10th World Congress on Structural and Multidisciplinary Optimization
May 19 -24, 2013, Orlando, Florida, USA

The key points approach: combining response surface methodology and reduced order modeling to achieve drastic reduction in surrogates construction cost

Christian Gogu[1], Jean-Charles Passieux[1]

[1] Université de Toulouse ; UPS, INSA, Mines Albi, ISAE ; ICA (Institut Clément Ader), Toulouse, France,
christian.gogu@gmail.com

1. Abstract
An efficient, thus popular approach for approximating the response of a computationally expensive model is through response surface methodology. Nevertheless, decreasing the surrogate construction cost and improving the handling of high dimensional output remain active topics of research. This paper describes a new approach for addressing both these challenges simultaneously for those types of codes involving the resolution of large systems of equations, such as typically finite element codes. The proposed method is based on the interaction between response surface methodology and reduced order modeling by projection, also known as reduced basis modeling. The novel idea is to carry out the full resolution of the system of equations only at a small, appropriately chosen, number of points. At all the remaining points, only an inexpensive reduced basis solution is calculated. A specific error criterion allows to control the quality of the approximation by determining whether additional full system resolutions are necessary or whether the reduced basis solutions are sufficient. An application example is provided involving the construction of a surrogate for the temperature field in a rocket engine combustion chamber wall. The proposed sequential surrogate construction strategy allowed a reduction by about an order of magnitude in the total system resolution time compared to the traditional response surface construction.

2. Keywords: response surface methodology, surrogate modeling, reduced basis modeling, proper orthogonal decomposition (POD), key points

3. Introduction

Computational codes allowed over the years to model increasingly complex phenomena. However, they often involve significant computational cost, which hinders their use in some applications requiring frequent calls to the simulation (e.g. optimization, statistical sampling). One way of reducing the computational cost is by using response surface methodology, also known as surrogate modeling, which aims at constructing an approximation of the simulation response based on a limited number of runs of the expensive simulation [1]-[6]. Multiple surrogate types can be used for fitting the samples, such as polynomial response surface approximations [7], kriging [8]-[10], neural networks [11]-[13] or support vector machines [14]-[16].
Some frequently encountered attributes of today's numerical simulations that render response surface construction more difficult are their high computational cost, the presence of a large number of input variables and the fact that the output of interest may not be a scalar but a high dimensional vector (e.g. the entire displacement field on a structure). The large number of variables is problematic due to the curse of dimensionality, in that the number of simulations required to construct the response surface grows exponentially. This problem is exacerbated when the computational cost of each simulation is high. The dimensionality of the output is problematic because it renders surrogate modeling of the full output difficult in the absence of additional assumptions.
These aspects thus pose the following challenges in terms of response surface construction [17]:

i. how to construct the surrogate model as efficiently as possible (i.e. with the least computational expenses)
ii. how to construct a surrogate model when the output quantity is not a scalar but a high-dimensional vector

(e.g. a pressure field map on an aircraft wing)

To address the first item multiple approaches have been proposed that are based on reducing the number of variables in the input space, which has the effect of decreasing the number of simulations required for the response surface construction. Among such approaches we could mention one-at-a-time (OAT) variable screening [18], global sensitivity analysis [19] or non-dimensional variable grouping [20]. The main purpose of these approaches is to remove variables that have negligible impact and regroup as efficiently as possible those that have. For additional details on these and other dimensionality reduction approaches for the input variables we refer the reader to the review in [17].

The second item (ii.), relative to the dimensionality of the output, can also be addressed by dimensionality reduction approaches, of course in the output space this time. For vector-based response, surrogate modeling techniques that take into account correlation between components are available [21] and can in some cases be more accurate than constructing response surfaces for each component independently. For a relatively small dimension of the output vector, methods such as co-kriging [22] or vector splines [23] are available. However, these are not practical for approximating the high dimensional pressure field around the wing of an aircraft or approximating heterogeneous displacements fields on a complex specimen since these fields are usually described by a vector with thousands to hundreds of thousand components. Fitting a surrogate for each component is even more time and resource intensive and might not take advantage of the correlation between neighboring points.
Reduced order modeling approaches by projection of the response on a reduced basis have proved to be efficient methods for achieving drastic dimensionality reduction in the output space. For example principal components analysis (PCA), also known as proper orthogonal decomposition (POD) or Karhunen-Loeve expansion, allows to determine a reduced-dimensional basis of the output space given a set of output simulation samples. Any output can then be projected on this basis and expressed by its basis coefficients. The challenge of obtaining a surrogate of high dimensional vector-type output quantities can then be solved by constructing response surfaces for the basis coefficients in terms of the design variables of interest. Such an approach has been successfully applied to the multidisciplinary design optimization of aircraft wings [24]-[26], to reliability based design optimization of automotive vehicles [27], to random fields uncertainty representation and propagation [28],[29].
The aim of this article is to present a new methodology that can address both points (i. and ii.) more efficiently than existing approaches for problems involving partial differential equations (typically problems solved by the finite elements method). Our method is based on the coupling of the reduced basis modeling approach with the construction phase of a response surface in order to achieve more efficient surrogate construction of complex multidimensional output. To do so, the idea is to (1) solve the full problem, but on a small number of points of the design of experiments, (2) to use these solutions to construct a reduced basis, and (3) to use this reduced basis to build approximate solutions for the other points of the design of experiments, while controlling the quality of the approximations.
The article is organized as follows. We provide in section 4 an overview of reduced order modeling by projection, i.e. reduced basis modeling, and of its coupling with surrogate modeling as proposed by our method. In section 5 we give an application example of the proposed methodology to a thermal design problem. Finally, we provide concluding remarks in section 6.

