



Multi-fidelity bayesian optimization using model-order reduction for viscoplastic structures

Stéphane Nachar, Pierre-Alain Boucard, David Néron, Christian Rey

► To cite this version:

Stéphane Nachar, Pierre-Alain Boucard, David Néron, Christian Rey. Multi-fidelity bayesian optimization using model-order reduction for viscoplastic structures. *Finite Elements in Analysis and Design*, 2020, 176, pp.103400. <10.1016/j.finel.2020.103400>. <hal-02919419>

HAL Id: hal-02919419

<https://hal.science/hal-02919419v1>

Submitted on 23 Aug 2022

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.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Multi-fidelity bayesian optimization using model-order reduction for viscoplastic structures

Stéphane Nachar^a, Pierre-Alain Boucard^a, David Néron^{a,*}, Rey Christian^b

^a *Université Paris-Saclay, ENS Paris-Saclay, CNRS,
LMT - Laboratoire de Mécanique et Technologie, 94235, Cachan, France*
^b *SAFRAN Tech
CS 80112, 78772 Magny-les-Hameaux, France*

Abstract

One of the main issues when dealing with the numerical optimization of mechanical structures is the balance between computation time and model accuracy. The work presented herein aims at accelerating global optimization by using the framework of Bayesian optimization on a quantity of interest together with multiple levels of fidelity. These multi-fidelity data are generated from a model-order reduction framework: the LATIN Proper Generalized Decomposition. Within this framework, a reduced-order basis is generated on-the-fly and re-exploited to reduce the computational cost of observations. This strategy is illustrated on two elasto-viscoplastic test-cases for which significant speedups can be observed.

Keywords: Bayesian optimization, Multi-fidelity kriging, Reduced-order models

Introduction

Structural optimization with high-fidelity computer experiments has a long tradition [1, 2, 3], using solver responses to build and optimize a quantity of interest (QoI) y for a given design space \mathcal{D} . In the recent years, there has been
5 a growing interest in Bayesian optimization [4], which has proven to be effective for several difficult reference functions [5]. Bayesian optimization typically works by modeling the costly objective function y by a cheaper surrogate one \hat{y} , build as a Gaussian process regressor [6, 7, 8]. Roughly, the uncertainty
10 of the approximation is used to construct an acquisition function which estimates the probability of improving the global optimum obtained and quantifies the exploitation-exploration trade-off. This acquisition function is maximized to determine the next query point. This approach tries to reduce the number of calculated points. This technique is commonly used in machine learning to optimize parameters of the model [9, 10].

*Corresponding author

Email address: david.neron@ens-paris-saclay.fr (David Néron)

15 When dealing with mechanical problems, each point in the design space can lead to several hours of computation in order to obtain the converged value of the QoI. However, if the point is far from the accurate optimum zone, the QoI might have been computed with an unnecessary accuracy. A solution to reduce the duration of the optimization process is to use further data sources and to
 20 generate a multi-fidelity metamodel. Low-fidelity points (obtained from low-cost calculations) are performed to estimate the region of the optimum. After a certain number of calculations, low-fidelity and high-fidelity information are both used for the next query points to ensure the obtained optimum, and to estimate the optimal additive correction on the low-fidelity data previously
 25 computed. Further developments have already made on this subject [11, 12, 13] and have been applied on structural optimization [14, 15].

Model reduction techniques constitute an efficient way to generate low-fidelity data by seeking the solution of the problem in a reduced-order basis (ROB), whose dimension is much lower than the original vector space. Using a
 30 [database of reduced order models](#) as a low-fidelity data is quite recent [16] and uses a *posteriori* approach which usually consists of defining this ROB by the decomposition of the solution of a surrogate model relevant to the initial model. *A priori* methods like Proper Generalized Decomposition (PGD) [17, 18] follow a different path by building progressively an approximation of the solution in a
 35 separated-variable form, without assuming any basis. This method seems more appropriate when no *a priori* information on the design space is available and was hitherto not tested in this multi-fidelity process.

The purpose of this paper is to show how to couple multi-fidelity kriging and model-order reduction to speedup the global optimization of the QoI. This
 40 strategy has already been performed for virtual charts generation [19]. Herein, optimization is made by calculating the QoI from the PGD modes generated on-the-fly with the LATIN-PGD framework [20] to deal with nonlinear problems. This framework generates the solution \mathbf{s} from an initial solution \mathbf{s}_0 by computing at each step a spatio-temporal correction in a separated variable
 45 form by PGD technique. Since the method characterizes the solution over its entire domain at each iteration, it is possible to compute an approximation of the QoI without going to convergence. This property is particularly suitable for a global optimization process requiring an exploration phase in the design space. Then, low-fidelity fields are obtained by stopping the solver before convergence,
 50 and high-fidelity information is obtained with converged fields. In addition, the solver ability to reuse information from previously calculated modes of the reduced-order basis is exploited using what is called the "multiparametric strategy" [21, 22]. This accelerates more and more computations as the process is moving forward. The coupling of these methods is done on viscoplastic test
 55 cases and a $8.4\times$ speedup is obtained, allowing a first demonstration of this approach.

Summary

We summarize herein the strategy in its whole and the related improvements so that the reader could find, as well as the main guidelines to follow the in-

60 teraction and the motivation of the various developments. First, the two main ingredients, that are multi-fidelity Bayesian optimization and model reduction methods, are recalled. On one hand, multi-fidelity Bayesian optimization is based on the generation of a metamodel using two different sources of information: firstly, only low-fidelity data are used to generate a first metamodel which
65 is then enriched by the simultaneous use of high and low-fidelity data (typically examples of low/high-fidelity data are: numerical simulation/real tests, coarse mesh/fine mesh, coarse model/accurate model ...). On the other hand, model reduction methods rely on the a priori or a posteriori construction of separated variables representations of the solution. The functions of these representations
70 are classically called modes. The main novelty introduced in this paper is the coupling of these two ingredients, and in a more detailed way, the improvements concerns three aspects:

- the first aspect concerns the enhancement of the multi-fidelity kriging strategy which led to a modification of the Evofusion method with the
75 taking into account of the variance on low-fidelity data;
- the second aspect concerns the use of LATIN-PGD as a tool for generating low and high-fidelity data and its coupling with Bayesian optimization. Low-fidelity data are generated by stopping the process of generating separate variable representations before convergence. High-fidelity data are
80 obtained at convergence of the method (low/high-fidelity data correspond to a truncated number of modes/converged number of modes);
- the third aspect is the use of a LATIN-PGD property which allows, from solutions generated for different sets of optimization parameters (that constitute a database of reduced-order models), to reuse the separate variable
85 representations to reduce the calculation costs of the solution associated with a new set of parameters.

Finally, in order to be able to deal with quasi-industrial test cases, an object-oriented implementation using Matlab/OpenMP (not detailed here) was carried out.

90 The development of the article is as follows. The bayesian optimization strategy for multi-fidelity data is described on Section 1, with a focus on multi-fidelity kriging methods, a modified Evofusion method and a non-usual acquisition function which take into account multi-fidelity data. Observations used in this optimization process are computed thanks to the LATIN-PGD framework presented
95 in Section 2. This framework is used in the context of viscoplastic problems and exploit the multiparametric strategy to start computation from fields interpolated from previously calculated solutions. This multi-fidelity bayesian optimization process using the LATIN-PGD framework gives significant speedup and results of this strategy is visible on Section 3 with a presentation of the test
100 cases.

1. Bayesian optimization strategy for multi-fidelity data

1.1. Bayesian optimization

Let us consider here a spatio-temporal mechanical problem defined by some input parameters \underline{x} in a design space \mathcal{D} . For each value $\underline{x} \in \mathcal{D}$, the solution of the problem allows to compute a given quantity of interest (QoI), denoted $y(\underline{x}) \in \mathbb{R}$, which is assumed to be scalar. The associated objective function $\underline{x} \mapsto y(\underline{x})$ is considered as a black-box function without analytical expression of y nor its derivatives.

In the case of minimization, this problem is formulated concisely as follows:

$$\text{Find } (\underline{x}^*, y^*) \in \mathcal{D} \times \mathbb{R}, \text{ solution of } \underline{x}^* = \underset{\underline{x} \in \mathcal{D}}{\operatorname{argmin}} y(\underline{x}) \text{ and } y^* = y(\underline{x}^*) \quad (1)$$

As in this preliminary study, we focus on unconstrained problems, global optimisation is presented without constraints. Indeed, In this context, literature reports that Bayesian optimization is a successful option.

A non-exhaustive list of global optimization methods are presented in [23]. Standard gradient-based optimizers and heuristics are particularly appreciated for their convergence properties but need many observed points. Among them, Bayesian optimization is commonly used in the case of black-box expensive optimization [24, 5, 25]. The black-box function y is modeled herein as a Gaussian process \hat{y} conditioned on observations, using gaussian process regression [8, 6, 7]. Initial observations are chosen from near-random sampling method like Latin Hypercube Sampling (LHS) [26] to prevent large areas without points. If needed, the training set is sequentially enriched by the global maximum of an acquisition function which estimates the probability of improving the global optimum. This function is build from the posterior distribution of \hat{y} and is a cheap function which can be globally optimized, unlike y .

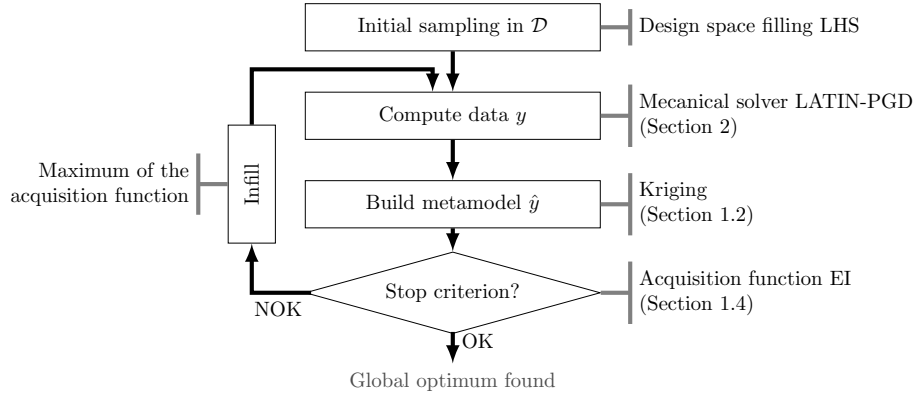


Figure 1: Classical bayesian optimization

125 1.2. Gaussian Process Regression (kriging)

N points \underline{x}_i are observed with $\forall i \in \llbracket 1, N \rrbracket$, $y_i = y(\underline{x}_i)$. The set of the N observed points is denoted as \mathbf{X} and the set of their corresponding responses \underline{y} . They are used to create the training set $(\mathbf{X}, \underline{y})$. The objective is to build a regressor from this training set. In the case of linear regression methods, the output $y(\underline{x})$ can be mapped by a linear combination of the input \hat{y} :

$$\hat{y}(\underline{x}) = \underline{x}^T \underline{\beta} + Z(\underline{x}) \quad (2)$$

$$\forall i \in \llbracket 1, N \rrbracket, y_i = \underline{x}_i^T \underline{\beta} + Z_i \quad (3)$$

The term $\underline{x}^T \underline{\beta}$ represents the scalar product between the input vector and the weights associated $\underline{\beta}$. The term Z is the estimated residual error between the linear prediction and the true response.

