Comparison of surrogate models for the actual global optimization of a 2D turbomachinery flow

J. PETER, M. MARCELET, S. BURGUBURU
ONERA
Department of CFD and Aeroacoustics
BP 72 - 29 av. de la Div. Leclerc, 92322 Chtillon
FRANCE

V. PEDIRODA
Universit degli Studi di Trieste
Dipartimento di Ingegneria Meccanica
via Valerio 10, 34100, Trieste
ITALY

Abstract: This article adresses the issue of selecting surrogate models suitable for the global optimization of 2D turbomachinery flows. As a first step towards this goal the analysis of a family of flows on a two-parameter design space is presented. Four types of surrogate models are considered : least square polynomials, artificial neural networks (multi-layer perceptron and radial basis function) and Kriging. Discussed is the ability of these surrogate functions to give a satisfactory description of the exact function of interest on the design space, during a global optimization. The number of CFD evaluations for an adequate description of the exact function is presented.

Key–Words: Turbomachinery, global optimization, surrogate model

1 Introduction

Shape optimization is one of the most important application of computational fluid dynamics (CFD). For example, the drag-reduction of an aircraft with constraints on the lift, geometry and momentums is a prominent issue of external aerodynamics. In the field of internal flows, minimization of total-pressure losses of a blade row is an important and classical issue. Despite the huge amount of work devoted to aerodynamic shape optimization during the three last decades, no specific algorithm has appeared to be really adequate for all problems or at least for a very wide range of problems.

Since the mid 70's and the landmark papers of Hicks and VanderPlaats, local optimization using the gradient of the functions of interest with respect to the design parameters have focused much attention [12]. In the late 80's and at the beginning of the 90's it appeared that those gradients could be computed by the so-called adjoint vector method [4] or direct differentiation method [5] instead of the costly finite difference method. Local optimization of a parametrized solid shape, combining adjoint vector method, a descent method -like feasible descent [7] - and some kind of mesh deformation tool became very popular. ONERA has developped both discrete adjoint vector and discrete direct method in the aerodynamic code $elsA$ [1, 3, 15], and demonstrated its ability to carry out optimization of 3D industrial configurations [2].

In other respects, several authors considered the issue of global aerodynamic optimization. Almost all types of global optimization strategies were considered with a significant emphasis on genetic algorithms [6]. The authors interested in this method had to face the bottleneck of the huge cost of the numerous exact evaluations of design requested by the optimization algorithm. To circumvent this issue, many authors replaced some of the exact evaluations (by the CFD and post-processing code) by those provided by a well defined surrogate-model [9, 10].

This article deals with the problem of global optimization for transonic flows [11], focusing on the actual search for the most adapted surrogate model for a turbomchinery design optimization problem. It is organized as follows. Geometry, governing equations and design space are presented in section 2. The surrogate models we consider are detailed in section 3. The ability of these surrogate functions to give a satisfactory description of the exact function of interest on the design space is discussed in section 4.

2 Description and analysis of the turbomachinery flow

2.1 Nominal geometry

The considered test case is derived from the stator blade of VEGA2 configuration, which is a classical stator rotor turbine configuration [13]. Due to the high cost of global optimisation, only a 2D geometry deduced from an appropriate projection of the 3D geometry at the hub, is studied. The 4 domains mesh

with matching joins is presented in figure 2, where the caracteristic length of the blade, $L = 90mm$, can be measured. The total number of mesh points is 11816. Periodicity conditions are applied at the lower and upper borders. The aerodynamic data of the

