner.

To tackle constrained Bayesian optimization, EGO was extended to Super-Efficient Global Optimization (SEGO) method [10]. SEGO uses surrogate models of the constraints to give an estimation of the search space Ω_f . The latter was enhanced to tackle multimodal and equality constraints with the Upper Trust Bound (*UTB*) criterion [25, 26].

The acquisition function that we use is the *WB2s* (Watson and Barnes 2nd criterion with scaling) [26] that is known to be more robust than the Expected improvement (*EI*) criterion; especially in high dimension. *WB2s* can be seen as smooth version compared to the *WB2* [10] criterion and is less multimodal compared to *EI*. Algorithm 1 details the SEGO optimization procedure.

Algorithm 1: SEGO for continuous inputs.

Result: Solution of the problem (1) over the continuous design space Ω .

Inputs: Initial DoE \mathcal{D}_0 and set $t = 0$. The search space Ω .

while the stopping criterion is not satisfied **do**

1. Build the surrogate model of the objective to obtain the mean and standard deviation prediction at a given point: (\hat{f}, s^f) from the DoE \mathcal{D}_t .
2. Build the surrogate model for each constraint i, j : $(\hat{g}_i, s_i^g), (\hat{h}_j, s_j^h)$ from the DoE \mathcal{D}_t and compute an estimation of the search space Ω_f .
3. Construct the acquisition function *WB2s*.
4. Maximize the acquisition function *WB2s* over Ω_f : $x_t = \arg \max_{x \in \Omega_f} \text{WB2s}(x)$.
5. Add $x_t, f(x_t), g(x_t), h(x_t)$ to the DoE \mathcal{D}_{t+1} . Increment t .

end

3 MIXED CATEGORICAL CONSTRAINED BAYESIAN OPTIMIZATION

To handle mixed categorical design variables, we propose to use the continuous relaxation method that has been recently shown to be well-suited for expensive discrete problems [15, 18]. The main drawback of such method is that it enlarges the dimension of the design space according to the size of the categorical space. To overcome such issue,

we propose to combine continuous relaxation with the well-known partial least-square procedure [20] to reduce the number of the GP hyperparameters.

For mixed categorical design variables, the proposed treatment relies on continuous relaxation. The design space $\Omega \times S \times \mathbb{F}^l$ is relaxed to a continuous space Ω' constructed on the following way:

- $\forall i \in \{1, \dots, m\}$, the variable z_i is relaxed within its bounds and treated as continuous.
- $\forall j \in \{1, \dots, l\}$, we use a relaxed one-hot encoding [27] for c_j and add L_j new continuous dimensions into Ω' .

Therefore, we get, after relaxation, a new design space $\Omega' \subseteq \mathbb{R}^{n'}$ where $n' = n + m + \sum_{j=1}^l L_j > n + m + l$.

The nature of the variables should be respected when evaluating a point in the relaxed space so we define the inverse operator *Project* that projects a point $X \in \Omega'$ to its closer point w^X in $\Omega \times S \times \mathbb{F}^l$. Namely, *Project* rounds the value of an integer variable z_i to the closer value among its L_i levels and, for a categorical variable c_j , *Project* selects the level which corresponding dimension value is the highest.

In this work, when building the Kriging model, the error correlation will be estimated using a squared exponential (or Gaussian) correlation kernel over the relaxed design space. We denote by $X^{w^i} = \text{Relax}(w^i)$ the relaxation in Ω' of a point $w^i \in \Omega \times S \times \mathbb{F}^l$. The mixed categorical kernel is

$$k(X^{w^i}, X^{w^j}) = \prod_{p=1}^{n'} \exp\left(-\theta_p (X_p^{w^i} - X_p^{w^j})^2\right), \theta_p \in \mathbb{R}^+ \quad (4)$$

This kernel relies on n' hyperparameters θ_p estimated by maximum of likelihood such that the more the number of variables n' for the problem, the more the number of hyperparameters to optimize. Reducing this number leads to a better estimation for the hyperparameters, a more convenient optimization of the likelihood and makes the model scalable for high-dimensional problems. To do so, the Partial Least Squares (PLS) method [28] searches the direction that maximizes the variance between the input and output variables. This is done by a projection into a smaller space spanned by the so-called principal components. The number of principal components d that corresponds to the new number of hyperparameters for KPLS is chosen to be much lower than n' . The resulting PLS squared exponential kernel is given by

$$k(X^{w^i}, X^{w^j}) = \prod_{q=1}^d \prod_{p=1}^{n'} \exp\left(-\theta_q (b_{*p}^q X_p^{w^i} - b_{*p}^q X_p^{w^j})^2\right), \forall \theta_q \in \mathbb{R}^+ \quad (5)$$

where $[b_{*p}^q]_{p,q}$ are scalars that measure the influence of the input variables on the output y^f . Combining this model construction with SEGO gives the method described in Algorithm 2.

Algorithm 2: SEGO for mixed categorical inputs using KPLS

Result: Solution of the problem (1) over the mixed categorical design space $\Omega \times S \times \mathbb{F}^l$.

Inputs: Initial DoE \mathcal{D}_0 and set $t = 0$. The search space $\Omega \times S \times \mathbb{F}^l$.

while the stopping criterion is not satisfied **do**

1. Relax continuously integer and categorical input variables to a real bounded space Ω' of dimension $n' = n + m + \sum_{j=1}^l L_j$. Namely, we continuously relax the mixed categorical DoE $\mathcal{D}_t = \{w^i\}_i \in (\Omega \times S \times \mathbb{F}^l)^{n_t}$ to a continuous DoE $\mathcal{D}'_t = \{X^{w^i}\}_i \in (\Omega')^{n_t}$ using the relaxation procedure *Relax*.
2. Use PLS to reduce the number of the hyperparameters from n' to d ($d \ll n'$), during the construction of the GP models (for the objective function f and the constraints g and h) related to the DoE \mathcal{D}'_t .
3. Build an estimation of the feasible domain $\Omega'_f \subset \Omega'$ with the criterion *UTB* and construct the acquisition function *WB2s*.
4. Maximize the acquisition function *WB2s* over Ω'_f :

$$X_t = \arg \max_{X \in \Omega'_f} WB2s(X).$$
5. Project the obtained continuous solution over $\Omega \times S \times \mathbb{F}^l$:

$$w_t = Project(X_t).$$
6. Add $w_t, f(w_t), g(w_t), h(w_t)$ to the DoE \mathcal{D}_{t+1} . Increment t .

end

4 RESULTS

In this section, we carry out experiments for unconstrained and constrained test cases, with several number of variables and an increasing complexity. We optimize analytical test cases as a benchmark study and then we solve aircraft design optimization problems as a validation and application of the present work.