Kriging [6, 7], also called Gaussian process regression [8], can be presented as an extension to linear regression and can also be understood as a form of Bayesian inference [27] by considering Z as a Gaussian process:

$$\hat{y}(\underline{x}) = \underline{f}(\underline{x})^T \underline{\beta} + Z(\underline{x}) \quad (4)$$

The covariance between any two samples is defined by the covariance function (or kernel) $k(\underline{x}, \underline{x}'; \underline{\theta})$, (hyper)parametrized by $\underline{\theta}$. The Gaussian process \hat{y} is conditional on observations, which means that by assuming the responses y_i are random variables, a joint distribution between the prediction and the observations is defined by equation (5):

$$\begin{pmatrix} \hat{y}(\underline{x}) \\ \underline{y} \end{pmatrix} \sim N \left(\begin{pmatrix} \underline{f}(\underline{x})^T \underline{\beta} \\ \mathbf{F}^T \underline{\beta} \end{pmatrix}, \begin{pmatrix} \mathbf{1} & \underline{k}^T(\underline{x}) \\ \underline{k}(\underline{x}) & \mathbf{C} \end{pmatrix} \right) \quad (5)$$

130 with \mathbf{F} the regression matrix such as $\mathbf{F}_{ij} = f_i(\underline{x}_j)$, \mathbf{C} the covariance matrix between observed data such as $\mathbf{C} = k(\underline{x}_i, \underline{x}_j; \underline{\theta})$ and $\underline{k}(\underline{x})$ the covariance vector between observed and predicted data such as $\underline{k}_j(\underline{x}) = k(\underline{x}_j, \underline{x}; \underline{\theta})$.

The predictive distribution is defined by $[\hat{y}(\underline{x}) | \hat{y}(\mathbf{X}) = \underline{y}, \underline{\beta}, \underline{\theta}]$ where notation $[A|B]$ stands for the distribution of A conditionally to B . So, conditionally to $\underline{\beta}$ and $\underline{\theta}$, the distribution $\underline{x} \rightarrow [\hat{y}(\underline{x}) | \hat{y}(\mathbf{X}) = \underline{y}, \underline{\beta}, \underline{\theta}]$ is Gaussian $N(\mu(\underline{x}), \sigma^2(\underline{x}))$ with:

$$\begin{cases} \mu(\underline{x}) = \underline{f}(\underline{x})^T \underline{\beta} + \underline{k}(\underline{x})^T \mathbf{C}^{-1} (\underline{y} - \mathbf{F} \underline{\beta}) \\ \sigma^2(\underline{x}) = (k(\underline{x}, \underline{x}) - \underline{k}(\underline{x})^T \mathbf{C}^{-1} \underline{k}(\underline{x})) \end{cases} \quad (6)$$

135 Kriging means that $\underline{x} \mapsto \mu(\underline{x})$ is the surrogate model that is used to approximate the objective function y . The kriging variance $\underline{x} \mapsto \sigma^2(\underline{x})$ represents the model mean squared error. The great interest of using kriging is the definition of the kriging variance as an error indicator of the metamodel on the whole design space. A 95% confidence interval can thus be computed corresponding to $\mu(\underline{x}) \pm 3\sigma(\underline{x})$ as shown in figure 2.

Parameters $\underline{\beta}$ and hyperparameters $\underline{\theta}$ can be obtained by integrating the parameter posterior distributions, or by maximizing the likelihood function associated with the Gaussian process. Here, the Leave-One-Out (LOO) cross-validation method is used. The idea is to generate n sub-training sets by extracting the observation (\underline{x}_i, y_i) , $i \in \llbracket 1, n \rrbracket$ and using this one as a validation set to monitor the performance. Hyperparameters are used to build metamodel liked with each sub-training sets and the LOO predictive log probability when leaving the training case i is computed :

$$\log p(y_i | \mathbf{X}, \underline{y}_{-i}, \underline{\theta}) = \frac{1}{2} \log(\sigma_i^2) + \frac{(y_i - \mu_i)^2}{\sigma_i^2} \quad (7)$$

where notation \underline{y}_{-i} means all targets except number i , and μ_i and σ_i^2 are respectively the mean and the variance of the metamodel build without the training case i on \underline{x}_i . Accordingly, the LOO log predictive probability is :

$$L_{LOO} = \sum_{i=1}^n \log p(y_i | \mathbf{X}, \underline{y}_{-i}, \underline{\theta}) \quad (8)$$

The objective is to minimize this mesure to obtain the optimal hyperparameters. It can be done by using bayesian optimization on hyperparameters [9].

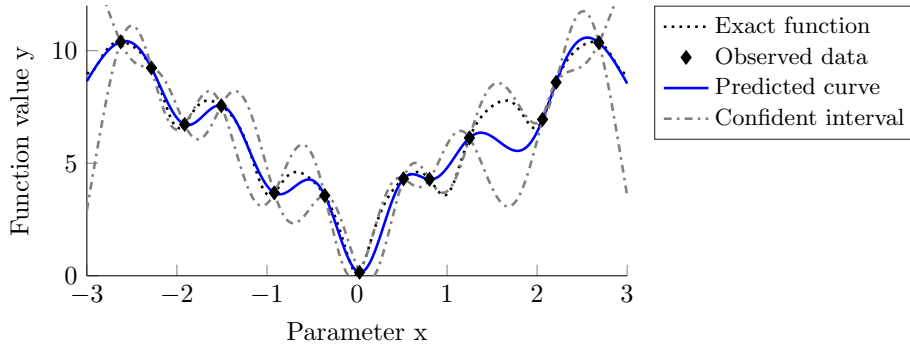


Figure 2: Kriging based on observations of the Ackley function
 $(x, y) \mapsto -20 \exp \left(-0.2 \sqrt{0.5(x^2 + y^2)} \right) - \exp(0.5(\cos 2\pi x + \cos 2\pi y)) + 20$

140 1.3. Data fusion with enhanced multi-fidelity kriging

A way to reduce the design step is to reuse this process for kriging by considering multiple solvers from the coarse one to the finest one and merge data observations. The reader can refer to [28] for a review of surrogate modeling and multi-fidelity approaches.

145 In the context of variable-fidelity modeling, a possibility to take into account low-fidelity observations $(\mathbf{X}_1, \underline{y}_1)$ and high-fidelity one $(\mathbf{X}_2, \underline{y}_2)$ is to use the cokriging approach. Cokriging modifies the covariance matrix $\underline{\mathbf{C}}$ (see Equation

(5)) and needs to establish a relation between the high-fidelity and low-fidelity data. Some hypotheses are assumed to compute the cross-correlation between these data and can be found in [13, 29, 30, 31]. All these methods modify the covariance matrix to take into account the solver quality and, even they allow expert judgement, are intrusive.

Other methods based on recursive metamodeling exist: Hierarchical Kriging [32] which replaces regression function f by a Gaussian process conditioned to low-fidelity data \hat{y}_1 , and Evofusion [11]. Evofusion method is a simple one and has given good results in a multi-fidelity benchmark [33]. Details of Evofusion algorithm can be seen in Algorithm 1 where y_{corr} is defined by:

$$y_{corr} = y_2 - \hat{y}_1(\mathbf{X}_2) \quad (9)$$

Algorithm 1: Evofusion Algorithm

Input: low-fidelity and high-fidelity observations $(\mathbf{X}_1, \underline{y}_1)$, $(\mathbf{X}_2, \underline{y}_2)$
 Build a low-fidelity metamodel with only low-fidelity data
 $(\mathbf{X}_1, \underline{y}_1) \longrightarrow \hat{y}_1$
 Compute the gap between the low-fidelity metamodel \hat{y}_1 and high-fidelity observations \underline{y}_2 on high-fidelity points \mathbf{X}_2 :
 $y_{corr} = y_2 - \hat{y}_1(\mathbf{X}_2)$
 Build a correction metamodel with correction data $(\mathbf{X}_2, \underline{y}_{corr}) \longrightarrow \hat{y}_{corr}$
 Modify low-fidelity observations with the correction metamodel:
 $\underline{y}_{1c} = \underline{y}_1 + \hat{y}_{corr}(\mathbf{X}_1)$
 Build the fused metamodel with corrected data and high-fidelity data:
 $(\mathbf{X}_1, \underline{y}_{1c}) \cup (\mathbf{X}_2, \underline{y}_2) \longrightarrow \hat{y}$
Output: Fused metamodel: \hat{y}

A modified version will be used herein. Corrected low-fidelity points were interpolated on the classical method. The correction made by the \hat{y}_{corr} error metamodel can be somewhat inaccurate due to a lack of information. Adding variance on these points will release the exploration component of the acquisition function (detailed in the next section) around LF points which are far away from a high-fidelity observation (and therefore the correction is potentially false in this area).

So the improvement is to consider an estimated variance σ_ϵ on low-fidelity data. This noise can be quantified by the variance associated with the error metamodel. Indeed, the sum of the random variables \tilde{y}_1 and $\hat{y}_c(\mathbf{X}_1)$ gives an associated variance $\sigma_\epsilon(\mathbf{X}_1)$ such that

$$\sigma_\epsilon(\mathbf{X}_1)^2 = \sigma(\tilde{y}_1)^2 + \sigma(\hat{y}_c(\mathbf{X}_1))^2 + 2 \text{Cov}(\tilde{y}_1, \hat{y}_c(\mathbf{X}_1)) \quad (10)$$

The independence of the random variables is assumed: $\text{Cov}(\tilde{y}_1, \hat{y}_c(\mathbf{X}_1)) = 0$. In addition, no measurement noise is considered on the LF observations initially: $\sigma(\tilde{y}_1) = 0$. We will therefore consider in the following $\sigma_\epsilon(\mathbf{X}_1) = \sigma(\hat{y}_c(\mathbf{X}_1))$.

This analysis is also appropriate for HF data with $\sigma(\hat{y}_c(\mathbf{X}_2)) = 0$ at these points.

Figure 3 presents the method through the next functions:

$$\begin{aligned} y_1 : x &\mapsto 0.5(6x - 2)^2 + \sin(12x - 4) + 10(x - 0.5) - 5 \\ y_2 : x &\mapsto 2y_1(x) - 20x + 20 \end{aligned} \quad (11)$$

Note that these two examples of function are simple one-variable functions which have been used by several authors to compare multi-fidelity approaches (*e.g.*, [34, 35]).