4. The key points response surface approach

4.1 Reduced basis modeling

Many numerical simulations in the engineering domain involve solving a partial differential equations problem. After space (and time) discretization, the problem often involves a (set of) large linear system(s) of equations.

$$K(\boldsymbol{u};\boldsymbol{\mu}) = \boldsymbol{F} \qquad (1)$$

with $\boldsymbol{u} \in \mathbb{R}^n$ the unknown state variables and $\boldsymbol{\mu} \in \mathbb{R}^p$ a set of p parameters of interest (material parameters, time...) so that $K : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$, n being the number of state variables. Let us assume that K is such that given any value of the set of parameters $\boldsymbol{\mu}$ a unique solution $\boldsymbol{u}=\boldsymbol{u}(\boldsymbol{\mu})$ exists.
Model order reduction is a family of approaches that aims at significantly decreasing the computational burden associated with the inversion of system (1). A particular class of model reduction techniques, denoted as reduced basis approaches (or reduced order modeling by projection), aims at reducing the number of state variables of the model by projection on a certain basis. Accordingly, an approximation of the solution is sought in a subspace $\mathcal{V}$ of dimension m (with usually $m<<n$), while enforcing the residual to be orthogonal to the same sub-space $\mathcal{V}$. Typically, $\mathcal{V}$ is defined by a so called *reduced-basis* $\boldsymbol{\Phi}=\{\boldsymbol{\Phi}_1,\ldots,\boldsymbol{\Phi}_m\}$.
The initial problem of Eq. 1 is rewritten, as defined by Galerkin conditions, projected onto the reduced basis:

$$\boldsymbol{\Phi}^T K(\Phi\boldsymbol{\alpha},\boldsymbol{\mu}) = \boldsymbol{\Phi}^T \boldsymbol{F} \qquad (2)$$

where $\boldsymbol{\alpha}$ are the reduced state variables, that is the coefficients of vector $\boldsymbol{u}$ expressed in the reduced basis $\boldsymbol{\Phi}$.
If K is linear with respect to its first variable $\boldsymbol{u}$, the problem of Eq. 1 can be written as:

$$K(\boldsymbol{\mu})\boldsymbol{u} = \boldsymbol{F} \qquad (3)$$

For example in structural mechanics $K(\boldsymbol{\mu})$ is the stiffness matrix, which usually depends on some parameters of interest $\boldsymbol{\mu}$ (e.g. material properties), $\boldsymbol{u}$ is the displacement vector and $\boldsymbol{F}$ the vector of the forces.
Similarly then, the projected problem of Eq. 2 can be written as:

$$\Phi^T K(\boldsymbol{\mu})\Phi\boldsymbol{\alpha} = \Phi^T \boldsymbol{F} \qquad (4)$$

At this point it is important to realize that Eq. 4 is equivalent to a reduced order model of the initial problem of Eq. 3. Indeed, solving the problem of Eq. 4 typically involves the inversion of a large system of equations of size *n*, the size of the stiffness matrix $K(\boldsymbol{\mu})$, which for large scale problems can easily reach hundreds of thousands. On the other hand solving the reduced order model of Eq. 4 involves the inversion of a much smaller system of equations of size *m*, the size of the projected stiffness matrix $\Phi^T K(\boldsymbol{\mu})\Phi$, which is equal to the dimensionality of the reduced basis *m* (typically *m*<<*n*, since *m* does usually not exceed a few dozen). Solving this reduced order model leads directly to $\boldsymbol{\alpha}$, the coefficients of the solution in the reduced basis.
The problem projected onto the reduced basis thus yields an approximate solution whose accuracy can be quantified by measuring the following residual:

$$e_{rb}^2 = \frac{\left\| K(\boldsymbol{\mu})\Phi\boldsymbol{\alpha} - \boldsymbol{F} \right\|^2}{\left\| \boldsymbol{F} \right\|^2} \qquad (5)$$

Up to now, the subspace V on which the problem is projected, or more precisely one of its basis Φ, was not specified and many different choices are possible for this projection.
The proper orthogonal decomposition (POD) can be a way to build a relevant reduced basis in the context of reduced order modeling by projection [30]-[32]. Indeed, POD [33] (also known as Karhunen Loeve decomposition [34],[35] or principal component analysis [36]) is an approach which consists in constructing a reduced basis from a set of solution, called snapshots. Mathematically, the extraction of the reduced basis from the snapshots is done by singular value decomposition. Generally, the snapshots are the results of full simulations on a set of points.
Note that classical POD-type approaches do not generally go all the way to solving or even formulating the reduced order model of Eq. 4. Indeed POD is often used only as a dimensionality reduction approach. Solutions $\boldsymbol{u}$ that were already calculated by solving the full size problem (Eq. 3) are projected on the POD-reduced basis and expressed in terms of their basis coefficients $\boldsymbol{\alpha}$ thus allowing to express initial solutions $\boldsymbol{u}$ with a drastically reduced dimensionality. Note however that these basis coefficients $\boldsymbol{\alpha}$ can also be obtained by solving the reduced order model of Eq. 4, that is by solving the initial problem projected on the reduced basis. This option is not typically used in POD because solving the reduced order model for solutions that are already in the reduced basis (thus for which the full solution is already available) has no interest. However the reduced order model has an interest for obtaining the basis coefficients $\boldsymbol{\alpha}$ at a new point which is not part of the reduced basis. In this case, the reduced order solution is only an approximation and the approximation error can be assessed by the metric provided in Eq. 5. The reduced basis modeling approach thus has two major assets: dimensionality reduction by the use of the basis coefficients and computational time reduction for approximating new solutions by the use of the reduced basis model. This paper presents a framework which uses both these aspects in order to more efficiently construct surrogate models.