4.1 Implementation choices

In order to compare, we used several optimization algorithms detailed hereafter: Bandit-BO, NSGA2, SEGO with Kriging, SEGO with Gower distance and the proposed method, SEGO coupling Kriging and PLS.

The Bandit-BO implementation used is the one by Nguyen et al. [16], we are not considering parallelization or batch evaluations. Bandit-BO creates a GP model for each arm, so it requires at least $2 \times N_c$ initial points, N_c being the number of categorical possibilities for the problem inputs. If we are not using Bandit-BO, for constrained optimization, we are using a continuous relaxed Latin hypercube sampling and then we project the output points to obtain the mixed integer DoE.

The NSGA2 [29] algorithm used is the implementation from the toolbox pymoo [30] with the default parameters (probability of crossover = 1, eta = 3). Fronts are not relevant in our study as we are considering single-objective optimization.

The optimization with SEGO is made from SEGOMOE [26] for both constrained and

unconstrained cases. For SEGO using Gower distance [17] (denoted by SEGO+GD), we are considering the implementation of the Surrogate Modeling Toolbox (SMT) [31], an open-source python toolbox where some computations associated to the present work have been done. The same holds for SEGO and Kriging coupled with PLS to reduce to d the number of hyperparameters (denoted SEGO+KPLS_ dD). As the PLS could potentially lead to numerical instabilities, we are using the homoscedastic noise that maximizes the likelihood as a so-called nugget. For SEGO using Kriging (denoted by SEGO+KRG), we also use the implementation from the toolbox SMT. For the constrained analytical test cases, we are using the *UTB* criterion [25]. Some adaptations have been done to Bandit-BO and NSGA2 to consider both integer and categorical variables. As NSGA2 can only consider integer variables, categorical variables are treated as integer ones. Contrarily, Bandit-BO can treat only categorical variables so integer variables are treated as categorical ones. For Kriging, Kriging with PLS or Gower distance, the hyperparameters are optimized with COBYLA [32] and the chosen model regression is constant. When optimizing with SEGO, the acquisition function is maximized using ISRES [33] to find some interesting starting points and SNOPT [34] to finalize the process based on these starting points. The squared exponential kernel is the only one considered for these methods.

In order to compute some statistical data (median and variance), we are doing 20 repetitions of the optimization process for a given method and an initial DoE size. We consider that a constraint is respected if the constraint violation is smaller than the threshold value 10^{-4} .

4.2 Benchmark test cases including unconstrained and constrained optimization problems

In this section, a benchmark of different problems is proposed in order to compare the efficiency of the proposed algorithm with some state-of-the-art methods. The first ten are analytical cases, with or without constraints in order to provide some data profiles and the last two concern some more complex applications for aircraft conceptual design.

4.2.1 Unconstrained optimization

To begin with, we validate our method on unconstrained problems up to 14 dimensions. The first analytical test case is a modified Branin function [14], denoted by “**Branin 5**”, where the first variable is an integer $x_1 \in \{-5, -4, \dots, 9, 10\}$ and the second one $x_2 \in [0, 10]$ is a continuous variable. As this problem is only 2-dimensional, SEGO-KRG is considered without the coupling with the PLS technique. For Bandit-BO, we represent x_1 as a bandit with 16 arms associated to the 16 integer values from 0 to 15. Therefore, an initial DoE of at least 32 points is required for Bandit-BO. A smaller initial DoE is also considered for NSGA2 and SEGO-KRG in order to compare the convergence according to the DoE size: 5 points or 32 points. For a given DoE size (5 or 32), 20 different initial DoE are obtained via Latin hypercube sampling in order to compute some statistical data (median and variance) about the convergence results. In Fig.1a, the medians and the associated quartiles (25% and 75%) of the 20 runs are illustrated for each of the three algorithms. The initial DoE is shown before the black dotted line. For the DoE with 32 initial points, we are doing 50 iterations of the methods Bandit-BO and SEGO-

KRG, for a total of 82 evaluations. For NSGA2, 200 iterations are performed. When the DoE size is reduced to 5 points, 200 iterations of SEGO-KRG are made and compared to 300 evaluations for NSGA2. These comparisons show that SEGO-KRG outperforms BANDIT-BO and NSGA2 in terms of convergence. Moreover, we find that a smaller initial DoE performs better for SEGO, that is a known result [35].

To compare the dispersion of the best results over the 20 runs, Fig.1b shows the boxplots for SEGO-KRG, Bandit-BO and NSGA2 after 50 iterations associated to a initial DoE of 32 points. These plots show that, not only SEGO-KRG converges better in median but also it is the only method to converge for every of the 20 initial DoE. However, Bandit-BO converges better than NSGA2 on this test case.