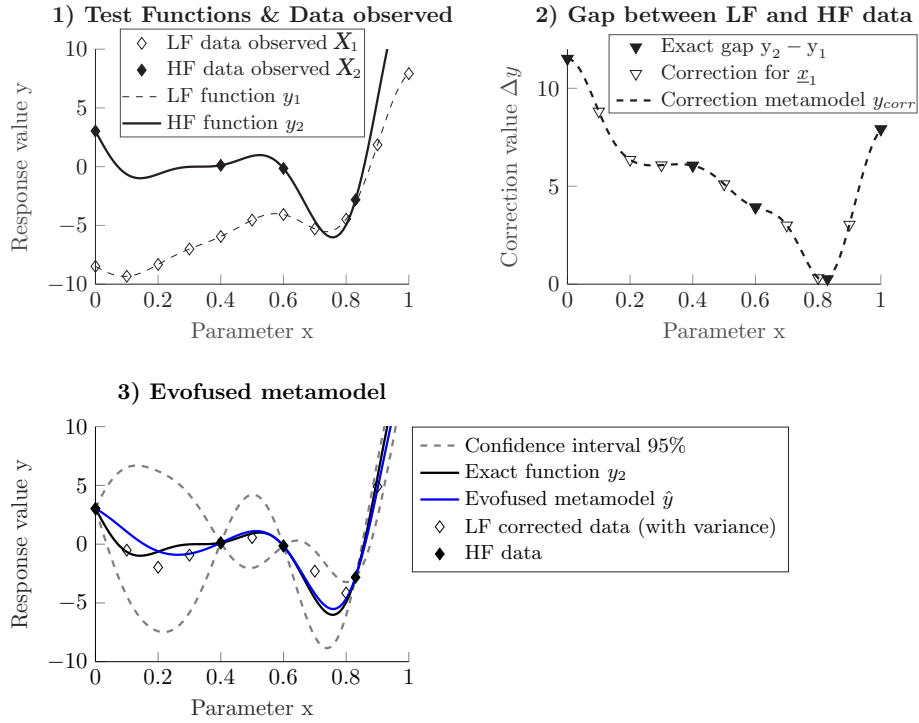


Figure 3: Illustration of the Evofusion method

1.4. Adapted acquisition function for multi-fidelity kriging

The role of the acquisition function is to guide the search for the optimum by considering a good exploitation-exploration trade-off. The new point will be around already known optimum zones (exploitation), or in wide unobserved areas (exploration). To do that, the expected improvement function [4] is the most widely used in the literature. It consists in calculating the average of the improvement function by integrating the density function of the improvement function I:

$$I(\underline{x}) = \langle \hat{y}(\underline{x}) - y^* \rangle_+ \quad (12)$$

170 with \hat{y} the Gaussian process $N(\mu(\underline{x}), \sigma(\underline{x}))$ which models y and y^* the maximum of observed values.

The new query point is found by maximizing the expected improvement:

$$\underline{x} = \operatorname{argmax} \mathbb{E}[\mathbf{I}(\underline{x})] \quad (13)$$

The likelihood of improvement $\mathbb{E}[\mathbf{I}(\underline{x})]$ can be computed by integrating the density function of \mathbf{I} :

$$\mathbb{E}[\mathbf{I}(\underline{x})] = \int_{I=0}^{I=\infty} I \frac{1}{\sqrt{2\pi}\sigma(\underline{x})} \exp\left(-\frac{(\mu(\underline{x}) - y^* - I)^2}{2\sigma^2(\underline{x})}\right) dI \quad (14)$$

and can be evaluated analytically:

$$\mathbb{E}[\mathbf{I}(\underline{x})] = (\mu(\underline{x}) - y^*)\Phi\left(\frac{\mu(\underline{x}) - y^*}{\sigma(\underline{x})}\right) + \sigma(\underline{x})\phi\left(\frac{\mu(\underline{x}) - y^*}{\sigma(\underline{x})}\right) \quad (15)$$

where Φ and ϕ denote respectively the probability density function and the cumulative distribution function of the standard normal distribution. The first term in analytical expression (15) is the exploration term, whose contributions
175 will be significant in the unobserved areas of \mathcal{D} . The second one is the exploitation term whose contributions will be significant in area around the observed optimum.

In the case of kriging without variance associated with observations, $\mathbf{EI}(\underline{x}) = 0$. As the function is positive, the points already sampled will never be recomputed. For this reason, it is essential that the low-fidelity data have an associated variance. In addition, the observed minimum is not necessarily accurate if it is derived from low-fidelity data. To address this last point, it is possible to ignore the observed minimum y^* but the minimum of the mean or a specific quantile of the Gaussian process at the observed points, noted μ^* [36]. This method is called “Expected improvement with plug-in”:

$$\mathbf{PI}(\underline{x}) = (\mu(\underline{x}) - \mu^*)\Phi\left(\frac{\mu(\underline{x}) - \mu^*}{\sigma(\underline{x})}\right) + \sigma(\underline{x})\phi\left(\frac{\mu(\underline{x}) - \mu^*}{\sigma(\underline{x})}\right) \quad (16)$$

This method does not take into account the noise of the future observation: the improvement is defined and its expectation is calculated as if the next evaluation
180 would be deterministic. Nevertheless, it already gives better results than \mathbf{EI} [36].

2. Fast goal-oriented QoI computations using the LATIN-PGD framework

2.1. Reference problem

For the sake of simplicity, let us consider the quasi-static and iso-thermal
185 evolution of a viscoelastic structure defined over the time-space domain $I \times \Omega$, under the assumption of small perturbations. The continuum body is submitted to surface forces \underline{F}_d on a first part of its boundary $\partial_2\Omega$ and to body forces \underline{f}_d in its

interior Ω . The continuum body is also submitted to prescribed displacements \underline{U}_d over the complementary part of $\partial\Omega$, denoted by $\partial_1\Omega$ (with $\partial_1\Omega \cap \partial_2\Omega = \emptyset$).

Movement within the structure results in a displacement field with respect to the initial configuration $\underline{U} : I \times \Omega \mapsto \mathbb{R}^3$. Infinitesimal strain theory is assumed and allows the linearization of the strain tensor $\underline{\varepsilon} = \underline{\varepsilon}(\underline{U}) = \frac{1}{2} [\nabla \underline{U} + \nabla^T \underline{U}]$. Internal forces are expressed by the stress tensor $\underline{\sigma}$ satisfying the following boundary value problem:

$$\operatorname{div}(\underline{\sigma}) + \underline{f}_d = \underline{0} \text{ on } \Omega, \quad \underline{\sigma} \cdot \underline{n} = \underline{F}_d \text{ on } \partial_2\Omega, \quad \underline{U} = \underline{U}_d \text{ on } \partial_1\Omega \quad (17)$$

The nonlinear behavior laws are taken into account with the equation (18) and will be in details on the following section, but formally, one has:

$$\forall (M, t) \in \Omega \times I, \underline{\sigma}(M, t) = \mathcal{B}(\underline{\varepsilon}(M, \tau), \tau \in [0, t]) \quad (18)$$

In the context of elasto-viscoplasticity, solving the boundary value problem given by (17) is equivalent to minimizing the energy functional $\mathcal{E} : \mathcal{U} \mapsto \mathbb{R}$ defined as:

$$\mathcal{E}(\underline{U}) = \int_{\Omega} \mathcal{B}(\underline{\varepsilon}) : \underline{\varepsilon} \, d\Omega - \int_{\Omega} \underline{f}_d \cdot \underline{U} \, d\Omega - \int_{\partial_2\Omega} \underline{F}_d \cdot \underline{U} \, dS \quad (19)$$

where \mathcal{U} is a functional space of the form:

$$\mathcal{U} = \{ \underline{U} \in H^{1,p}(\Omega) / \mathcal{E}(\underline{U}) < +\infty, \underline{U} = \underline{U}_d \text{ sur } \partial_1\Omega \}$$

190 with $p \geq 1$ and $H^{1,p}(\Omega)$ the Sobolev space of functions in $L^p(\Omega, \mathbb{R}^3)$ with partial derivatives in $L^p(\Omega, \mathbb{R}^3)$.

A weak formulation of the boundary value problem given by (19) can be written *i.e.* find $\underline{U} \in \mathcal{U}$ such that

$$\forall t \in I, \forall \underline{U}^* \in \mathcal{U}^*, \mathcal{R}(\underline{U}(M, t); \underline{U}^*) = 0 \quad (20)$$

with \mathcal{U}^* the space of functions in \mathcal{U} that vanish on the boundary $\partial_1\Omega$ with $\underline{U}_d = \underline{0}$, and the residual function \mathcal{R} is defined by:

$$\int_{I \times \Omega} \mathcal{B}(\underline{\varepsilon}(\underline{U})) : \underline{\varepsilon}(\underline{U}^*) \, d\Omega dt - \int_{I \times \Omega} \underline{f}_d \cdot \underline{U}^* \, d\Omega dt - \int_{I \times \partial_2\Omega} \underline{F}_d \cdot \underline{U}^* \, ds dt \quad (21)$$

The resolution of this nonlinear problem will be done using the LATIN method explained on section 2.2.

2.2. The Chaboche elasto-viscoplasticity behavior law

Unified viscoplastic framework previously presented in [37] is considered. In this constitutive law, strain $\underline{\varepsilon}$ is split between elastic reversible strain $\underline{\varepsilon}_e$ and plastic strain $\underline{\varepsilon}_p$ with $\underline{\varepsilon} = \underline{\varepsilon}_p + \underline{\varepsilon}_e$. Stress is driven by $\underline{\sigma} = \mathbf{C}\underline{\varepsilon}_e$ where \mathbf{C} is the Hooke operator. Plastic behavior zone appears when stress goes over an elastic limit f . This limit is usually represented by an ellipsis in deviatoric stress principal component space also called yield surface. Size and origin of

the ellipsis are driven respectively by isotropic hardening R (drag effect) and a unique linear kinematic hardening X :

$$f = (\boldsymbol{\sigma} - R)_{eq} - \sigma_0 \quad (22)$$

195 where $J_2 = (\boldsymbol{\sigma} - R)_{eq}$ is the Von Mises equivalent stress and $\sigma_0 = \sigma_y - R$ is the yield surface size. We define the plastic strain p and the primal field linked with X , $\boldsymbol{\alpha}$. Primal fields $(p, \boldsymbol{\alpha})$ are associated with (R, X) .

The Norton-Hoff law drives the plastic strain p :

$$\dot{p} = \left\langle \frac{f}{k} \right\rangle_+^N \quad (23)$$

where k , N are material dependant scalars, $\langle \cdot \rangle_+$ are Macaulay brackets.

State laws are:

$$\boldsymbol{\sigma} = \mathbf{C}\boldsymbol{\varepsilon}_e \quad (24)$$

$$X = \frac{2}{3}C\boldsymbol{\alpha} \quad (25)$$

$$R = R_\infty(1 - e^{-bp}) \quad (26)$$

where C, R_∞, b are material dependant scalars.

Defining a pseudo-dissipation potential F , the evolution equations can be expressed:

$$\frac{d}{dt} \begin{bmatrix} \boldsymbol{\varepsilon}_p \\ -\boldsymbol{\alpha} \\ -p \end{bmatrix} = \left\langle \frac{f}{k} \right\rangle_+^N \begin{bmatrix} \sqrt{\frac{3}{2}}\mathbb{N} \\ -\sqrt{\frac{3}{2}}\mathbb{N} + \frac{3\gamma}{2C}X \\ -1 \end{bmatrix} \quad (27)$$

with \mathbb{N} the unitary normal vector defined as

$$\mathbb{N} = \sqrt{\frac{3}{2}} \frac{\boldsymbol{\sigma}_D - X}{(\boldsymbol{\sigma}_D - X)_{eq}}, \quad (\mathbb{N})_{eq} = 1 \quad (28)$$

200

2.3. LATIN nonlinear solver for goal-oriented data

The LATIN framework [20] is a powerful method to obtain the linear equilibrium of the structure with the respect of nonlinear behavior laws. Here we recall the main principles of the method, the details of which can be found in [38].