4.2 Combining response surface methodology with reduced order modeling

The reduced basis modeling approach allows to significantly reduce the dimensionality of a response by expressing it in terms of its basis coefficients α in the reduced basis Φ. To create an approximation of the high dimensional response we will thus construct surrogate models of the basis coefficients α. Such an approach, though not novel, addresses challenge ii. presented in the introduction, namely constructing surrogate models of high dimensional responses.
However the issue of the construction cost, mentioned in challenge i. of the introduction, remains a hurdle. We thus propose to further use the reduced basis modeling to also decrease the surrogate construction cost. This is achieved by appropriately choosing some key points of the design of experiments at which the full problem is solved. At all the other points only the reduced order problem (full problem projected on the reduced order basis) needs to be solved. Figure 1 provides the flow-chart of the proposed procedure, that we call the key points response surface approach.

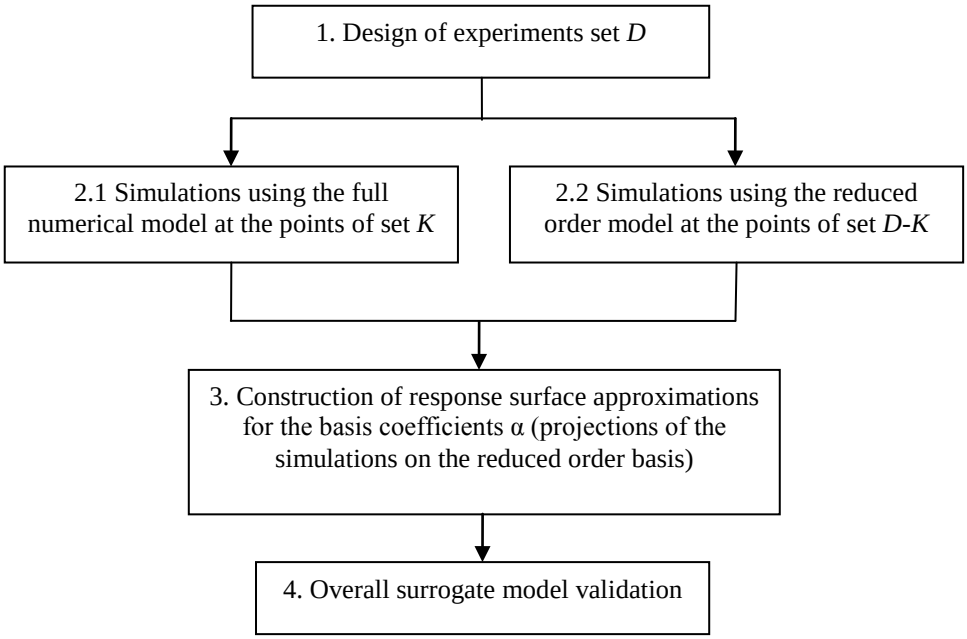


Figure 1. Flow-chart of the key points response surface construction steps.

In the first step the set of points *D* representing the design of experiments is defined. This set of points will serve both for the reduced dimensional basis extraction using the key points procedure as well as for the construction of the response surface approximations. Since both the key points procedure and the surrogate modeling benefit from a space-filling design, latin hypercube sampling will be used to construct the set D but other sampling methods are possible.

In the second step, full or reduced order simulations are run at the points of the design of experiments. The set of key points, defining the reduced dimensional basis Φ are denoted by *K*. The full numerical model is evaluated at the points of set *K*. On the other hand, at the remaining points (set *D-K*) an approximation of the response is calculated using the reduced order model (problem solved projected on the reduced dimensional basis Φ). Typically the cardinality of *K* is one to several orders of magnitude smaller than that of *D* as will be illustrated in the two application examples in sections 3 and 4. This is precisely where the interest of the proposed method lies: if the full resolutions need to be carried out only at the points *K* and that the inexpensive reduced basis model can be used at all the remaining points, this will lead to substantial computational savings for the construction of the surrogate model. Note also that the reduced order approximation satisfies the error criteria e_{rb} (cf. Eq. 5) imposed for the reduced basis modeling, which allows to control the quality of the approximation. Accordingly the value of *K* is not chosen beforehand but it is determined automatically by the proposed construction procedure such as to verify the error criterion.

At step 3 the full simulation results of the set *K* are projected onto the basis B to obtain their basis coefficients α. For the reduced order simulations of the set *D-K* we already have the coefficients α. Response surface approximations of the basis coefficients α are then constructed.

Finally the accuracy of the approximations is verified in step 4. Two error measures are already available. The first is the criterion e_{rb} on the residuals that was imposed during the reduced order modeling (cf. Eq. 5). The second consists in classical RSA error measures, such as root mean square errors or cross validation errors. An additional cross validation error measure can be considered on the whole procedure itself. The leave-one-out cross validation would consist in applying the key points procedure (steps 1-3 of Figure 1) not on the entire design of experiment set D but on a subset by leaving one point out. The procedure's prediction would then be compared to the simulation response at the point that was left out. Note that in the approach we propose we only have the exact finite element solution at the key points (at all the remaining points we only have the reduced basis solution). However since the key points were determined to be the most relevant solutions for constructing the reduced basis it seems appropriate to do the cross validation over these key points only. By doing all possible permutations of leaving a key point out and taking the root mean square of the errors in the prediction we obtain an error measure for the proposed surrogate modeling procedure.

4.3 Construction of the approximation subspace: choosing the key points

At this point we need to choose how to construct the reduced basis. In our approach, similarly to proper orthogonal decomposition (POD), the construction of the reduced basis requires full simulations, but only for a small set of samples, called key points. The choice of these key points is of importance because it has a strong effect on the accuracy of the response surface. Indeed, if the number of key points is too small or poorly chosen, the solutions computed by the reduced model will be inaccurate. Conversely, if the number of key points is too large, the basis would be large which reduces the numerical efficiency. In this paper, we investigate two ways to automatically choose a set of key points, as small as possible, but preserving an imposed accuracy of the reduced order model approximation.