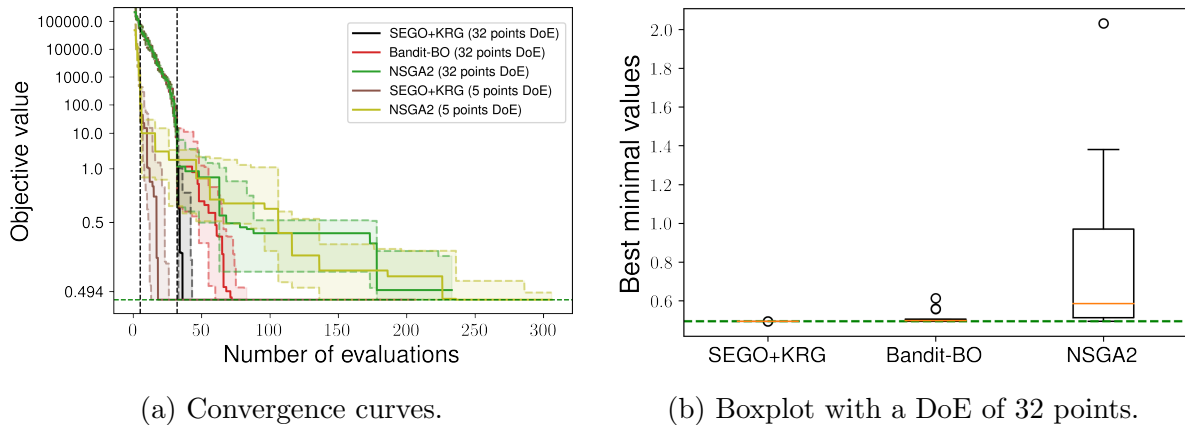
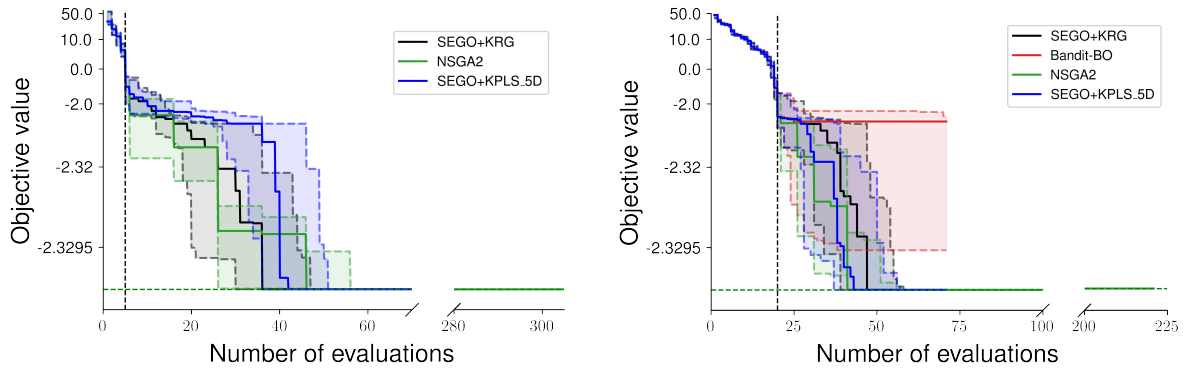


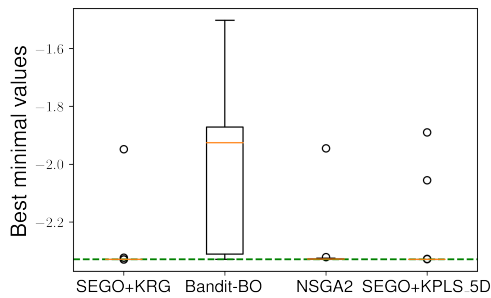
Figure 1: “Branin 5” obtained optimization results. The Boxplots are generated, after 50 iterations, using the 20 best points.

The second analytical test case is a toy function that consists in the choice between a set of 10 1-dimensional continuous functions [13], denoted by “Set 1”, where the first variable is a categorical variable $x_1 \in \{0, 1, \dots, 9\}$ and the second one $x_2 \in [0, 1]$ is a continuous variable. This problem only has 2 variables but, as the relaxed space is in dimension 11, SEGO with KPLS is then considered to compare with SEGO with Kriging such that we can see how reducing the number of spatial correlations from 11 to 5 in the model can impact the optimization process. For Bandit-BO, we represent x_1 as a bandit with 10 arms from 0 to 9 and an initial DoE with a minimum of 20 points is required. As previously for NSGA2 and SEGO-KRG, two initial DoE sizes are considered with 5 points and 20 points. For a given DoE size, 20 different initial DoE are obtained to compute statistics on the convergence results. In Fig.2, the medians and the associated quartiles (25% and 75%) of the 20 runs are illustrated for each of the four algorithms. The initial DoE is shown before the black dotted line. For the DoE with 20 initial points in Fig.2b, 50 iterations of the methods Bandit-BO, SEGO+KPLS_5D and, SEGO-KRG are performed for a total of 70 evaluations and for NSGA2 200 iterations are performed. When the DoE size is reduced to 5 points in Fig.2a, 200 iterations of SEGO-KRG are done and compared to 300 evaluations for NSGA2. These convergence plots show that the smaller the DoE, the faster the convergence. Also, KPLS slows the convergence at the start but this dimension reduction does not change the convergence overall and the incumbent is even better at the end with KPLS than without. On this test case, Bandit-BO method does not perform well. To compare the dispersion of the best results

over the 20 runs, Fig.2c shows the boxplots for SEGO-KRG, Bandit-BO, NSGA2 and SEGO+KPLS_5D after 50 iterations associated to a initial DoE of 20 points. These plots show that SEGO+KPLS does not converge properly because 2 runs are outliers that did not have converged whereas with NSGA2 and SEGO-KRG, there is only one outlier. Nevertheless, NSGA2 is not as precise as SEGO-KRG on average.



(a) Convergence curves using 5 points initial DoE (b) Convergence curves using 20 points initial DoE.



(c) Boxplots with a DoE of 20 points.

Figure 2: “Set 1” obtained optimization results.

4.2.2 Constrained optimization

For constrained optimization, Bandit-BO can not be considered as it can not deal with constraints, so SEGO is only compared to NSGA2. The first constrained test case is a modified Branin function [11] with one constraint, denoted by “**Branin 3**”, where the two first variables are categorical variables with 2 levels each such that there is 4 possible Branin function variations and the two last variables are the continuous ones.

This problem has 4 variables in the initial space and 6 in the relaxed one, so SEGO with Kriging is applied without any dimension reduction technique. According to the previous experiments, a small initial DoE with 5 points is considered in order to obtain better results for a given number of evaluations. In Fig.3a, the medians and the associated quartiles (25% and 75%) of the 20 runs are illustrated for NSGA2 and SEGO-KRG. The initial DoE is shown before the black dotted line, then 50 iterations of SEGO-KRG are performed and compared with 200 iterations of NSGA2. To compare the dispersion of the best results over the 20 runs, Fig.3b shows the boxplots for SEGO-KRG and NSGA2 after 50 iterations. On this low-dimension constrained case, the mixed integer version of

SEGO with Kriging is shown to perform well and be adapted to the constrained mixed optimization.