205

Equations are separated into two sub-problems which define the two manifolds A_d and Γ . In our case, A_d contains solutions which satisfy kinematic and static admissibility and state laws. These linear equations are global in time and space. On the other hand, the variety Γ contains the nonlinear behaviour equations. $A_{d,0}$ is similar to A_d with boundary conditions set equal to zero. Unlike Newton-like techniques, this approach provides a complete solution at each

210

225 Steps of the method are briefly described below and figure 5 gives a schematic representation.

- **Elastic initialization:** The algorithm is classically initialized by computing the elastic solution of the problem, such that $\mathbf{s}_0 \in A_d$, but \mathbf{s}_0 can be enriched by adding $A_{d,0}$ fields interpolated from previous computations.
- **Local stage:** Knowing a solution \mathbf{s} in A_d , local stage consists in finding a local solution $\hat{\mathbf{s}}$ in Γ using search direction $(\hat{\mathbf{s}} - \mathbf{s}) \in \mathbf{E}^+$.

$$\mathbf{E}^+(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) + (\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon}) = \mathbf{0} \quad (29)$$

230 The local problem can be solved at each time steps and at each Gauss point. Here, the search direction is chosen to have $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}$.

- **Global stage:** The linear stage consists in finding $\mathbf{s} \in A_d$ knowing $\hat{\mathbf{s}} \in \Gamma$ and using the search direction \mathbf{E}^- (30).

$$\mathbf{E}^-(\boldsymbol{\sigma} - \hat{\boldsymbol{\sigma}}) - (\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}) = \mathbf{0} \quad (30)$$

The corresponding sub-problem is linear but global in space and time. This problem is equivalent to the search of a correction $\delta\mathbf{s} \in A_{d,0}$ of the solution $\hat{\mathbf{s}} \in A_d$ calculated at the previous convergence iteration. Equation 30 can be written as:

$$\mathbf{E}^-((\tilde{\boldsymbol{\sigma}} + \delta\boldsymbol{\sigma}) - \hat{\boldsymbol{\sigma}}) - ((\tilde{\boldsymbol{\varepsilon}} + \delta\boldsymbol{\varepsilon}) - \hat{\boldsymbol{\varepsilon}}) = \mathbf{0} \quad (31)$$

The chosen error indicator is a stagnation criterion on consecutive solutions \mathbf{s} and $\hat{\mathbf{s}}$:

$$\nu_{i+1} = \frac{\|\mathbf{s}_{i+1} - \hat{\mathbf{s}}_{i+\frac{1}{2}}\|_{\mathbf{C}}}{\frac{1}{2}\|\mathbf{s}_{i+1} + \hat{\mathbf{s}}_{i+\frac{1}{2}}\|_{\mathbf{C}}}, \quad \|\mathbf{s}\|_{\mathbf{C}} = \int_{\Omega \times I} \frac{1}{2} (\boldsymbol{\varepsilon} : \mathbf{C} \boldsymbol{\varepsilon} + \boldsymbol{\sigma} : \mathbf{C}^{-1} \boldsymbol{\sigma}) \, d\Omega dt \quad (32)$$

where $\|\mathbf{s}\|_{\mathbf{C}}$ is the norm associated to the Hooke operator.

This method is particularly adapted to our process:

- An interrupted Newton-Raphson resolution gives the converged solution up to the time step i_t . In the case of the LATIN method, an approximation of the complete spatio-temporal solution is available. This property allows the solver to be considered as a goal-oriented one and will be used to generate multi-fidelity QoI data. 235
- It is possible to approximate the resolution at the global stage by searching fields correction $\delta\mathbf{s}$ like separated variable functions. The computational cost of the global stage is reduced and the set of fields \mathbf{s} can be written as a PGD reduced-order model. This aspect will be explained in section 2.4. 240
- The initialization of the Newton-Raphson method by an already computed spatio-temporal approximation is complex. By its inherent structure, the LATIN method naturally allows the initialization of the calculation by a complete pre-calculated spatio-temporal solution. This property allows a enhanced multiparametric strategy detailed in section 2.5. 245

2.4. On-the-fly model-order reduction with the PGD method

The Proper Generalized Decomposition is a model reduction method. Initiated under the name of radial approximation [45] as part of the LATIN framework for space-time decomposition, it is also used for rheology problems [46], for an approximate resolution of the Navier-Stokes equation [47] or in image correlation [48]. The method has been generalized with multiparametric use [49] — or even in large dimensions [50] — with possible use of hyper-reduction methods [51]. The method is also particularly suitable for application cases such as optimization [52], uncertainty quantification [53] in dynamic data-driven application systems (DDDAS) [54] or in haptics [55].

The objective is to compute the best correction of an approximation of the solution in a separated variable form. Modes database is no longer imposed as in Proper Orthogonal Decomposition and the solution of the problem is sought at the same time as the best basis for its representation. The approximation performed is enriched until a given error criterion is reached.

We assume that a PGD approximation of order $m-1$ has already been computed:

$$\underline{U}(t, M) \approx \underline{U}_0(t, M) + \underbrace{\sum_{i=1}^{m-1} \lambda_i(t) \cdot \underline{\Lambda}_i(M)}_{\underline{U}_c(t, M)} \quad (33)$$

with \underline{U}_0 the elastic solution such as $\underline{U}_0 \in \mathcal{U}$ and $\forall i, (\underline{\Lambda}_i, \lambda_i) \in \mathcal{U}^* \times \mathcal{I}$, with $\mathcal{I} = L^2(I, \mathbb{R})$ the space of the sumable square functions in I . In the PGD approach, neither functions $\underline{\Lambda}_k$ nor functions λ_k are given initially, and both families are computed on-the-fly. At the global linear stage of the LATIN method, the correction of field $\delta \underline{U}$ can be found with a separated-variable form:

$$\delta \underline{U} = \underline{\Lambda} \cdot \lambda \quad (34)$$

And can be obtained by minimizing the global search direction (20). It can be expressed as [18],

$$\text{Find } \delta \underline{U}(t, M) \in \mathcal{U}^* \text{ such that } \forall \underline{U}^* \in \mathcal{U}^*, a(\delta \underline{U}, \underline{U}^*) = \ell(\underline{U}^*) \quad (35)$$

where a is a continuous coercive bilinear operator according to $H_0^1(\Omega) \otimes \mathcal{I}$ and ℓ is a continuous linear operator according to $H_0^1(\Omega) \otimes \mathcal{I}$.

The new pair $(\underline{\Lambda}, \lambda) \in \mathcal{U}^* \times \mathcal{I}$ is defined as the one that verifies the following double Galerkin orthogonality criterion:

$$a(\tilde{u}_{m-1} + \underline{\Lambda} \lambda, \underline{\Lambda} \lambda^* + \underline{\Lambda}^* \lambda) = \ell(\underline{\Lambda} \lambda^* + \underline{\Lambda}^* \lambda), \quad \forall \lambda^* \in \mathcal{I}^*, \forall \underline{\Lambda}^* \in \mathcal{I} \quad (36)$$

We can thus define the two following applications:

- $S_m : \mathcal{I} \mapsto \mathcal{U}$ is the application that maps a time function λ into a space function $\underline{\Lambda} = S_m(\lambda)$ defined as:

$$a(\tilde{u}_{m-1} + \underline{\Lambda} \lambda, \underline{\Lambda}^* \lambda) = \ell(\underline{\Lambda}^* \lambda), \quad \forall \underline{\Lambda}^* \in \mathcal{U}^* \quad (37)$$

It is associated to a space problem.

- $T_m : \mathcal{U} \mapsto \mathcal{I}$ is the application that maps a space function $\underline{\Lambda}$ into a time function $\lambda = T_m(\underline{\Lambda})$ defined as:

$$a(\tilde{u}_{m-1} + \underline{\Lambda}\lambda, +\lambda^*\underline{\Lambda}) = \ell(\lambda^*\underline{\Lambda}), \quad \forall \lambda^* \in \mathcal{I} \quad (38)$$

It is associated to a time problem (scalar ODE).

A pair $(\underline{\Lambda}, \lambda)$ verifies (36) if and only if $\underline{\Lambda} = S_m(\lambda)$ and $\lambda = T_m(\underline{\Lambda})$, which is a nonlinear problem which could be solved with a fixed-point algorithm. The interested reader can refer to [56] for a review of the different algorithms to solve a linear problem with the PGD and [57] for the special case of the viscoplastic LATIN algorithm.

So with the LATIN-PGD framework, at each convergence step, a reduced-order approximation of the complete spatio-temporal solution is obtained and allow us to compute an approximation of the QoI firstly. By this aspect, a space and time basis can also be used to start future computations.

2.5. Fast computation of multiple solutions with LATIN multiparametric strategy and PGD interpolation

For surrogate modelling, the computation of a quantity of interest y_i is done sequentially on certain points \underline{x}_i in the design space \mathcal{D} . The computation of multiple solutions is fully compatible with a major feature of the LATIN framework: its multiparametric strategy [58] which allows to start a new calculation with fields created from previous computations. Its objective is to provide very quickly the solution of nonlinear evolution problems for several parameter values of the model and reduce the number of iterations to reach the required error estimator level.

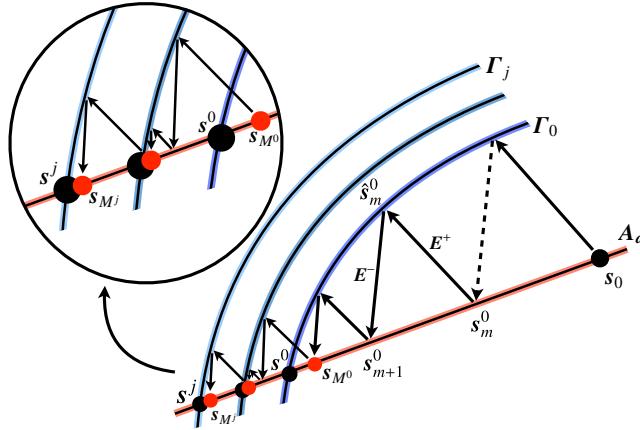


Figure 6: Schematic representation of the multiparametric strategy [59]. For computing solution parametrized by \underline{x}_j , the solution of a similar problem parametrized by \underline{x}_{j-1} can be used to reduce the iteration number.

We assume that for $i \in \llbracket 1, j-1 \rrbracket$, displacement solution \underline{U}_i linked with the problem parametrized by \underline{x}_i has already been computed. Each solution is expressed from its space and time basis generated at the same time as the solution.

For the calculation of the query parameter \underline{x}_j , an approximation of $\mathbf{s} \in A_d$ from previous computations is considered by interpolating the displacement field \underline{U}_j . Inverse distance weighting is considered, using euclidian distance in the design space \mathcal{D} :

$$\underline{U}_j(t, M) \approx \underline{U}_{j,0}(t, M) + \sum_{k=0}^{j-1} \omega_k \underline{U}_{k,c}(t, M) \quad (39)$$

290 with $\forall k \in \llbracket 1, j-1 \rrbracket$, $\omega_k = \phi(\|\underline{x}_j - \underline{x}_k\|_{\mathcal{D}})$, $\|\bullet\|_{\mathcal{D}}$ the euclidian norm of \mathcal{D} and $\phi: d \mapsto \frac{1}{d^p}$.