The first approach consists in sequentially browsing the points $\boldsymbol{\mu}_i$ of the DoE. For the first point $\boldsymbol{\mu}_1$ of the DoE, the full simulation always needs to be carried out and its result $\boldsymbol{u_1}$ becomes the first vector of the reduced basis. Then at the point $\boldsymbol{\mu}_i$ it is assumed that one has already a reduced basis of size m_i. The problem for parameter $\boldsymbol{\mu}_i$ is then solved by projection on this reduced basis. This corresponds to the inversion of a small system of size m_i whose computational cost is low (often negligible) compared to that of the full simulation. The accuracy of the approximate solution thus constructed is evaluated with a measure of the residual error e_{rb}. If this indicator is below a certain threshold ε_{rb}, then we move on to the next parameter $\boldsymbol{\mu}_{i+1}$. Otherwise, the complete problem is solved for this point and the associated solution is orthogonalized as shown in Eq. 6, normalized and added to the basis.

$$\Phi_k = u_k - \sum_{i=1}^{k-1} \langle u_k, \Phi_i \rangle \Phi_i \qquad (6)$$

Where $\langle \bullet, \bullet \rangle$ denotes the L^2 scalar product.

The algorithm for this approach is presented below (Algorithm 1).

Algorithm 1 Sequential approach

$U_1 \leftarrow$ full simulation of point μ_1.
Reduced basis initialization : $B = U_1$.
for $i \in D$ **do**
  $\text{error}_i \leftarrow$ Reduced basis simulation of point μ_i
  if $\text{error}_i > e_{rb}$ **then**
    $U_i \leftarrow$ full simulation of point μ_i.
    New key point $\mu_i \in K$ and reduced basis enrichment : $B \leftarrow \{B, U_i\}$.
  end if
end for

This approach, called sequential approach in the sequel, has the advantage of considering points only once, so it has a certain advantage in terms of computation time. However, the choice of reference points (i.e. the key points) is highly dependent on the way the DoE is being browsed. The size of the subspace basis will therefore certainly not be minimal.

A second approach of selecting the key points, inspired from [39], consist in building these reference points iteratively at the points where the error is maximum. At iteration k, the basis $V^k=\{\Phi_1,\ldots,\Phi_{k-1}\}$ constructed from the full simulation of the problems with the reference sets of parameters $\left\{\boldsymbol{\mu}_1^{ref},\ldots,\boldsymbol{\mu}_{k-1}^{ref}\right\}$ are supposed to be known. Then an approximation of the solution of all the points of the DoE is computed with the reduced model projected onto the orthonormal basis V^k. The point where the error e_{rb} is the highest is chosen to be the next key point as shown in Eq. 7.

$$\boldsymbol{\mu}_k^{ref} = \arg\max_{\mu \in DoE}(e_{RB}(\boldsymbol{\mu})) \qquad (7)$$

A simulation of the problem with the set of parameter $\boldsymbol{\mu}_k^{ref}$ is then performed using the full model. The corresponding solution $\boldsymbol{u_k^{ref}}$ is orthonormalized and added to the reduced basis. Then one proceeds to the next iteration. The algorithm presented below (Algorithm 2) stops when the error is less than the threshold ε_{rb} everywhere. Unlike the previous approach, the entire DoE is covered at each iteration, but this method, that we call

maximum residual approach in the sequel, is supposed to reduce the cardinal of the reduced basis.

Algorithm 2 iterative max. residual approach

$U_1 \leftarrow$ full simulation of point μ_1.
Reduced basis initialization : $B = U_1$.
while error$> e_{rb}$ **do**
  for $i \in D$ and $error_i > e_{rb}$ **do**
    error$_i \leftarrow$ Reduced basis simulation of point μ_i
  end for
  k=argmax$(\{error_i\}_{i \in D})$;
  $U_k \leftarrow$ full simulation of point μ_k.
  New key point $\mu_k \in K$ and reduced basis enrichment : $B \leftarrow \{B, U_k\}$.
end while

5. Application example: surrogate of a thermal field

In order to illustrate the applicability and efficiency of the proposed methodology an application to the construction of a thermal field surrogate is considered in this section using kriging. Note that only the surrogate construction is sought in this section independently of any identification or other application problem that might involve the thermal field. Indeed the proposed surrogate methodology is independent of the surrogate use that might follow.

The application considered here is the heat transfer through the combustion chamber wall of a rocket engine [44],[45]. A schematic of a typical regeneratively cooled liquid hydrogen (LH2) liquid oxygen (LOX) rocket engine is illustrated in Figure 2. The regenerative cooling takes place when the liquid hydrogen (LH2) at a temperature of 40K flows through cooling channels in the combustion chamber wall before entering the injectors as shown in Figure 2. Determining the temperature field in the combustion chamber wall is important in order to account for the high thermally induced stresses that arise and which can lead to broken cooling channel walls (cf. Figure 2).