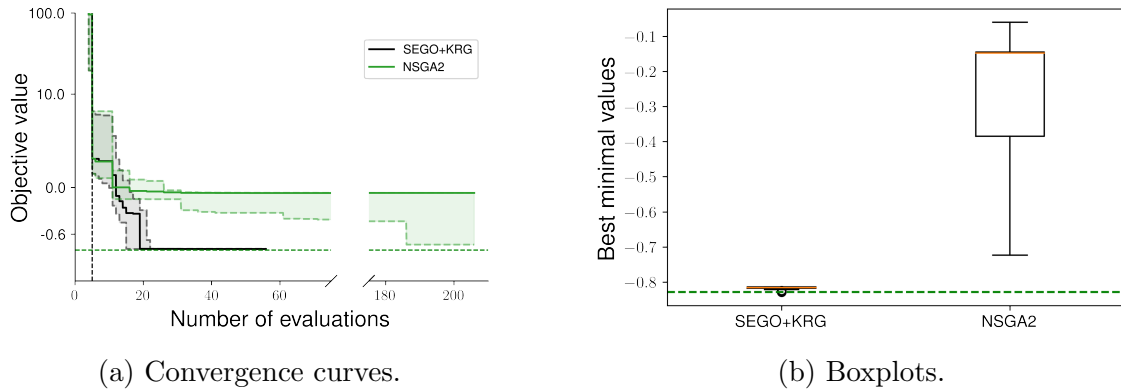


Figure 3: “Branin 3” optimization results.

The second constrained test case is an augmented dimension modified Branin function [11] with one constraint, denoted by “Branin 4”. The firsts two variables are categorical with 2 levels each and the 10 last variables are continuous. Therefore, we can see the test case as a set of 4 continuous augmented 10-dimensional Branin functions. This problem has 12 variables in the initial space and 14 in the relaxed space. SEGO with KPLS is considered with 2 principal components for the PLS technique to compare the dimension reduction with SEGO-KRG. As previously an initial DoE with 5 points is chosen and shown before the dotted line in Fig.4a. The medians and the associated quartiles (25% and 75%) of the 20 runs are illustrated for NSGA2 and SEGO. On this graph, 50 iterations of SEGO-KRG are compared with 200 iterations of NSGA2. To compare the dispersion of the best results over the 20 runs, Fig.4b shows the boxplots for SEGO-KRG, SEGO-KPLS_2D and NSGA2 after 50 iterations. On this high-dimension constrained case, our method is still efficient.

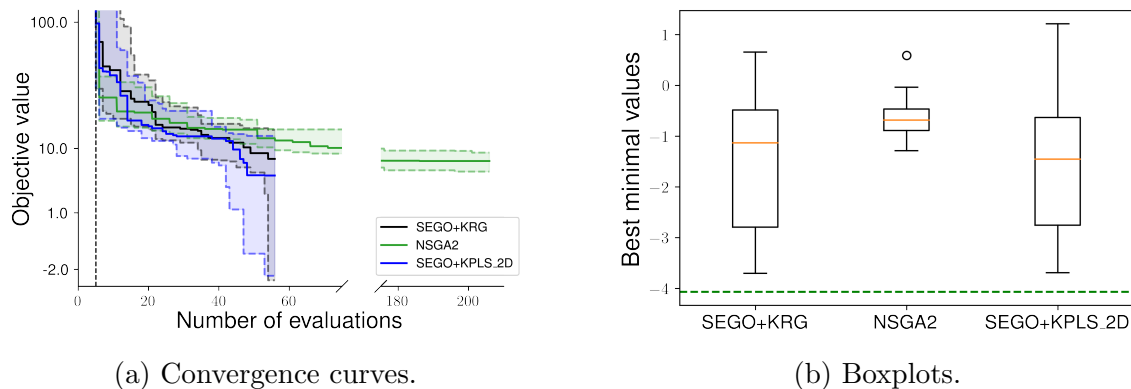


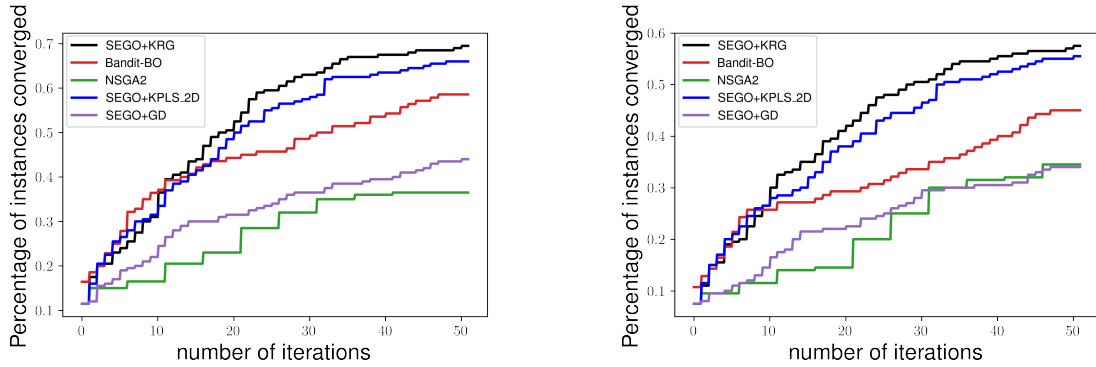
Figure 4: “Branin 4” optimization results.

These four analytical cases have shown that SEGO performs better than both Bandit-BO for mixed integer Bayesian Optimization and NSGA2 for constrained optimization. It has been shown that the PLS technique allows the method to be scalable for high-dimension and even to give better results than Kriging by carrying favor to the most relevant space.

4.2.3 Overall benchmark results

For 10 different test cases (3 constrained and 7 unconstrained) we are considering 20 runs with different DoE sampled by Latin hypercube sampling for a total of 200 instances. The resulting percentages of instances that have converged after a given budget for every method are plotted on Fig.5. For unconstrained test cases, in order to compare with Bandit-BO, the size of the initial DoE is given by $\min\{5, 2 \times N_c\}$ where N_c is the number of categorical possibilities and for constrained test cases, we took 5 points for the initial DoE size. This allow us to compare Bandit-BO, SEGO, SEGO with Gower Distance, SEGO with KPLS and NSGA2. Some tests are in dimension 2, so in order to compare with the same number of hyperparameters for all tests, we had to choose between 1 or 2 principal components for KPLS. As the number of points increases, the projected points are really closed to each other, so, to insure a certain stability, we choose KPLS with 2 principal components and noise evaluation, denoted by KPLS_2D in the following.