Furthermore, a global space basis $\mathbf{U} = [\underline{\Lambda}_1, \underline{\Lambda}_2, \dots, \underline{\Lambda}_{m_s}] \in \mathbb{R}^{d_s \times m_s}$ can be considered by orthonormalizing the concatenation of space bases previously obtained, such as $\forall k \in \llbracket 1, j-1 \rrbracket$:

$$\underline{U}_{k,c}(t, M) = \sum_{i=0}^{m_s} \lambda_i^{(k)}(t) \underline{\Lambda}_i(M) \quad (40)$$

So, by considering (39) and (40),

$$\underline{U}_j(t, M) \approx \underline{U}_{j,0}(t, M) + \sum_{i=0}^{m_s} \tilde{\lambda}_i(t) \underline{\Lambda}_i(M) \quad (41)$$

with $\forall i \in \llbracket 1, m_s \rrbracket$ $\tilde{\lambda}_i = \sum_{k=0}^{j-1} \omega_k \lambda_i^{(k)}$. Algorithm 2 summarizes the different steps of the multiparametric strategy implemented.

295 It is important to note that we assume here that the design variables \underline{x}_i modify only the manifold Γ . Thus the design variables are typically material coefficients or structural loads. In the developments proposed here, it is not possible to take into account variations in parameters that would modify in particular the kinematic admissibility conditions. However, manifold A_d can be modified for example when optimizing the geometry, and in this context, one
300 could then integrate the geometrical parameters in the PGD decomposition [60].

This strategy is tested on a turbine blade test case (see section 3.2). A 3D solver implementation was developed for the test using MATLAB. To reach fast and parallel computation, LATIN operators are organized as *nd*-arrays and a parallel implementation of Einstein summation exploits variable broadcasting to
305 deal with time- and spatial-dependent operators. This computational paradigm allows to be as fast as commercial softwares. 200 equidistant points of the design space $[\alpha, T] = [81^\circ, 100^\circ] \times [881^\circ\text{C}, 900^\circ\text{C}]$ are computed without the strategy, and with the strategy by an ordered path (the 200 values are sorted on the basis of a criterion of Euclidean distance in the design space, which leads to the
310 smallest possible evolutions of the parameters between 2 successive runs), and

Algorithm 2: Multiparametric strategy

Input: n points $\mu_1, \dots, \mu_n \in \mathcal{D}$

For each point μ_j

For each previous point μ_k

 Calculation of the contribution: $\omega_k = \phi(d(\mu_j, \mu_k))$

Field interpolation: $\underline{U}_j(t, M) \leftarrow \sum_{k=1}^{j-1} \omega_k \underline{U}_k(t, M)$

Generation of the common spatial base

 Concatenation of spatial modes:

$[\Xi, \Sigma, \mathbf{V}] \xleftarrow{\text{Thin SVD}} [\Xi_1, \dots, \Xi_{j-1}]$

Calculation of the associated elastic field: $\underline{U}_{j,0}(t, M)$

For each spatial mode $\Lambda_i^{(\Xi)}$

 Compute correction of time functions $\bar{\lambda}_i \leftarrow \left\langle \Lambda_i^{(\Xi)}, \underline{\tilde{U}}_j - \underline{U}_{j,0} \right\rangle_S$

LATIN Correction: $\underline{U}_j \in \mathbf{A}_{dj} \cap \Gamma_j$

Output: Mechanical solution for the n points

15 random paths. In the later case, the 200 sets of design variables are used in a random order, leading to potentially large changes in the design variables between 2 successive runs. Given the randomness, this random path is repeated 15 times and an average of the results over the 15 paths is calculated.

Points computed	2	5	10	20	50	100	200
Ordered path	4.25×	4.24×	4.46×	4.86×	4.95×	4.87×	5.05×
Random path	2.09×	2.13×	2.22×	2.56×	3.13×	3.51×	3.78×

Table 1: Speedup provided by the multiparametric strategy on the turbine blade test case

315 As a result of the previous study, it can be deduced that with the LATIN-PGD framework, multi-fidelity QoI values can be obtained easily and also that the multiparametric strategy allows good speedups. The objective is now to see if coupling multi-fidelity bayesian optimisation with the multiparametric strategy allows to simultaneously reduce the time computation and provide a
320 good estimation of the global optimum.

3. Global optimization by coupling multi-fidelity kriging and reduced-order models

3.1. Coupling algorithm

325 The different algorithmic blocks presented in the previous sections are assembled in algorithm 7. Some parameters influence the calculation time and the accuracy of the optimum obtained. Mainly two of them can be discussed: the number of low-fidelity observations n_{BF} before enrichment with high-fidelity data, and the quality of low-fidelity observations which will be driven by the

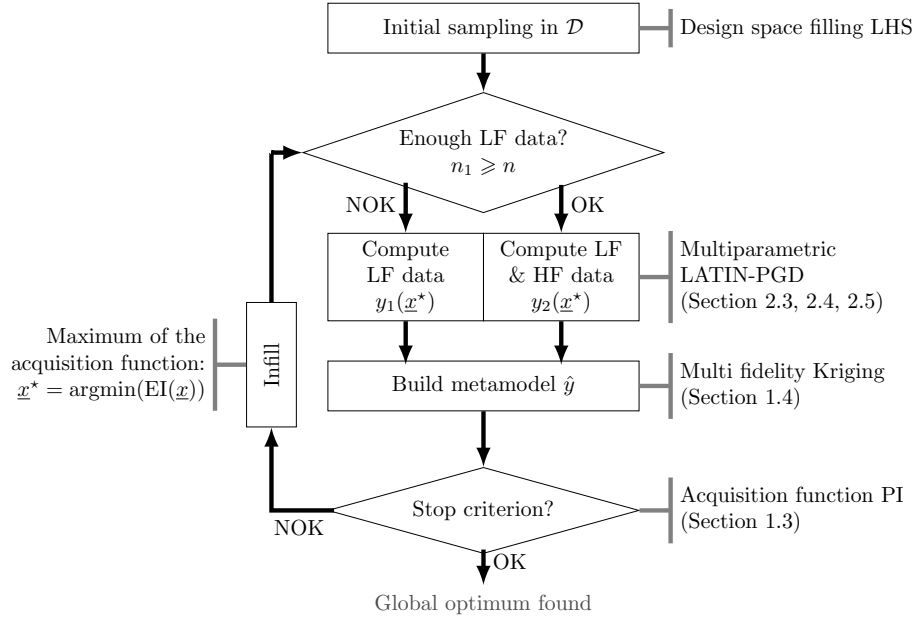


Figure 7: Enriched bayesian optimization

LATIN indicator ν_{BF} which will be the stopping criterion of the solver. To determine the pairs of parameters (n_{BF}, ν_{BF}) that allow to reduce calculation costs while reaching the optimum, an important testing campaign was carried out. Results are presented on two test cases on following sections. A compromise has been derived from this study between percentage of cases where the global optimum area was found and reduction of the calculation costs.

3.2. Presentation of the damper test-case

This first test case is freely inspired from a damper part of the Vulcain engine of the Ariane 5 launcher [18]. The typical dimensions in directions (x, y, z) are $45 \times 70 \times 50$ mm. The mechanical part is clamped on the upper, left and right sides (Blue parts in figure 8). The turbine nose is visible on the bottom left of figure 8. This one is loaded on the front and rear sides (green and red in the figure). The two loading pressures have the same intensity $P(t)$ described by figure 9 with $P_{max} = 80$ MPa, but their directions are different. On each side, the loading direction is driven by two angles (θ, ϕ) that vary in $[0^\circ, 90^\circ]$. The description of the angle-driven loading conditions is shown in figure 8. The material is a 316 Steel at 600°K . Material behavior is described by the Chaboche constitutive law (see Section 2.2) and Table 3 for the material coefficients.

The aim here is to obtain the worse loading case for the structure by dealing with the maximum of the Von Mises stress during one loading cycle. All the loading options describe a 4D design space $\mathcal{D} = [0^\circ, 90^\circ]^4$.

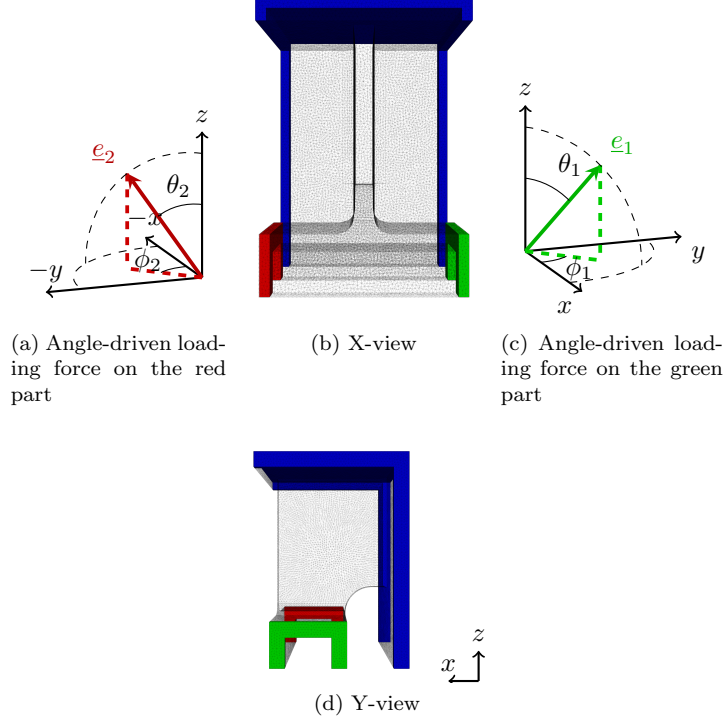


Figure 8: Damper test case

Objective function	$y(\underline{x}) = \max_{I \times \Omega} \sigma_{VonMises}$
Parameters	$\underline{x} = (\theta_1, \phi_1, \theta_2, \phi_2) \in \mathcal{D} = [0^\circ, 90^\circ]^4$
Space element type	Linear triangular (30,000 DOFs)
Loading cycles	One cycle (10s — 41 time steps)
Boundary conditions	Clamped on blue side faces $\underline{P}_1(t) = P(t) \underline{e}_1$ on green side $\underline{P}_2(t) = P(t) \underline{e}_2$ on red side

Table 2: Second test case characteristics

E	ν	N	k	σ_0	C	γ	R_∞	b
137.6 GPa	0.3	14	150 MPa.s ^{1/n}	20 MPa	37.2 GPa	300	80 MPa	10

Table 3: Elastic-viscoplastic constitutive coefficients for the second test case

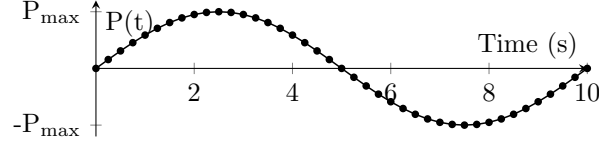


Figure 9: Pressure evolution $P(t)$

3.3. Strategy parameters estimation on the damper test-case

For each of the (n_1, ν_1) pairs tested, 100 optimizations are performed with 100 different random choice of initial points. Moreover, we consider a maximum of 100 high-fidelity points (which means that we are limited in terms of the maximum calculation time allowed for the study, which may lead to not reaching the success criterion because it would have required more high-fidelity points). As the objective function is to find the maximum of the Mises stress, a pre-study (very costly in computing time) has determined that beyond 113 MPa we can consider that the maximum is reached. Thus, we will consider as a success an optimization which leads to a Mises stress higher than 113 MPa, *i.e.*:

$$y_2(\underline{x}^*) > 113 \text{ MPa} \quad (42)$$

Table 4 shows the percentage of test-cases that converged to the global minimum area. Table 5 gives computation time to obtain the success criterion and Table 6 gives the impact of the random choice of initial points on the computation time. In the following tables, n/dim means n points per dimension of the design space. For example here, the design space is dimension 4, so $10/\text{dim}$ corresponds to $10 \times 4 = 40$ points.