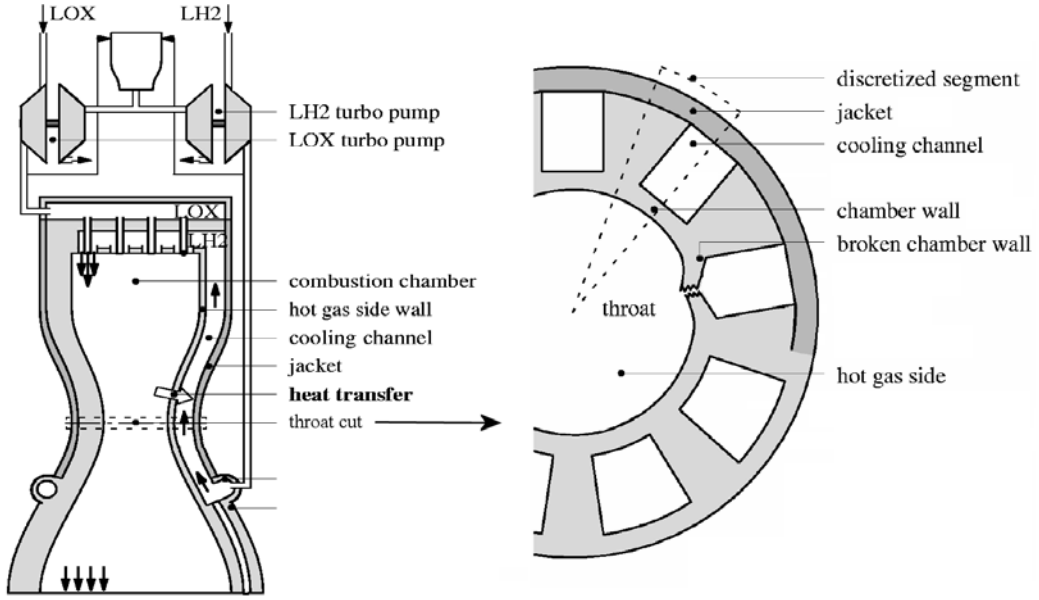


Figure 2: Schematic of a regeneratively cooled rocket engine combustion chamber. The drawings are not to scale.

Due to manufacturing constraints the combustion chamber wall is made of two different sections: an internal side usually made of a copper alloy in which the cooling channels are machined and an external jacket usually made of a Ni alloy (cf. Figure 2). Convective heat exchange can take place between the combustion chamber gases and the inner side of the combustion chamber wall, between the liquid hydrogen and the cooling channel side of the wall and between the outer environmental temperature and the external side of the combustion chamber wall. The corresponding heat transfer problem is thus parameterized by the following parameters: the conductivity of the inner side of the wall (k_{Cu}), the conductivity of the jacket (k_{Ni}), the temperature of the gases on the inner side of the combustion chamber (T_{hot}), the temperature on the outer side of the combustion chamber (T_{out}), the temperature of the cooling fluid (T_{cool}), the film convection coefficient on the inner side of the combustion chamber (h_{hot}), the film

convection coefficient on the outer side of the combustion chamber (h_{out}) and the film convection coefficient on the cooling channel side (h_{cool}).
We seek in this application example a surrogate of the temperature field in the combustion chamber wall as a function of the previous eight parameters during the stationary regime of the rocket engine.
For the design of experiments a finite element model of the combustion chamber wall is used. By symmetry considerations only the discretized segment illustrated in Figure 2 (right side) is modeled. An in-house finite element solver coded in Matlab is used, but the use of any commercial software is possible given that the "stiffness matrix" can be accessed. The problem's boundary conditions and finite element mesh is illustrated in Figure 3.

Figure 3: Schematic of a regeneratively cooled rocket engine combustion chamber. The drawings are not to scale.

The bounds of the design of experiments are provided in Table 1 and a total of 660 points were sampled by latin hypercube.

Table 1. Bounds for the design of experiments of the thermal problem.

Parameter	k_{Cu} (W/(mK))	k_{Ni} (W/(mK))	T_{hot} (K)	T_{out} (K)	T_{cool} (K)	h_{hot} (kW/(m^{2}K))	h_{out} (kW/(m^{2}K))	h_{cool} (kW/(m^{2}K))
Lower bound	50	200	700	260	30	20	6	100
Upper bound	130	400	950	310	70	40	20	350

The key points approach is applied with an error criterion $e_{rb}=10^{-3}$. The sequential approach leads to a total of 15 basis vectors. The convergence of the residuals error with the experiment number is illustrated in Figure 4 together with the error criterion threshold (red dotted line).

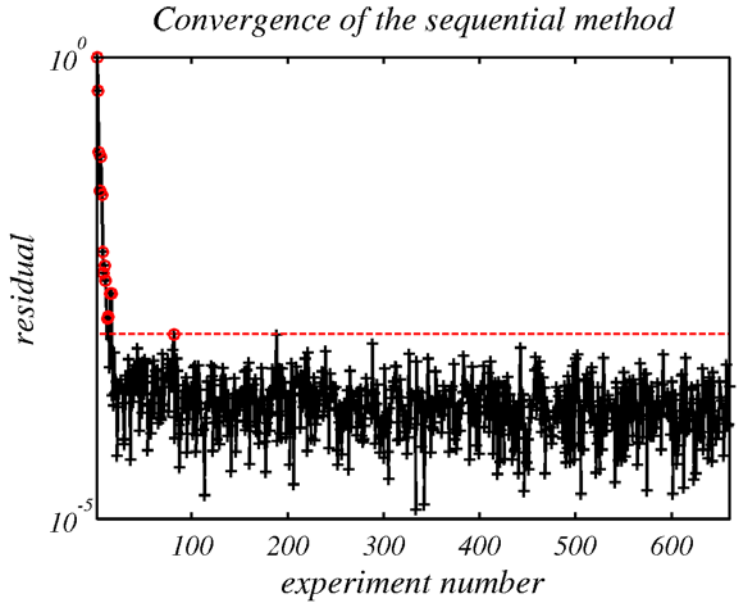


Figure 4: Plot of the temperature surrogate approximation error (log scale) for the sequential approach as a function of the DoE experiment number. The error criterion threshold is represented by the red dotted line.

The vast majority of the 660 points of the DoE are evaluated using the computationally inexpensive reduced basis model, only 15 full system resolutions being required. An overview and graphical representation of the first three temperature modes (basis vectors) is presented in Appendix 1.