These 10 test cases were extracted from several state-of-the-art papers [11–14, 17, 36, 37] from dimension 2 to 12, with integer, continuous and/or categorical variables. For constrained case, we keep only the inputs that gives a constraint violation smaller than 10^{-4} . A problem is considered solved if the error to the known solution is smaller than 2% on Fig.5a and smaller than 0.5% on Fig.5b. The mean error after the 50 iterations can be found in Tab.1. SEGO with PLS gives the smaller errors on constrained test cases but, for unconstrained ones, SEGO-KRG performs the best. As we can see on the data profiles, SEGO-KRG and SEGO+KPLS_2D are similar and outperform the three others methods. However, SEGO with PLS being less efficient on unconstrained test cases, SEGO-KRG gives better results over all tests. These preliminary results are promising and some more realistic applications are considered in the next section.



(a) Data profiles for a tolerance of 2%

(b) Data profiles for a tolerance of 0.5%

Figure 5: Data profiles generated using 10 analytical test cases

Table 1: Mean errors of each method after 50 iterations.

ERRORS	NSGA2	Bandit-BO	SEGO+KRG	SEGO+KPLS_2D	SEGO+GD
Unconstrained test cases	14.29%	5.84%	2.27%	5.74%	8.41%
Constrained test cases	58.27%	-	27.61%	25.41%	47.83%

4.3 Aircraft design test cases

For the core MDO application, we apply FAST-OAD on two different aircraft design problems. The first one named “CERAS” is a classic well-known aircraft and the second “DRAGON” is a more innovative new one that is currently under development. This time, as the evaluations are expensive, we are doing only 10 runs instead of 20. For each run, we draw a random starting DoE of 5 points. As we have no equality constraint and a small budget, we will not force the constraints to be as large as possible and we will use only the mean prediction of the constraint models and not the *UTB* criterion [25]. Also, to have realistic results, the constraints violation will be forced to be less than 10^{-3} .

4.3.1 CERAS

To validate the results of section 4.2 on a real test case, we are considering the data from the CEntRAL Reference Aircraft System (“CERAS”) based on an Airbus A320 aircraft. The problem to solve is a constrained optimization problem with 6 continuous design variables, 2 categorical variables and 2 integer ones, for a total of 10 design variables. The presented version of SEGO (with or without the PLS technique) is used as an optimizer in a Multidisciplinary feasible (MDF) architecture where the MDA is computed with FAST-OAD. The optimization problem is described in Tab.2 where the total number of variables is reported using the relaxation technique.

Table 2: Definition of the “CERAS” optimization problem.

	Function/variable	Nature	Quantity	Range
Minimize	Fuel mass	cont	1	
with respect to	x position of MAC	cont	1	[16., 18.] (<i>m</i>)
	Wing aspect ratio	cont	1	[5., 11.]
	Vertical tail aspect ratio	cont	1	[1.5, 6.]
	Horizontal tail aspect ratio	cont	1	[1.5, 6.]
	Wing taper aspect ratio	cont	1	[0., 1.]
	Angle for swept wing	cont	1	[20., 30.] ($^{\circ}$)
	Total continuous variables		6	
	Cruise altitude	discrete	1	{30k,32k,34k,36k} (<i>ft</i>)
	Number of engines	discrete	1	{2,3,4}
	Total discrete variables		2	
	Tail geometry	cat	2 levels	{T-tail, no T-tail}
	Engine position	cat	2 levels	{front or rear engines}
	Total categorical variables		2	
	Total relaxed variables		12	
subject to	$0.05 < \text{Static margin} < 0.1$	cont	2	
	Total constraints		2	

We are testing 3 methods, NSGA2, SEGO with Kriging and SEGO with KPLS over 10 DoEs and 200 iterations. As for the analytical test cases, Fig.6a shows that PLS method speeds up the optimization process at the start. However, the boxplots of the final results on Fig.6b show that, at the end, the KPLS and Kriging versions give similar results while KPLS uses 3 times less correlation length hyperparameters. The Kriging version starts slower but catches up with KPLS quickly and for every given budget Kriging gives better results than NSGA2. The best configuration is obtained after a SEGO+KPLS_4D optimization, the optimal result is given in Tab.3. The optimal aircraft geometries obtained

using each of the three methods and the baseline are plotted in Fig.7. As SEGO with Kriging is almost the same as the best, SEGO+KRG and SEGO+KPLS are grouped on the geometries. With KPLS we obtain a constraint margin of 0.500 and an objective of 16722.73 kg.

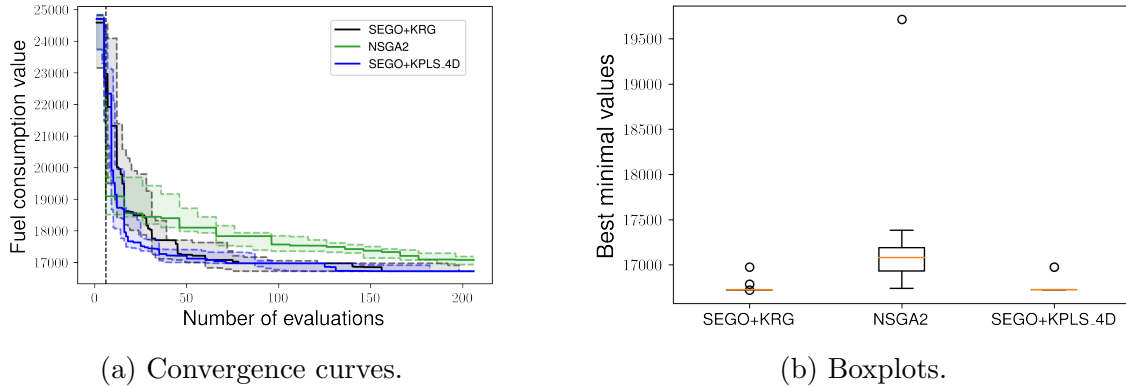


Figure 6: “CERAS” optimization results using a DoE of 5 points. The Boxplots are generated, after 200 iterations, using the 10 best points.

Table 3: “CERAS” Optimal aircraft configuration.