$\nu_1 \backslash n_1$	10^{-3}	10^{-2}	10^{-1}
6/dim	56%	83%	75%
10/dim	58%	84%	74%
12/dim	63%	87%	74%
14/dim	59%	87%	77%

Table 4: Percentage of cases where the global minimum area was found

To reduce computation time, the best strategy is to consider firstly the computation of 4×6 low-fidelity points, obtained by considering $\nu_1 = 10^{-1}$ as the solver stopping criterion. This result is hardly surprising when looking at curve 10.

This figure is the empirical cumulative distribution function obtained by computing 20,000 QoI values from the 4D design space. The curve obtained for $\nu_1 = 10^{-1}$ allows to identify more easily the optimum area than information given by high-fidelity values ν_2 . With this choice, the gain brought by

$\nu_1 \backslash n_1$	10^{-3}	10^{-2}	10^{-1}
6/dim	27'40	9'48	3'14
10/dim	36'02	10'09	3'23
12/dim	42'38	10'27	3'26
14/dim	43'22	11'14	3'38

Table 5: Average calculation time — Only high-fidelity data: 1h01

$\nu_1 \backslash n_1$	10^{-3}	10^{-2}	10^{-1}
6/dim	47%	87%	14%
10/dim	72%	73%	28%
12/dim	106%	78%	30%
14/dim	107%	91%	42%

Table 6: Normalized variance of computation time — Only HF: 34%

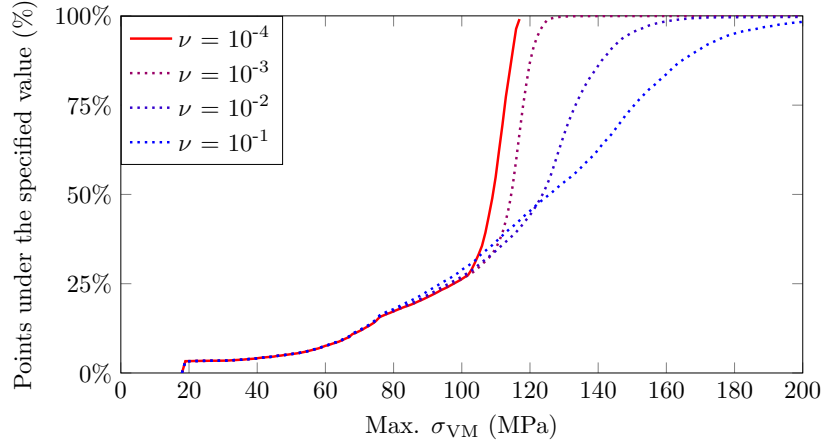


Figure 10: Empirical QoI cumulative distribution function

multi-fidelity is 19 times. The gain provided by the multiparametric method is estimated by testing multiple random paths of 30 points in the design space. The contribution of the multiparametric strategy allows fields to be calculated 2.5 times faster, which allows to estimate that the gain provided by coupling methods is around 47.5 times. Nevertheless, using $\nu_1 = 10^{-2}$ seems to be a good compromise between calculation time and success and this gain needs to be validated *a posteriori* by performing the calculation without the contribution of the multiparametric strategy, with only high-fidelity data. The next test-case overcomes this last issue.

3.4. Presentation of the turbine blade test-case

The second test-case is an aircraft turbine blade. The mesh comes from [61]. This blade is considered as clamped on its base, with blocked movement along the y-axis on some bottom sides (in purple in figure 11). This one has a centrifugal load and a uniform load on the upper surface (in red in Figure 11). The load direction is in the surface plane and is driven by angle $\alpha \in [0^\circ, 180^\circ]$ as seen in Figure 11c. The intensity of loads are shown in Figure 12. The material

of this test-case is an Inconel 601 at temperature $T \in [800^\circ\text{C}, 900^\circ\text{C}]$. Operating temperature will influence material parameters. Three sets of material parameters are given in table 7 and other values will be interpolated from them.

The aim here is to obtain $\max_{I \times \Omega} \sigma_{VonMises} = 180 \text{ MPa}$. To do that, a minimization is done on the 2D-design space with $(\alpha, T) \in [0^\circ, 180^\circ] \times [800^\circ\text{C}, 900^\circ\text{C}]$.

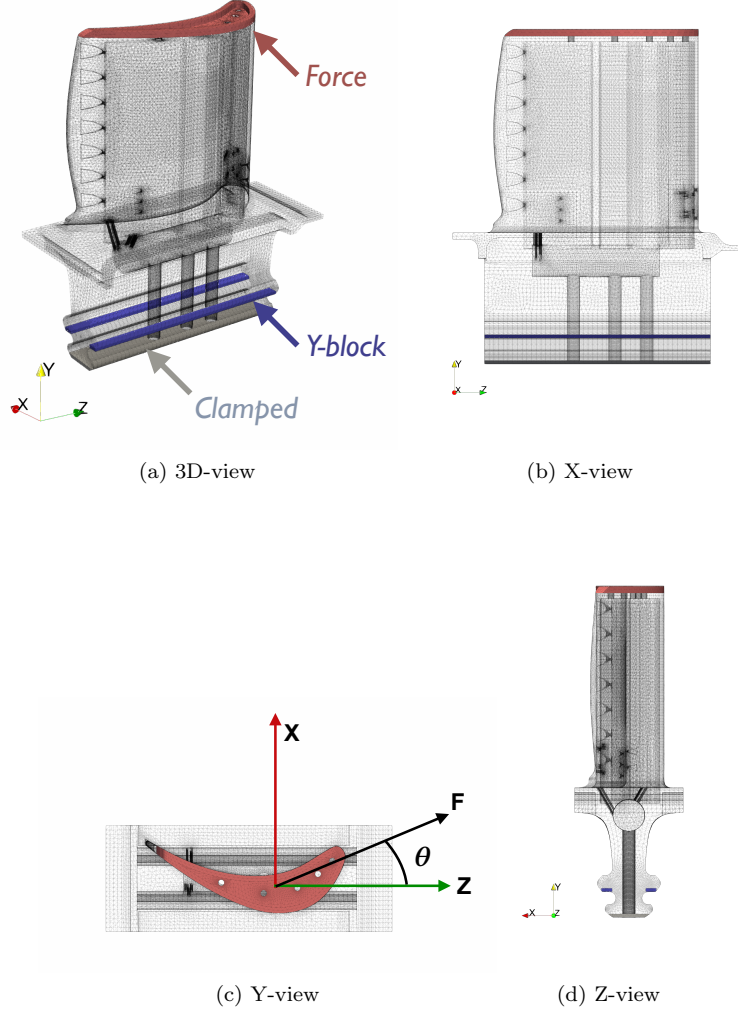


Figure 11: Different views of the structure

181 loading cases are tested with $T = 900^\circ\text{C}$ and $\alpha \in [0^\circ, 180^\circ]$ to quantify the influence of the error indicator explained in section 2.3 on the computation time, on the number of PGD modes generated, and the error on the quantity of interest. Results are presented in Figure 13 with an overkill solution. The

T	n	K	σ_0	C	γ	R_∞	b
800°C	14	630 MPa.s ^{1/n}	80 MPa	615 GPa	1.53.10 ⁶ s ⁻¹	80 MPa	300
850°C	11	560 MPa.s ^{1/n}	71 MPa	497 GPa	1.36.10 ⁶ s ⁻¹	70 MPa	250
900°C	9	490 MPa.s ^{1/n}	60 MPa	362 GPa	1.2.10 ⁶ s ⁻¹	60 MPa	200

Table 7: Elastic-viscoplastic constitutive coefficients for the first test-case (with $E = 210$ GPa and $\nu = 0.28$)

Objective function	$y(\underline{x}) = \max_{I \times \Omega} \sigma_{VonMises} - 180 $
Parameters	$\underline{x} = (\alpha, T) \in [0^\circ, 180^\circ] \times [850^\circ\text{C}, 950^\circ\text{C}]$
Space element type	Quadratic triangular (330,000DOFs)
Boundary conditions	Angle-driven load & Centrifugal load Clamping on bottom part Y-Block on bottom sides

Table 8: First test-case characteristics

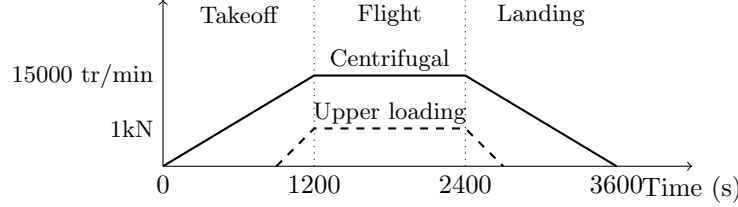


Figure 12: Loadings time evolution

overkill solution used here is a solution computed with an excessively small value of the LATIN indicator (equal to 10^{-8}), so that we can use this solution as a reference. These results allow us to consider LATIN Indicator 10^{-4} as the stopping criterion for converged field.

Figure 13a shows that even with low-fidelity data like $\nu_1 = 10^{-1}$, the QoI curves gives good trend and a good localization of the optimum area. A balance between QoI error and computation time can be found: if we consider low-fidelity data with $\nu_1 = 10^{-2}$, Figure 13b shows that the mean of the QoI error is around 8%, but computation time is 4.6 times faster than with data obtained with $\nu_2 = 10^{-4}$. So 4 observations can be obtained faster than obtaining a single high-fidelity observation, which allows to find faster the optimum zone. Nevertheless, high-fidelity information is still mandatory to verify the obtained optimum.