The maximum residuals error approach also requires 14 iterations (and thus 15 basis vectors) to reach the 10^{-3} error criterion as shown in Figure 5. The figure plots the approximation error indicator based on the residual norm e_{rb} as a function of the iteration number. For iteration i, the approximation errors of the whole set of experiments are plotted as + symbols. The new key point (red circle) at each iteration is the one which corresponds to this maximum residual error. At the end of the procedure all residuals are below the threshold considered.
Given that the same number of basis vectors were obtained by the sequential and the maximum residuals error approach we chose to consider in the rest of this section only the sequential approach basis vectors due to the lower computational effort for obtaining them.

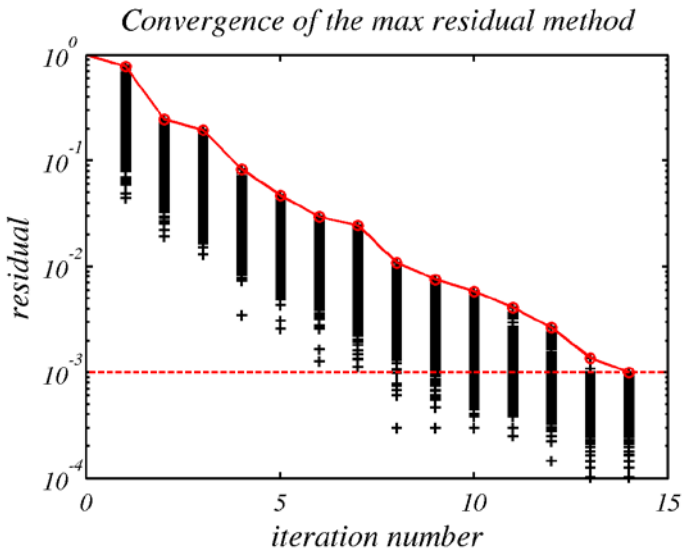


Figure 5: Plot of the temperature surrogate approximation error (log scale) as a function of the iteration number in the maximum residuals approach.

The reduced basis being constructed the next step consists in fitting a surrogate model to the basis coefficients. A kriging model was adjusted for each temperature basis coefficient using Surrogates Toolbox [42] (which itself utilizes a modified version of the DACE toolbox).
To validate the entire key points surrogate model cross validation is used again as described in section 2.2. The root mean square of the cross validation error over the 15 key points was found to be 0.105%. The individual errors at each of the cross validation points was quantified by the norm of the error in the temperature field divided by the norm of the temperature field, which thus led to relative errors.
Finally we provide an overview of the computational savings achievable with the proposed method for this thermal application problem. Table 2 gives the relative computational times needed for the resolution of the set of systems for the various methods discussed earlier. Note that the computational time was normalized such that the cost of the brute force approach is equal to 1.

Table 2. Relative computational cost associated to the resolution of the thermal DoE with the different techniques discussed in the paper.

	Computational time	Nb full systems	Nb projected systems	Size of the reduced-basis
Brute force	1	660	0	-
Max Res	0.27	15	7533	15
Sequential	0.1	15	659	15

For the brute force method (used as a reference), which consists in computing a full resolution for each experiment of the DoE, the resolution time is equal to one, while the maximum residual method leads to a speed-up of 3.7 and the sequential approach to a speed-up of 10, which is again a significant efficiency improvement. Note that for this problem there is no advantage in using the maximum residual error method since it finds the same reduced basis size as the sequential method.

6. Conclusions

The present article addressed the problem of constructing efficient surrogate models of high dimensional output. We proposed an approach based on the combination of reduced basis modeling and response surface methodology. The reduced basis modeling aims at solving the numerical problem projected on a reduced-dimensional basis, constructed sequentially based on a design of experiments. While such a reduced order modeling approach can be used as a metamodel by itself, even quicker response evaluations can be achieved by coupling the approach with response surface methodology. The proposed approach seeks to construct surrogate models of the basis coefficients that require a reduced number of expensive function evaluations made possible by the key points approach: the full scale (expensive) problem is only solved at a small number of key DoE points, while the reduced order model is used at all the others. Note that this approach is more code-intrusive than classical response surface construction, since the "stiffness" matrix needs to be available. It has however the potential of significant computational cost savings.

An application example was presented where a surrogate of a thermal field was sought. The proposed method allowed here to reduce the computational cost of the construction phase of the surrogate by an order of magnitude. It is to be noted that we applied here our approach to a problem where the entire high dimensional response was sought (the full temperature field), but note that the presented procedure may also be efficient for scalar responses that are extracted out of a higher dimensional output (e.g. maximum stress extracted out of the full stress map in the structure, or maximum temperature). In such a case our approach would allow to track the maximum even if it changes location on the structure. Compared to classical response surface approximation construction that require full scale problem solving at each design of experiment point our approach still has the potential to greatly reduce the number of required full scale problems solved.

Finally it is important to note that the larger the problem (in terms of number of nodes) the higher the potential of the proposed method in terms of surrogate construction speed-up.

7. Acknowledgements

Financial support by the by the French National Research Agency (ANR) through the COSINUS program under the grant ANR-09-COSI-005 is gratefully acknowledged.

8. References

[1] Sacks J, Welch WJ, Mitchell TJ, and Wynn HP, "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4 (4), pp. 409-435, 1989.

[2] Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, and Tucker PK, "Surrogate-based Analysis and Optimization," *Progress in Aerospace Sciences*, Vol. 41, pp. 1-28, 2005.

[3] Kleijnen JPC, Sanchez SM, Lucas TW, and Cioppa TM, "A user's guide to the brave new world of designing simulation experiments," *INFORMS Journal on Computing*, Vol. 17, No. 3, pp. 263-289, 2005.