Name	Nature	Value
Fuel mass	cont	16722.55 kg
static margin	cont	0.0495
x position of MAC	cont	16.2825 m
Wing aspect ratio	cont	11
VT aspect ratio	cont	6
HT aspect ratio	cont	6
Wing taper ratio	cont	0.5099
Wing sweep angle	cont	30.0 °
Cruise altitude	discrete	36,000 ft
Tail geometry	cat	T-tail
Engine position	cat	Front engine
Number of engines	discrete	2

From an aircraft design point of view, the obtained optimized values can be surprising for some parameters, but are logical given the limits of current models in FAST-OAD:

- as the flight ceilings (buffeting, aerodynamics, ...) are not computed, the algorithm logically goes toward high altitudes,
- the mass of a strongly swept wing is not properly computed, and does not bring enough penalty on such configuration,
- similarly, the models do not compute the additional mass that should be required for a T-tail with such a large aspect ratio on vertical tail.

However, the capabilities of the proposed algorithm are very promising for a conceptual design stage where a lot of architecture choices are still undetermined, leading to a large

combinatorial problem. In that perspective, the ability of the method to capture the right trends regarding number of engines (integer) and their position (categorical) is a good perspective, that needs to be confirmed with an even more complex design case using many categorical and integer variables. The “DRAGON” hybrid aircraft case described below provides this feature.

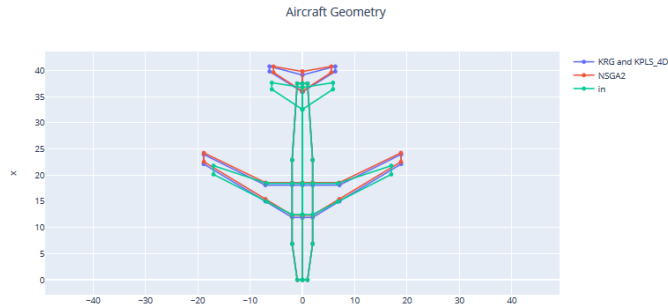


Figure 7: “CERAS” best configurations geometries.

4.3.2 DRAGON

The “DRAGON” aircraft concept in Fig.8 has been introduced by ONERA in 2019 [38] within the scope of the European CleanSky 2 program ¹ which sets the objective of 30% reduction of CO₂ emissions by 2035 with respect to 2014 state of the art. ONERA responded to this objective by proposing a distributed electric propulsion aircraft that improves the aircraft fuel consumption essentially by increasing the propulsive efficiency. Such efficiency increase is obtained through improvement of the bypass ratio by distributing a large number of small electric fans on the pressure side on the wing rather than having large diameter turbofans. This design choice avoids the problem associated with large under-wing turbofans and at the same time allows the aircraft to travel at transonic speed. Thus the design mission set for “DRAGON” is 150 passengers over 2750 Nautical Miles at Mach 0.78.



Figure 8: “DRAGON” aircraft concept.

The employment of a distributed propulsion comes at a certain cost; a turbo-electric propulsive chain is necessary to power the electric fans which brings additional complexity and weight. Typically, turboshafts coupled to electric generators are generating the electrical power on board the aircraft. The power is then carried to the electric fans through an electric architecture ensuring robustness to single component failure. This safety feature is obtained with redundant components as depicted in Fig.9. The baseline configuration

¹<https://www.cleansky.eu/technology-evaluator>

is two turboshafts, four generators, four propulsion buses with cross-feed and forty fans. This configuration was selected for an initial study as it satisfies the safety criterion. However it was not designed to optimize aircraft weight. The turboelectric propulsive chain being an important weight penalty, it is of particular interest to optimize the chain and particularly the number and type of each component, characterized by some discrete or particular values.

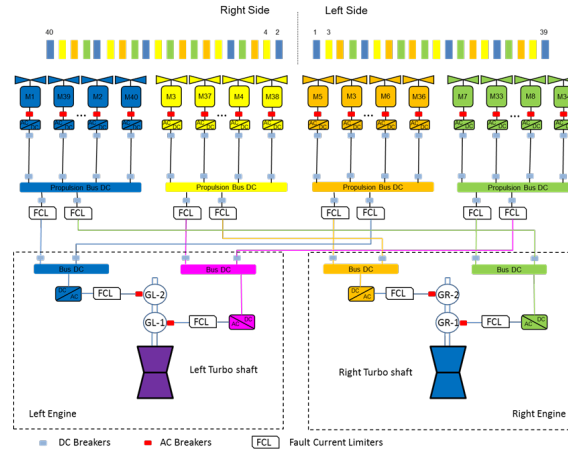


Figure 9: “DRAGON” turboelectric propulsive architecture.

To know how optimizing the fuel mass will impact the aircraft design, we are considering the problem described in Tab.4 where the Mean Average Chord is referred as MAC and the Takeoff Field Length as TOFL. We consider this problem as a test case involving a large number of mixed integer variables and four constraints. The problem to solve is a constrained optimization problem with 8 continuous design variables and 4 integer ones, for a total of 12 design variables.

Table 4: Definition of the “DRAGON” optimization problem.

	Function/variable	Nature	Quantity	Range
Minimize	Fuel mass	cont	1	
with respect to	Cruise Mach number	cont	1	[0.6, 0.8]
	TOFL for sizing	cont	1	[1800., 2500.] (m)
	Climb reserve for sizing	cont	1	[300., 800.](ft/min)
	Fan operating pressure ratio	cont	1	[1.05, 1.45]
	Span	cont	1	[30., 36.] (m)
	Angle for swept wing	cont	1	[20., 40.] (°)
	Wing taper aspect ratio	cont	1	[0.2, 0.5]
	Engine length to diameter ratio	cont	1	[1., 5.]
	Total continuous variables		8	
	Electric motors and fan number	discrete	1	{16,24,32,40}
	Electric cores number	discrete	1	{2,4,6,8}
	Generators per turbomachine	discrete	1	{1,2,4,8}
	Turbomachines number	discrete	1	{2,3,4}
	Total discrete variables		4	
	Total relaxed variables		12	
subject to	Climb time < 29 (s)	cont	1	
	TOFL < 2200 (m)	cont	1	
	Climb reserve > 500 (ft/min)	cont	1	
	Fan radius < 0.1 MAC (m)	cont	1	
	Total constraints		4	

To validate our method, we are comparing, in Fig.10, SEGO with KPLS (4 hyper-parameters) with its more expensive version using Kriging and with NSGA2. The best method after 50 iterations is the proposed one involving PLS regression as shown in the boxplots in Fig.10b. In Fig.10a, after 200 iterations, we still find that SEGO is better than NSGA2 and that the PLS helps for the convergence. After 50 iterations, SEGO seems to have almost converged whereas NSGA2 takes twice the number of iterations.