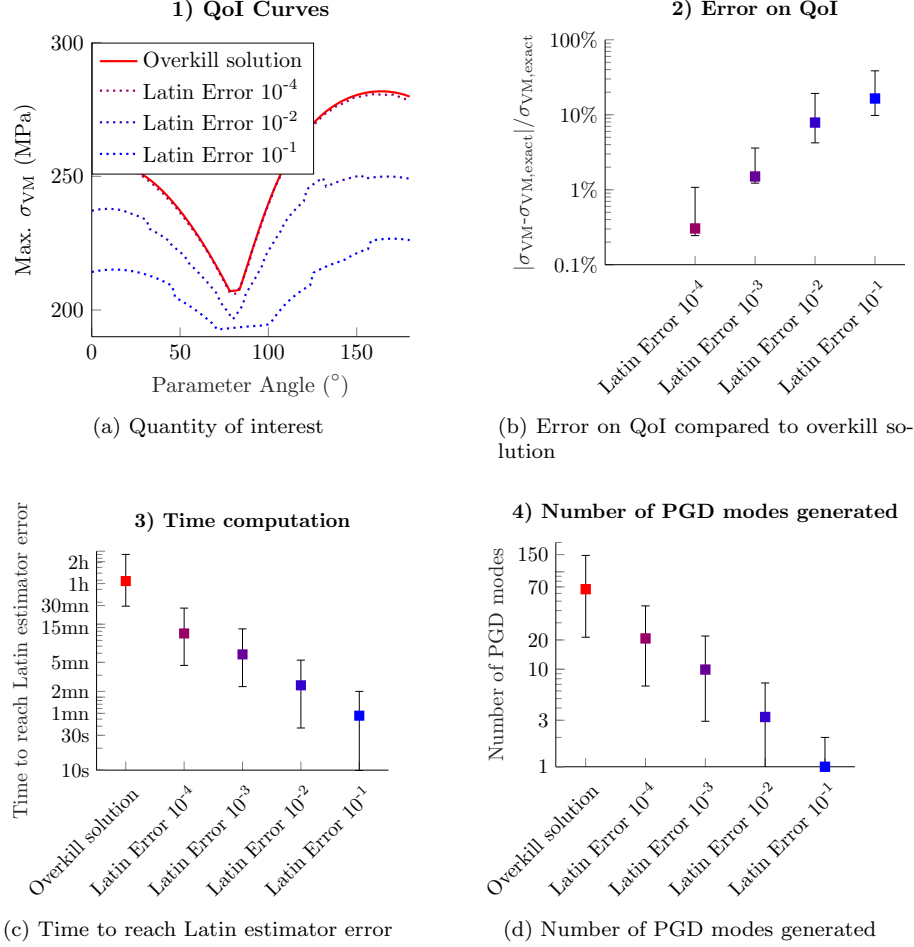


Figure 13: Influence of the value of the LATIN error indicator as stopping solver criterion

3.5. Results of the optimization test on the turbine blade test-case

To distinguish the gain provided by the multi-fidelity data fusion and the multiparametric strategy, three computation batches are performed. Each batch considers 20 optimisations made with different initial sampling to estimate the average time savings. The first batch ϕ_0 is performed with only exact data without the multiparametric strategy. The second one ϕ_1 is made with only exact data and the multiparametric strategy, and the last one ϕ_2 is obtained by coupling methods. Each optimisation is made by initially computing 10 points per dimension with LF stopping criterion $\nu_1 = 10^{-2}$.

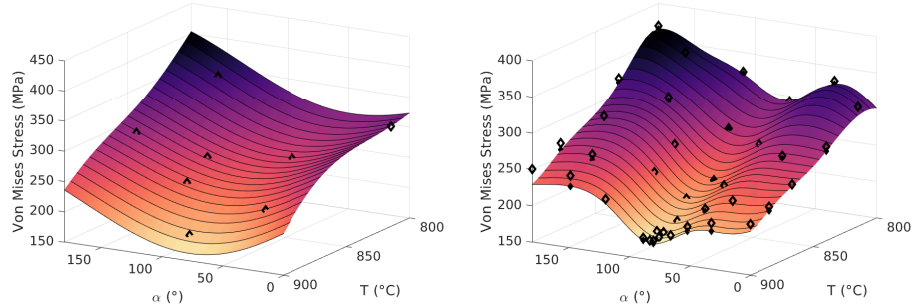
The success criterion corresponds to the case where the exact solution cal-

culated at least estimated is less than 180 MPa:

$$y_2(\underline{x}^*) < 180 \text{ MPa} \quad (43)$$

In this case, the success criterion is met each time. The ϕ_0 phase is performed in 6h05, the ϕ_1 phase in 2h51 and finally the ϕ_2 phase in 44 minutes. The multiparametric strategy provides a gain of $2.1\times$ and the multi-fidelity strategy
415 a gain of $3.9\times$ for a total gain of $8.4\times$. 17 high-fidelity points were requested on average by the algorithm to pass the objective (In the worst case, 24 high-fidelity points were requested).

Figure 14a shows one of the 20 initial samplings used to start the bayesian optimisation. Every low-fidelity observations are interpolated since we don't
420 have any information about QoI error on these points. Figure 14b shows the last metamodel generated by the bayesian optimisation. Low-fidelity observations were corrected by high-fidelity information and the correction metamodel. The low-fidelity corrected point in $(180^\circ, 900^\circ\text{C})$ is not interpolant as we consider the lack of information on the correction metamodel and so its variance. High-
425 fidelity points are filled on the figure. On the same position in the design space, low-fidelity points are also visible to understand the gap between low-fidelity and high-fidelity data.



(a) The initial metamodel (with only low-fidelity data) (b) The final metamodel (High-fidelity data are filled diamonds)

Figure 14: Metamodels generated during the optimization process

This figure can be compared to Figure 15a which is the QoI reference function to give Figure 15b. The QoI error between the final metamodel generated by the
430 optimization and the reference shows that the QoI error is almost zero around the optimum, but quite important far from the optimum zone.

Conclusion

The aim of this work is to accelerate the global optimization of mechanical structures to allow the use of more complex and accurate models within the

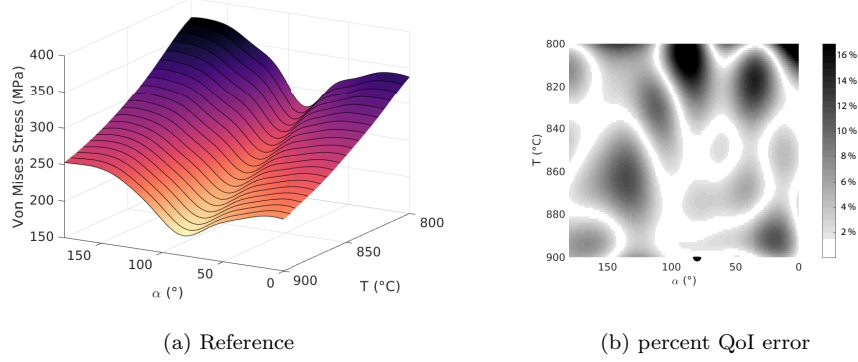


Figure 15: Comparison between the final metamodel generated by the optimization and the reference

design offices. To achieve this goal, the bayesian optimization process has been improved to the use of multi-fidelity data by considering an adapted acquisition function. The surrogate model generation step also uses an improved version of the multi-fidelity kriging method Evofusion, which adds an estimated variance on low-fidelity data.

The multi-fidelity data are not obtained by calling a given quality model, but by using a solver allowing the computation of mechanical fields with an adapted level of fidelity. The proposed solver is based on the LATIN-PGD framework, which has the triple advantage of a fast computation of the spatio-temporal fields in a separated-variable form, giving an approximation of the complete solution and the amount of interest at each iteration, and allowing to start computation from an approximation interpolated from previous computations with the multiparametric strategy.

The two families of methods presented will thus work together to bring significant time savings on the global optimization of a quantity of interest. The generalization of the enrichment strategy requires the determination of the number of calculated low-fidelity points, as well as of the solver stopping criterion in this low-fidelity case. A parametric study campaign has been performed to estimate these two parameters. 10 low-fidelity points per dimension with a LATIN stopping criterion $\nu_1 = 10^{-2}$ will be firstly considered before adding high-fidelity points. With these parameters, the optimization of the airplane blade test-case was obtained in 44 minutes instead of 6 hours, which corresponds to a gain of $8.4\times$.

In the short term, the strategy is expected to be applied for geometric optimization problems. The main difficulty concerns the multiparametric strategy and the interpolation of fields for different geometries. This difficulty can be overcome by morphing around a single mesh [62], by using XFEM methods [63], or by isogeometric parameterization of the structure [64]. In this last case, tools already developed can be used using an IGA-FEM coupling method [65].

Thereupon, the strategy can also be improved by a better choice of the
 465 acquisition function. A comparison of the different acquisition functions should
 be made in the same way as [36] in the case of mechanical problems. Several
 enrichment criteria, including the Approximate Knowledge Gradient (AKG)
 criterion and the Augmented Expected Improvement (AEI) criterion, were not
 compared to the others. However, it appears that the AKG function can help
 470 to obtain the optimum faster.

4. Bibliography

- [1] P. Duysinx, M. P. Bendsøe, Topology Optimization of Continuum Structures with Local Stress Constraints, *International Journal for Numerical Methods in Engineering* 43 (1998) 1453–1478.
- 475 [2] G. Allaire, Shape Optimization by the Homogenization Method, Applied Mathematical Sciences, Springer-Verlag, New York, 2002.
- [3] S. Choi, J. J. Alonso, S. Kim, I. M. Kroo, M. Wintzer, Two-Level Multi-Fidelity Design Optimization Studies for Supersonic Jets, in: 43rd AIAA Aerospace Sciences Meeting and Exhibit, AIAA Paper 2005 - 0531, Reno, Nevada, 2005, p. 2.
 480
- [4] J. Mockus, V. Tiesis, A. Zilinskas, The application of Bayesian methods for seeking the extremum, *Towards global optimization* 2 (117-129) (1978) 2.
- [5] D. R. Jones, M. Schonlau, W. J. Welch, Efficient global optimization of expensive black-box functions, *Journal of Global optimization* 13 (4) (1998) 455–492.
 485
- [6] D. G. Krige, A statistical approach to some mine valuation and allied problems on the Witwatersrand: By DG Krige (1951).
- [7] G. Matheron, Principles of geostatistics, *Economic geology* 58 (8) (1963) 1246–1266.
 490
- [8] C. E. Rasmussen, C. K. Williams, Gaussian processes in machine learning, *Lecture notes in computer science* 3176 (2004) 63–71.
- [9] F. Hutter, H. H. Hoos, K. Leyton-Brown, Sequential Model-Based Optimization for General Algorithm Configuration, in: C. A. C. Coello (Ed.), *Learning and Intelligent Optimization, Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2011, pp. 507–523.
 495
- [10] K. Kandasamy, G. Dasarathy, J. Oliva, J. Schneider, B. Póczos, Multi-fidelity gaussian process bandit optimisation, *Journal of Artificial Intelligence Research* 66 (2019) 151–196. doi:10.1613/jair.1.11288.
 500 URL <http://dx.doi.org/10.1613/jair.1.11288>

- [11] A. I. Forrester, N. W. Bressloff, A. J. Keane, Optimization using surrogate models and partially converged computational fluid dynamics simulations, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 462 (2011) (2006) 2177–2204. doi:10.1098/rspa.2006.1679.
- 505 [12] Z. Han, R. Zimmerman, S. Görtz, Alternative Cokriging Method for Variable-Fidelity Surrogate Modeling, *AIAA Journal* 50 (5) (2012) 1205–1210. doi:10.2514/1.J051243.
- [13] R. Zimmermann, Z.-H. Han, Simplified Cross-Correlation Estimation for Multi-Fidelity Surrogate Cokriging Models, *Advances and Applications in Mathematical Sciences* 7 (2) (2010) 181–201.
- 510 [14] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2017) 301–320 doi:10.1111/j.1467-9868.2005.00503.x.
- [15] H. K. Yong, L. Wang, D. J. J. Toal, A. J. Keane, F. Stanley, Multi-fidelity Kriging-assisted structural optimization of whole engine models employing medial meshes, *Structural and Multidisciplinary Optimization* (2019). doi:10.1007/s00158-019-02242-6.
- 515 [16] T. Benamara, P. Breitskopf, I. Lepot, C. Sainvitu, P. Villon, Multi-fidelity POD surrogate-assisted optimization: Concept and aero-design study, *Structural and Multidisciplinary Optimization* 56 (6) (2017) 1387–1412. doi:10.1007/s00158-017-1730-4.
- 520 [17] F. Chinesta, R. Keunings, A. Leygue, *The Proper Generalized Decomposition for Advanced Numerical Simulations*, SpringerBriefs in Applied Sciences and Technology, Springer International Publishing, Cham, 2014.
- [18] D. Néron, P.-A. Boucard, N. Relun, Time-space PGD for the rapid solution of 3D nonlinear parametrized problems in the many-query context, *International Journal for Numerical Methods in Engineering* 103 (4) (2015) 275–292. doi:10.1002/nme.4893.
- 525 [19] S. Nachar, P.-A. Boucard, D. Néron, F. Bordeu, Coupling multi-fidelity kriging & model-order reduction for the construction of virtual charts, *Computational Mechanics* (2019).
- 530 [20] P. Ladevèze, *Nonlinear Computational Structural Mechanics: New Approaches and Non-Incremental Methods of Calculation*, Mechanical Engineering Series, Springer Verlag, 1999.
- [21] P.-A. Boucard, P. Ladevèze, A multiple solution method for non-linear structural mechanics, *Mechanical Engineering* 50 (5) (1999) 317–328.
- 535 [22] C. Heyberger, P.-A. Boucard, D. Néron, Multiparametric analysis within the proper generalized decomposition framework, *Computational Mechanics* 49 (3) (2012) 277–289. doi:10.1007/s00466-011-0646-x.