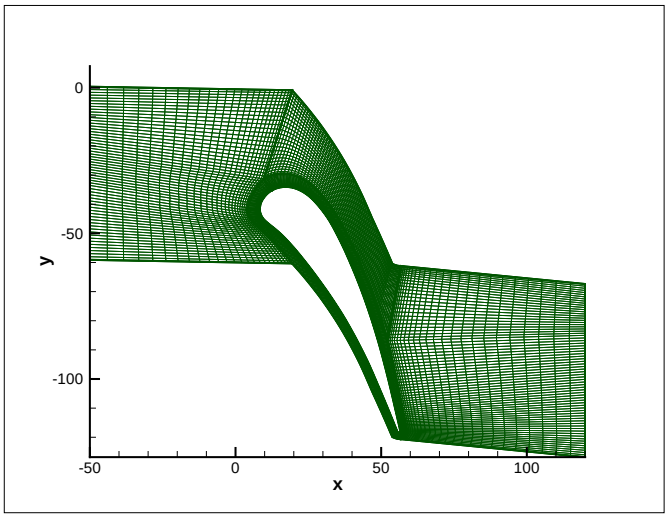

Figure 1: mesh of nominal 2D configuration

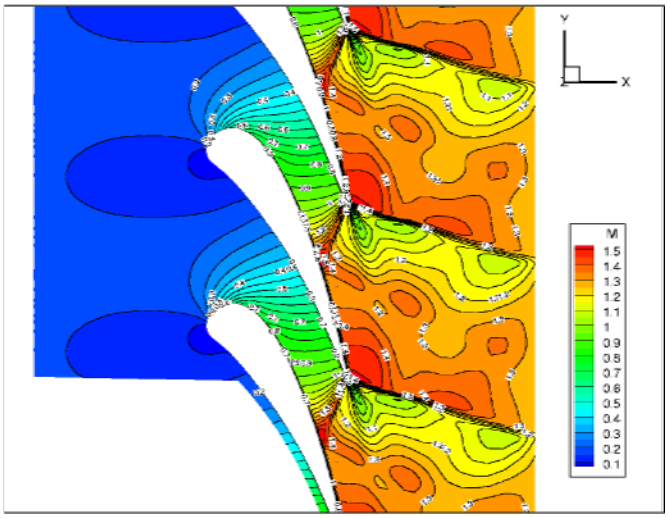

Figure 2: iso Mach-number lines of nominal 2D configuration

subsonic inlet are $T_i = 288,27K$, $p_i = 101325Pa$. The direction of the flow at the inlet is also constrained along the x axis. At the outlet the static pressure is fixed. Its ratio to the inlet total pressure is $p_{s\ exit}/p_i = 0.35$. The eddy viscosity is determined by Sutherland law The Reynolds number of the flow based on the stagnation condition is then $Re = (\rho_i a_i L)/\mu(T_i) = 2,09.10^5$

2.2 Flow computation

The Reynolds averaged Navier-Stokes equations are considered. The turbulent viscosity is computed by Smith $k - l$ model. The seven-equation non-linear system is solved numericaly by the ONERA finite-volume cell-centred code for structured meshes, called $elsA$ [15]. Second order Roe-flux (using MUSCL approach with Van Albada limiting function) is used for mean flow convective term, the first order Roe flux is used for turbulent variable convective term, centred fluxes with interface centred evaluation of gradients are used for both diffusive terms. Centred formula is used for the source term of turbulent variables equations. More details can be found in reference [14], which also indicates a good comparison with experimental data for the original 3D geometry.
Due to the low value of the static pressure at the exit, the flow is sonic at the narrowest section between two blades, near the trailing edge (just like in a shocked nozzle). Two strong shock lines (one going along the x axis, the other being oblique) start from the trailing edge of the blade. A view of the iso-Mach number lines in presented.

2.3 Design space

The geometric deformation of the blade consists in moving the trailing edge along both x and y axis. The leading edge is fixed. The deformed shape of the blade is defined by a smooth algebraic function of the curvilinear coordinate. The displacement is damped out from the solid shape to the fixed boundary of the blade domain (see mesh plot). The maximum displacement in each direction is ± 0.4 mm. The displacement along the x axis is the first design parameter α_1, the displacement along the y axis is the second design parameter α_2. The main output of the computation is the total pressure at the exit, computed by integration on the exit surface. Its non dimensional value (actual value divided by inlet value) varies from 0.918 to .924, on the design space. Of course such low values appear because of the strong shocks. This variation is large enough to define an optimization problem.
A large regular sampling of design space with 21×21 points is considered. All corresponding flows are computed with exactly the same numerical parameters. The explicit space residual of the scheme is decreased for all design by four to five orders of magnitude for all computations. The plot of the exit total pressure on the design space is presented. It was checked that the variation of total pressure when the design changes corresponds to a change in the strength of the oblique shock.

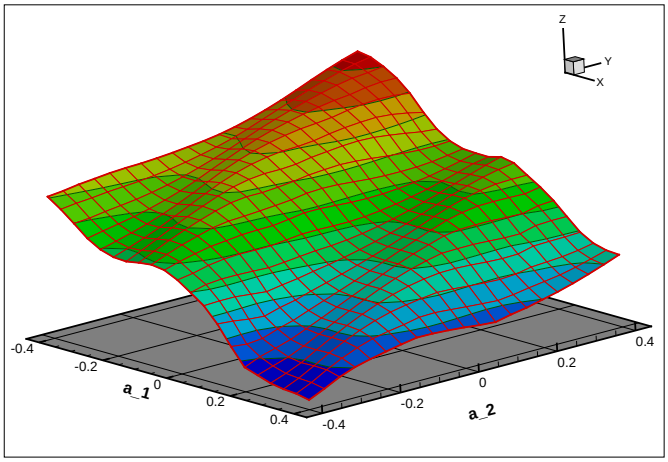

Figure 3: exit total pressure as function of design parameters (axis plane does not correspond to zero)