Like for “CERAS”, we find that the best configuration found with SEGO and Kriging with and without PLS regression are almost the same. The best configuration found is described in Tab.5.

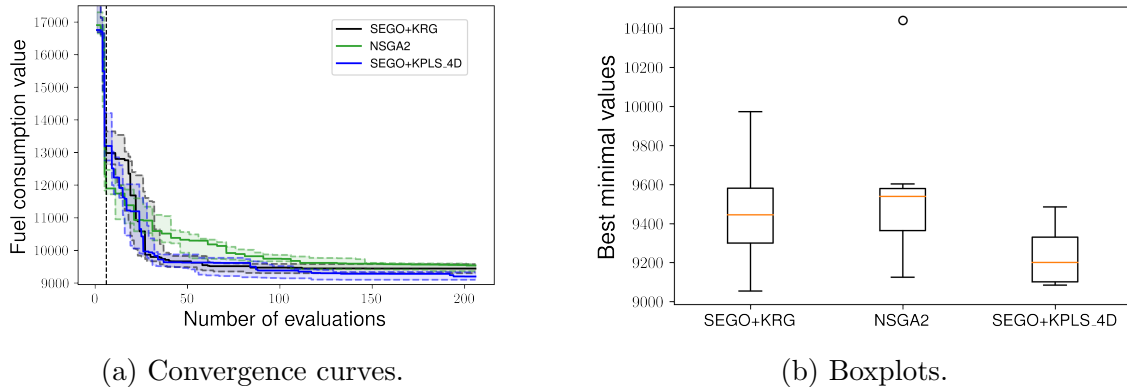


Figure 10: “DRAGON” optimization results using a DoE of 5 points. The Boxplots are generated, after 200 iterations, using the 10 best points.

Table 5: “DRAGON” Optimal aircraft configuration.

Name	Nature	Value
Fuel mass	cont	9055 kg
Climb time	cont	28.95 s
TOFL	cont	1865.67 m
Climb reserve at beginning of cruise	cont	583.165 ft/min
fan radius	cont	0.1 MAC
Cruise Mach number	cont	0.7233
TOFL for sizing	cont	2091.72 m
Climb reserve for sizing	cont	572.12 ft/min
Fan operating pressure ratio	cont	1.099
Span	cont	36.0 m
Angle for swept wing	cont	20.0°
Wing taper aspect ratio	cont	0.298
Engine length to diameter ratio	cont	5.0
Electric motors and fan number	discrete	40
Electric cores number	discrete	6
Generators per turbomachine	discrete	1
Turbomachines number	discrete	2

The observation of the optimization results from an aircraft design point of view leads to surprising conclusions, ultimately resulting in questioning the validity of the models utilized for the design. Among which:

- The optimal fan operating pressure ratio slightly less than 1.1 reveals a default in the coarse drag model used to assess the additional installation drag of distributed

electric propulsion.

- The sizing rules of the architecture is based on the default architecture and do not apply as easily to this new arrangement. Specifically connecting 40 motors to 6 cores of equal power is not straight forward.
- Additionally the safety analysis and cabling design for this new architecture have to be performed offline and may not be satisfying. Especially since the generator redundancy is removed by the optimization.

However, we can notice the clear direction taken by the optimizer to use the maximum number of motors with a low fan operating pressure ratio and constrained fan diameter. This also drives the motor length to diameter ratio to high values giving a clear indication for the choice of electric motor technology. A trade off is being made between propulsion efficiency, installation drag and space limitation which represents one of the challenge of DRAGON. The fact that such a trade off with continuous and discrete variables can be made is a great advantage for the study of aircraft with distributed electric propulsion. In future work, the flexibility of the optimization could allows us to view the architecture not only defined with discrete variables but provide a selection of architectures as categorical variables. This can help identify the most efficient architecture and most important redundancy or interconnection to install.

5 CONCLUSION AND PERSPECTIVES

To conclude, we have observed on both analytical and industrial cases than SEGO with KPLS is well-suited and efficient for a mixed integer high-dimensional constrained efficient global optimization problem. We have seen that continuous relaxation allows straightforward use of continuous GP but can be impractical as it increases the computational effort required to build the surrogate model. By using the PLS regression, it is possible to reduce the computational cost and makes the continuous relaxation affordable in practical contexts. This method has been applied to aircraft design optimization and contributed to the development of an MDA tool designing future aircraft configurations.

Future works will involve mixed integer Gaussian kernels for high-dimensional black-box problems to compare with this method. The number of components for PLS should be chosen from a global strategy and not beforehand and we expect KPLS to be useful for mixed integer Gaussian Kernels dimension reduction. For the Aircraft Design, the MDA for the “DRAGON” concept will soon be treated through overall aircraft design within FAST-OAD. The FAST-OAD Tool is still in development and will be made even more flexible and general.

ACKNOWLEDGEMENTS

This work is part of the activities of ONERA - ISAE - ENAC joint research group.

REFERENCES

- [1] E. Duriez and J. Morlier. Hale multidisciplinary design optimization with a focus on eco-material selection. In *Aerospace Europe Conference*, 2020.
- [2] A. Lambe and J. R. R. A. Martins. A unified description of mdo architectures. In *9th World Congress on Structural and Multidisciplinary Optimization*, 2011.