[4] Chen VCP, Tsui K-L, Barton RR, and Meckesheimer M, "A Review on Design, Modeling and Applications of Computer Experiments," *IIE Transactions*, Vol. 38, No. 4, pp. 273–291, 2006.

[5] Forrester AIJ and Keane AJ, "Recent advances in surrogate-based optimization," *Progress in Aerospace Sciences*, Vol. 45, No. 1-3, pp. 50-79, 2009.

[6] Simpson TW, Toropov V, Balabanov V, and Viana FAC, "Metamodeling in Structural and Multidisciplinary Optimization: How Far We Have Come – or Not," *Structural and Multidisciplinary Optimization*, Accepted for publication, 2010.

[7] Myers, R. H., and Montgomery, D. C., *Response Surface Methodology: Process and Product in Optimization Using Designed Experiments.* John Wiley & Sons, New York, 2002.

[8] Stein ML, Interpolation of Spatial Data: Some Theory for Kriging, Springer, 1999.

[9] Simpson TW, Mauery TM, Korte JJ and Mistree F, "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal*, 39(12), 2233-2241, 2001.

[10] Kleijnen JPC, "Kriging metamodeling in simulation: A review," *European Journal of Operational Research*, Vol. 192, No. 3, pp. 707-716, 2009.

[11] Park J and Sandberg IW, "Universal Approximation Using Radial-Basis-Function Networks," *Neural Computation*, Vol. 3, No. 2, pp. 246-257, 1991.

[12] Cheng B and Titterington DM, "Neural networks: a review from a statistical perspective," *Statistical Science*, Vol. 9, No. 1, pp. 2-54, 1994.

[13] Goel T and Stander N, "Comparing Three Error Criteria for Selecting Radial Basis Function Network Topology," *Computer Methods in Applied Mechanics and Engineering*, Vol. 198, Issues 27-29, pp. 2137-2150, 2009.

[14] Mika S and Tsuda K, An introduction to kernel-based learning algorithms, *IEEE Transactions on Neural Networks*, Vol. 12, No. 2, pp. 181-201, 2001.

[15] Scholkopf B and Smola AJ, Learning with Kernels, The MIT Press, Cambridge, USA, 2002.

[16] Smola AJ and Scholkopf B, "A tutorial on support vector regression," *Statistics and Computing*, Vol. 14, No. 3, pp. 199-222, 2004.

[17] Viana, F.A.C., Gogu, C., Haftka, R. T., 2010, "Making the most out of surrogate models: tricks of the trade", *ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Montreal, Canada, 15-18th August 2010

[18] Daniel, C., 1973, "One-at-a-Time Plans," *Journal of the American Statistical Association*, 68, pp. 353-360.

[19] Sobol I., 1993, "Sensitivity estimates for non-linear mathematical models". Math Modeling Comput Exp, Vol. 4, pp. 407–14.

[20] Gogu C, Haftka RT, Bapanapalli SK, and Sankar BV, 2009, "Dimensionality Reduction Approach for Response Surface Approximations: Application to Thermal Design," *AIAA Journal*, Vol. 47, No. 7, pp. 1700-1708.

[21] Khuri AI and Cornell JA, *Response Surfaces: Designs and Analyses*, 2nd edition, Dekker, Chapter 7, 1996.

[22] Myers, DE, "Matrix formulation of co-kriging", *Mathematical Geology*, Vol. 14, No. 3, pp. 249-257, 1982.

[23] Yee, TW, "Vector splines and other vector smoothers," Proc. Computational Statistics COMPSTAT 2000. Bethlehem, J.G., Van der Heijden, P.G.M. (eds.), Physica-Verlag, Heidelberg, 2000.

[24] Coelho RF, Breitkopf P, and Vayssade, CK-L, "Model reduction for multidisciplinary optimization-application to a 2D wing," *Structural and Multidisciplinary Optimization*, Vol. 37, No. 1, pp. 29-48, 2008.

[25] Coelho RF, Breitkopf P, Vayssade CK-L, and Villon P, "Bi-level model reduction for coupled problems," *Structural and Multidisciplinary Optimization*, Vol. 39, No. 4, pp. 401-418, 2009.

[26] De Lucas, S., Vega, J.M., Velazquez, A., "Aeronautic Conceptual Design Optimization Method Based on High-Order Singular Value Decomposition", AIAA Journal,Vol.49, No.12, pp. 2713-2725, 2011.

[27] Wehrwein D. and Mourelatos Z.P., "Reliability based design optimization of dynamic vehicle performance using bond graphs and time dependent metamodels," SAE Technical Paper 2006-01-0109, 2006, doi:10.4271/2006-01-0109.

[28] Yin, X., Lee, S., Chen, W., Liu, W.K. and Horstemeyer, M.F., "Efficient random field uncertainty propagation in design using multiscale analysis," *Journal of Mechanical Design*, Vol. 131 (2), 021006 (10 pages), 2009.

[29] Singh G. and Grandhi, R.V., "Propagation of structural random field uncertainty using improved dimension reduction method," 50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Palm Springs, California, AIAA paper 2009-2254, 2009.

[30] P. Kerfriden, P. Gosselet, S. Adhikaric, and S.P.A. Bordas. “Bridging proper orthogonal decomposition methods and augmented newton–krylov algorithms : An adaptive model order reduction for highly nonlinear mechanical problems.” *Comput. Meth. Appl. Mech. Eng.*, 200(5–8) :850–866, 2011.

[31] P. Krysl, S. Lall and J.E. Marsen. “Dimensional model reduction in non-linear finite element dynamics of solids and structures.” *International Journal for Numerical Methods in Engineering*, 51:479:504, 2001.

[32] D. Ryckelynck, “A priori hypereduction method : an adaptive approach”, *Journal of Computational Physics*, 202(1):346 - 366, 2005.