- 540 [23] J. Nocedal, S. J. Wright, Numerical Optimization, 2nd Edition, Springer, 2006.
- [24] J. Močkus, On Bayesian Methods for Seeking the Extremum, in: G. I. Marchuk (Ed.), Optimization Techniques IFIP Technical Conference: Novosibirsk, July 1–7, 1974, Lecture Notes in Computer Science, Springer
545 Berlin Heidelberg, Berlin, Heidelberg, 1975, pp. 400–404. doi:10.1007/978-3-662-38527-2.
- [25] D. R. Jones, A Taxonomy of Global Optimization Methods Based on Response Surfaces, Journal of Global Optimization 21 (4) (2001) 345–383. doi:10.1023/A:1012771025575.
- 550 [26] M. D. McKay, R. J. Beckman, W. J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 42 (1) (2000) 55–61.
- [27] C. K. I. Williams, Prediction with Gaussian processes: From linear regression to linear prediction and beyond, 1998. doi:10.1007/978-94-011-5014-9_23.
555
- [28] M. De Lozzo, Substitution de modèle et approche multifidélité en expérimentation numérique, Journal de la Société Française de Statistique 156 (4) (2015) 21–55.
- [29] Z.-H. Han, R. Zimmermann, S. Görtz, A New Cokriging Method for Variable-Fidelity Surrogate Modeling of Aerodynamic Data, in: 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, AIAA 2010-1225, Orlando, FL, 2010.
560
- [30] M. C. Kennedy, A. O’Hagan, Predicting the output from a complex computer code when fast approximations are available, Biometrika 87 (1) (2000) 1–13. doi:10.1093/biomet/87.1.1.
565
- [31] L. Le Gratiet, Bayesian Analysis of Hierarchical Multifidelity Codes, SIAM/ASA Journal on Uncertainty Quantification 1 (1) (2013) 244–269. doi:10.1137/120884122.
- [32] Z.-H. Han, S. Görtz, Hierarchical kriging model for variable-fidelity surrogate modeling, AIAA journal 50 (9) (2012) 1885–1896.
570
- [33] N. Courrier, P.-A. Boucard, B. Soulier, Variable-fidelity modeling of structural analysis of assemblies, Journal of Global Optimization 64 (3) (2016) 577–613. doi:10.1007/s10898-015-0345-9.
- [34] A. I. Forrester, A. Sóbester, A. J. Keane, Multi-fidelity optimization via surrogate modelling, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 463 (2088) (2007) 3251–3269. doi:10.1098/rspa.2007.1900.
575

- [35] C. Durantin, J. Rouxel, J.-A. Désidéri, A. Glière, Multifidelity surrogate modeling based on radial basis functions, *Structural and Multidisciplinary Optimization* 56 (5) (2017) 1061–1075. doi:10.1007/s00158-017-1703-7.
- [36] V. Picheny, T. Wagner, D. Ginsbourger, A benchmark of kriging-based infill criteria for noisy optimization, *Structural and Multidisciplinary Optimization* 48 (3) (2013) 607–626. doi:10.1007/s00158-013-0919-4.
- [37] J. Lemaitre, J.-L. Chaboche, *Mechanics of Solid Materials*, Cambridge university press, 1994.
- [38] N. Relun, D. Néron, P. A. Boucard, A model reduction technique based on the PGD for elastic-viscoplastic computational analysis, *Computational Mechanics* 51 (1) (2013) 83–92. doi:10.1007/s00466-012-0706-x.
- [39] P. Ladevèze, A. Nouy, O. Loiseau, A multiscale computational approach for contact problems, *Computer Methods in Applied Mechanics and Engineering* 191 (43) (2002) 4869–4891. doi:10.1016/S0045-7825(02)00406-1.
- [40] L. Champaney, J. Y. Cognard, D. Dureisseix, P. Ladevèze, Large scale applications on parallel computers of a mixed domain decomposition method, *Computational Mechanics* 19 (4) (1997) 253–263. doi:10.1007/s004660050174.
- [41] D. Néron, D. Dureisseix, A computational strategy for poroelastic problems with a time interface between coupled physics, *International Journal for Numerical Methods in Engineering* 73 (6) (2008) 783–804.
- [42] O. Allix, P. Ladevèze, Interlaminar interface modelling for the prediction of delamination, *Composite Structures* 22 (4) (1992) 235–242. doi:10.1016/0263-8223(92)90060-P.
- [43] H. Lemoussu, P. A. Boucard, P. Ladevèze, A 3D shock computational strategy for real assembly and shock attenuator, *Advances in Engineering Software* 33 (7) (2002) 517–526. doi:10.1016/S0965-9978(02)00074-1.
- [44] L. Champaney, Une nouvelle approche modulaire pour l’analyse d’assemblages de structures tridimensionnelles, Ph.D. thesis, ENS-Cachan, Cachan, France (1996).
- [45] P. Ladevèze, Sur une famille d’algorithmes en mécanique des structures, *Comptes-rendus des séances de l’Académie des sciences. Série 2, Mécanique-physique, chimie, sciences de l’univers, sciences de la terre* 300 (2) (1985) 41–44.
- [46] F. Chinesta, P. Ladeveze, E. Cueto, A Short Review on Model Order Reduction Based on Proper Generalized Decomposition, *Archives of Computational Methods in Engineering* 18 (4) (2011) 395. doi:10.1007/s11831-011-9064-7.

- [47] A. Dumon, C. Allery, A. Ammar, Proper general decomposition (PGD) for the resolution of Navier–Stokes equations, *Journal of Computational Physics* 230 (4) (2011) 1387–1407. doi:10.1016/j.jcp.2010.11.010.
- [48] J.-C. Passieux, J.-N. Périé, High resolution digital image correlation using proper generalized decomposition: PGD-DIC, *International Journal for Numerical Methods in Engineering* 92 (6) (2012) 531–550. doi:10.1002/nme.4349.
- [49] C. Heyberger, P. A. Boucard, D. Néron, A rational strategy for the resolution of parametrized problems in the PGD framework, *Computer Methods in Applied Mechanics and Engineering* 259 (2013) 40–49. doi:10.1016/j.cma.2013.03.002.
- [50] C. Paillet, D. Néron, P. Ladevèze, A door to model reduction in high-dimensional parameter space, *Comptes Rendus Mécanique* 346 (7) (2018) 524–531. doi:10.1016/j.crme.2018.04.009.
- [51] M. Capaldo, P. A. Guidault, D. Néron, P. Ladevèze, The Reference Point Method, a “hyperreduction” technique: Application to PGD-based nonlinear model reduction, *Computer Methods in Applied Mechanics and Engineering* 322 (2017) 483–514. doi:10.1016/j.cma.2017.04.033.
- [52] F. Schmidt, N. Pirc, M. Mongeau, F. Chinesta, Efficient mold cooling optimization by using model reduction, *International Journal of Material Forming* 4 (1) (2011) 73–82. doi:10.1007/s12289-010-0988-5.
- [53] A. Nouy, A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations, *Computer Methods in Applied Mechanics and Engineering* 196 (45) (2007) 4521–4537. doi:10.1016/j.cma.2007.05.016.
- [54] C. Ghnatios, F. Masson, A. Huerta, A. Leygue, E. Cueto, F. Chinesta, Proper Generalized Decomposition based dynamic data-driven control of thermal processes, *Computer Methods in Applied Mechanics and Engineering* 213–216 (2012) 29–41. doi:10.1016/j.cma.2011.11.018.
- [55] F. Bordeu, F. Chinesta, A. Leygue, E. Cueto, S. Niroomandi, Réduction de modèle par PGD appliqué la simulation en temps réel de solide déformables, in: 10e Colloque National En Calcul Des Structures, Giens, France, 2011.
- [56] A. Nouy, A priori model reduction through Proper Generalized Decomposition for solving time-dependent partial differential equations, *Computer Methods in Applied Mechanics and Engineering* 199 (23–24) (2010) 1603–1626. doi:10.1016/j.cma.2010.01.009.
- [57] M. Bhattacharyya, A. Fau, U. Nackenhorst, D. Néron, P. Ladevèze, A multi-temporal scale model reduction approach for the computation of fatigue damage, *Computer Methods in Applied Mechanics and Engineering* 340 (2018) 630–656. doi:10.1016/j.cma.2018.06.004.

- [58] P.-A. Boucard, L. Champaney, Approche multirésolution pour l'étude paramétrique d'assemblages par contact et frottement., *Revue Européenne des Eléments Finis* 13 (5/7) (2004) 437–448.
- 660 [59] M. Vitse, D. Néron, P.-A. Boucard, Virtual charts of solutions for parametrized nonlinear equations, *Computational Mechanics* 54 (6) (2014) 1529–1539. doi:10.1007/s00466-014-1073-6.
- [60] A. Courard, D. Néron, P. Ladevèze, L. Ballere, Integration of pgd-virtual charts into an engineering design process, *Computational Mechanics* 57 (4) (2016) 637–651. doi:10.1007/s00466-015-1246-y.
- 665 [61] M. Blanchard, Méthode global/local non-intrusive pour les simulations cycliques non-linéaires, Ph.D. thesis, Université Paris-Saclay (Jan. 2018).
- [62] A. Courard, D. Néron, P. Ladevèze, L. Ballere, Integration of PGD-virtual charts into an engineering design process, *Computational Mechanics* 57 (4) (2016) 637–651. doi:10.1007/s00466-015-1246-y.
- 670 [63] L. Noël, L. V. Miegroet, P. Duysinx, Analytical sensitivity analysis using the extended finite element method in shape optimization of bimaterial structures, *International Journal for Numerical Methods in Engineering* 107 (8) (2016) 669–695. doi:10.1002/nme.5181.
- 675 [64] T. Hughes, J. Cottrell, Y. Bazilevs., Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement., *Computer Methods in Applied Mechanics and Engineering* 194 (39-41) (2005) 4135–4195.
- [65] M. Tirvaudey, R. Bouclier, J.-C. Passieux, L. Chamoin, Non-invasive implementation of nonlinear Isogeometric Analysis in an industrial FE software, *Engineering Computations* In press (2019).