- [3] A. Lambe and J. R. R. A. Martins. Extensions to the design structure matrix for the description of multidisciplinary design, analysis, and optimization processes. *Structural and Multidisciplinary Optimization*, 46:273–284, 2012.
- [4] J. R. R. A. Martins and A. Lambe. Multidisciplinary design optimization: A survey of architectures. *AIAA Journal*, 51:2049–2075, 2013.
- [5] J. Mockus. *Bayesian approach to global optimization: theory and applications*, volume 37. Springer Science & Business Media, 2012.
- [6] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492, 1998.
- [7] J. Močkus. On bayesian methods for seeking the extremum. In *Optimization Techniques IFIP Technical Conference Novosibirsk*, 1974.
- [8] C. E. Rasmussen and J. Quiñonero-Candela. A unifying view of sparse approximate gaussian process regression. *Journal of Machine Learning Research*, 6:1939–1959, 2005.
- [9] A. Forrester, A. Sobester, and A. Keane. *Engineering Design via Surrogate Modelling: A Practical Guide*. Wiley, 2008. ISBN 9780470770795.
- [10] M. J. Sasena, P. Papalambros, and P. Goovaerts. Exploration of metamodeling sampling criteria for constrained global optimization. *Engineering Optimization*, 34:263–278, 2002.
- [11] J. Pelamatti, L. Brevault, M. Balesdent, E.-G. Talbi, and Y. Guerin. Efficient global optimization of constrained mixed variable problems. *Journal of Global Optimization*, 73:583–613, 2019. doi:10.1007/s10898-018-0715-1.
- [12] O. Roustant, E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn. Group kernels for gaussian process metamodels with categorical inputs. *arXiv e-prints*, 2018.
- [13] M. M. Zuniga and D. Sinoquet. Global optimization for mixed categorical-continuous variables based on gaussian process models with a randomized categorical space exploration step. *INFOR: Information Systems and Operational Research*, 58:1–32, 2020. doi:10.1080/03155986.2020.1730677.
- [14] S. Roy, W. A. Crossley, B. K. Stanford, K. T. Moore, and J. S. Gray. A mixed integer efficient global optimization algorithm with multiple infill strategy - applied to a wing topology optimization problem. In *AIAA Scitech 2019 Forum*, 2019. doi:10.2514/6.2019-2356.
- [15] E. C. Garrido-Merchán and D. Hernández-Lobato. Dealing with categorical and integer-valued variables in bayesian optimization with gaussian processes. *Neurocomputing*, 380:20–35, 2020.

- [16] D. Nguyen, S. Gupta, S. Rana, A. Shilto, and S. Venkatesh. Bayesian optimization for categorical and category-specific continuous inputs. In *AAAI-20 Technical Tracks*, 2020. doi:10.1609/aaai.v34i04.5971.
- [17] M. Halstrup. *Black-Box Optimization of Mixed Discrete-Continuous Optimization Problems*. PhD thesis, TU Dortmund, 2016.
- [18] R. Karlsson, L. Bliet, S. Verwer, and M. de Weerd. Continuous surrogate-based optimization algorithms are well-suited for expensive discrete problems. *arXiv e-prints*, 2020.
- [19] M. Bouhleb, N. Bartoli, R. Regis, A. Otsmane, and J. Morlier. Efficient global optimization for high-dimensional constrained problems by using the kriging models combined with the partial least squares method. *Engineering Optimization*, 50:2038–2053, 2018.
- [20] A. M. Bouhleb, N. Bartoli, A. Otsmane, and J. Morlier. Improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction. *Structural and Multidisciplinary Optimization*, 53:935–952, 2016.
- [21] C. David, S. Delbecq, S. Defoort, P. Schmollgruber, E. Benard, and V. Pommier-Budinger. From FAST to FAST-OAD: An open source framework for rapid overall aircraft design. *IOP Conference Series: Materials Science and Engineering*, 1024:012062, jan 2021. doi:10.1088/1757-899x/1024/1/012062.
- [22] S. H. Kim and F. Boukouvala. Surrogate-based optimization for mixed-integer non-linear problems. *Computers & Chemical Engineering*, 140, 2020.
- [23] D. Duvenaud. *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge, 2014.
- [24] R. J. Rossi. *Mathematical statistics: an introduction to likelihood based inference*. John Wiley & Sons, 2018.
- [25] R. Priem, N. Bartoli, Y. Diouane, and A. Sgueglia. Upper trust bound feasibility criterion for mixed constrained bayesian optimization with application to aircraft design. *Aerospace Science and Technology*, 105:105980, 2020.
- [26] N. Bartoli, T. Lefebvre, S. Dubreuil, R. Olivanti, R. Priem, N. Bons, J. R. R. A. Martins, and J. Morlier. Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design. *Aerospace Science and technology*, 90:85–102, 2019.
- [27] D. Golovin, B. Solnik, S. Moitra, G. Kochanski, J. Karro, and D. Sculley. Google vizier: A service for black-box optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 1487–1495, New York, NY, USA, 2017. Association for Computing Machinery. doi:10.1145/3097983.3098043. URL <https://doi.org/10.1145/3097983.3098043>.
- [28] H. Wold. Soft modelling by latent variables: The non-linear iterative partial least squares (nipals) approach. *Journal of Applied Probability*, 12:117–142, 1975.

- [29] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6: 182–197, 2002.
- [30] J. Blank and K. Deb. Pymoo: Multi-objective optimization in python. *IEEE Access*, 8:89497–89509, 2020.
- [31] M. A. Bouhleb, J. T. Hwang, N. Bartoli, R. Lafage, J. Morlier, and J. R. R. A. Martins. A python surrogate modeling framework with derivatives. *Advances in Engineering Software*, 135, 2019. doi:10.1016/j.advengsoft.2019.03.005.
- [32] M. J. D. Powell. *A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation*, pages 51–67. Springer Netherlands, 1994.
- [33] T. P. Runarsson and X. Yao. Search biases in constrained evolutionary optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 35:233–243, 2005.
- [34] P. E. Gill, W. Murray, and M. A. Saunders. Snopt: An sqp algorithm for large-scale constrained optimization. *SIAM Review*, 47:99–131, 2005.
- [35] R. L. Riche and V. Picheny. Revisiting bayesian optimization in the light of the coco benchmark. 2021.
- [36] J. Pelamatti, L. Brevault, M. Balesdent, E.-G. Talbi, and Y. Guerin. *High-Performance Simulation-Based Optimization*, chapter Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, pages 189–224. Springer International Publishing, 2020.
- [37] C. Vanaret. A global method for mixed categorical optimization with catalogs. *arXiv e-prints*, 2021.
- [38] P. Schmollgruber, C. Döll, J. Hermetz, R. Liaboef, M. Ridel, I. Cafarelli, O. Atinault, C. François, and B. Paluch. Multidisciplinary Exploration of DRAGON: an ONERA Hybrid Electric Distributed Propulsion Concept. In *AIAA Scitech 2019*, 2019.