[33] K. Kunisch and S. Volkwein. “Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics.” *SIAM Journal on Numerical Analysis*, 40(2) :492–515, 2002.

[34] K. Karhunen. “Uber lineare methoden für wahrscheinigkeitsrechnung.” *Annales of Academic Science Fennicae Series A1 Mathematical Physics*, 37:3-79, 1943.

[35] M.M. Loeve. *Probability Theory*. Van Nostrand, Princeton, NY. 1955.

[36] H. Hotteling. “Analysis of Complex of statistical variables into principal component.” *Journal of Educational Psychology*, 24:417--441, 1933.

[37] A. Nouy. “A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations.” *Comput. Methods Appl. Mech. Eng.*, 196(45-48):4521–4537, 2007.

[38] J.-C. Passieux, P. Ladeveze, and D. Neron. “A scalable time-space multiscale domain decomposition method : adaptive time scale separation.” *Comput. Mech.*, 46(4) :621–633, 2010.

[39] Grepl, M.A., Maday, Y., Nguyen, N.C., Patera, A.T., “Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations”, *ESAIM: Mathematical Modelling and Numerical Analysis*, Vol. 41, pp. 575-605, 2007, DOI: 10.1051/m2an:2007031.

[40] P. Kerfiden, J.-C. Passieux and S.P.A. Bordas, “ Local/global model order reduction strategy for the simulation of quasi-brittle fracture”, *International Journal for Numerical Methods in Engineering*, 89(2):154-179, 2012.

[41] A.I.J. Forrester and A.J. Keane, “Recent advances in surrogate-based optimization”, *Progress in Aerospace Sciences*, 45:50-79, 2009.

[42] F.A.C. Viana, “SURROGATES Toolbox User’s Guide”, version 2.1, http://sites.google.com/site/fchegury/surrogatestoolbox, 2010.

[43] Allen, D. M., "Mean square error of prediction as a criterion for selecting variables." *Technometrics,* 13, 469-475,1971.

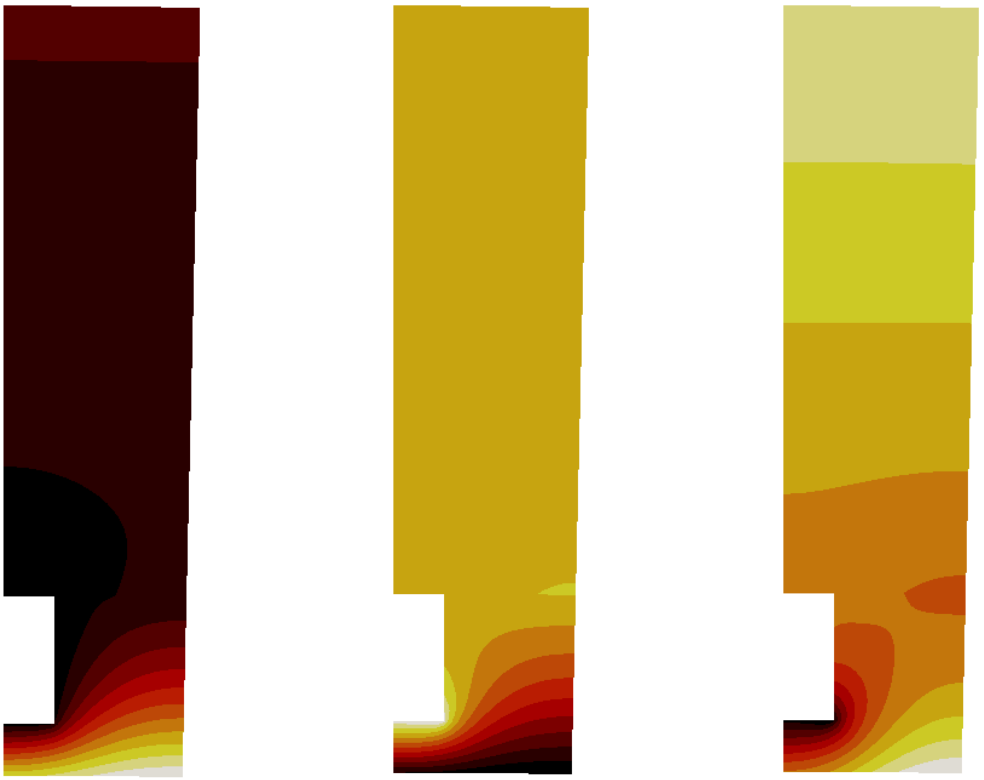
[44] Riccius, J.R., Haidn, O.J. and Zametaev, E.B., “Influence of time dependent effects on the estimated life time of liquid rocket combustion chamber walls,” AIAA 2004-3670, *40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, Fort Lauderdale, Florida , July 2004 .

[45] Riccius, J., Gogu, C., Zametaev, E., Haidn, O., “Liquid Rocket Engine Chamber Wall Optimization Using Plane Strain and Generalized Plane Strain Models”, AIAA paper 2006-4366, *42nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, Sacramento, CA, July 2006.

Appendix 1: Temperature modal basis

The basis vectors of the thermal reduced basis can be represented graphically and have following physical interpretation: the basis vectors can be seen as temperature modes characterizing the effect of the variations in the eight problem parameters considered (conductivities, film convection coefficients and wall temperatures). We illustrate the first three modes in Figure 6.

By construction the first mode is the dominant mode, meaning that alone it will explain most of the variations of the temperature fields when the eight parameters of the problem are varied. Of course the subsequent modes are required to catch finer details in the variations of the temperature field and thus meat the 10^{-3} error criterion.



Mode 1 Mode 2 Mode 3

Figure 6: First three temperature modes obtained using the key points basis construction procedure.