3 Brief description of the surrogate models

The exact function of interest (exit total pressure for the application) is $\mathcal{J}(\alpha)$. The description of the surrogate models is limited to the case of a two-component design vector α. The number of available exact evaluations of function $\mathcal{J}(\alpha)$ is noted n_s. The mean square error (MSE) on the sampling between the exact function $\mathcal{J}(\alpha)$ and the surrogate model $\overline{\mathcal{J}}(\alpha)$ is denoted by $\mathcal{E}$.

$$\mathcal{E} = \frac{1}{2}\sum_{i=1}^{n_s}\left(\mathcal{J}(\alpha_1^i,\alpha_2^i) - \overline{\mathcal{J}}(\alpha_1^i,\alpha_2^i)\right)^2$$

The vector of the exact evaluations is $\mathcal{J}_s$. $\mathcal{J}_s = \left[\ \mathcal{J}(\alpha_1),\quad \ldots\quad ,\mathcal{J}(\alpha_{n_s})\ \right]^T$

3.1 Least square polynomials

This method is both simple and well-known. Hence its presentation is limited to a degree two polynoms, altough polynoms of degree two, four, six and eight have been considered for the application. Suppose

$$\overline{\mathcal{J}}(\alpha) = \Psi_0+\Psi_1\alpha_1+\Psi_2\alpha_2+\Psi_{11}\alpha_1^2+\Psi_{12}\alpha_1\alpha_2+\Psi_{22}\alpha_2^2$$

The coefficients of the polynom are found by minimising the MSE on the sampling. This leads to

$$\Psi = (X^TX)^{-1}X^T\mathcal{J}_s$$

With $\Psi = [\Psi_0\ \ \Psi_1\ \ \Psi_2\ \ \Psi_{11}\ \ \Psi_{12}\ \ \Psi_{22}]^T$ and

$$X = \begin{bmatrix} 1 & \alpha_1^1 & \alpha_2^1 & (\alpha_1^1)^2 & \alpha_1^1\alpha_2^1 & (\alpha_2^1)^2 \\ 1 & \alpha_1^2 & \alpha_2^2 & (\alpha_1^2)^2 & \alpha_1^2\alpha_2^2 & (\alpha_2^2)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha_1^{n_s} & \alpha_2^{n_s} & (\alpha_1^{n_s})^2 & \alpha_1^{n_s}\alpha_2^{n_s} & (\alpha_2^{n_s})^2 \end{bmatrix}$$

3.2 Multilayer perceptron

Although a wide range of multi-layer perceptrons can be conceived, refering to the universal approximation theorem for neural networks[16], we have decided to use the multi-layer perceptron with just one hidden layer pictured in Fig 4. The activation function of the hidden layer units is the sigmoide function, and the final output of the network is simply a weighted sum of the hidden layer outputs. A bias value is added to the inputs and to the outputs of the hidden layer. Given a set of n_s exact computed responses, the learning process aims at determining the set of the $4n_c+1$ unknown coefficients $\omega = [a_{11},\ldots,b_{n_c}]$ so that the mean squared error $\mathcal{E}$ is minimal, where

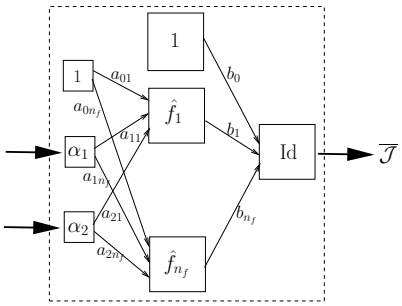

Figure 4: two layers perceptrons

$$\overline{\mathcal{J}}(\alpha_1,\alpha_2) = b_0 + \sum_{j=1}^{n_c} b_j tanh\left(a_{0j} + a_{1j}\alpha_1 + a_{2j}\alpha_2\right)$$

To find such a set of weights, a steepest descent optimization is performed. Once the gradient of the MSE with respect to the unknown coefficients is calculated, the Wolfe method is used to minimize the MSE in the gradient opposite direction. An iterative process is carried out till the gradient value is small enough. The initialization of the unknown coefficients set is an important matter for the method. In the present work, the initial guess of the gradient based search is chosen randomly.

3.3 Radial basis function network

The radial basis function network[17] used in this study is composed of n_c radial functions $\hat{f}_i$

$$\hat{f}_i(\alpha_1,\alpha_2) = exp\left(-\frac{1}{2}\frac{(\alpha_1-\alpha_1^i)^2+(\alpha_2-\alpha_2^i)^2}{r_i^2}\right)$$

where (α_1^i,α_2^i) and r_i are respectively the center and the radius of the radial function. The output of the network is given by the following formula:

$$\overline{\mathcal{J}}(\alpha_1,\alpha_2) = \sum_{i=1}^{n_c} a_i\hat{f}_i(\alpha_1,\alpha_2)$$

where $A = [a_1, a_2, \ldots, a_{n_c}]$ is a set of coefficients to determine depending on the exact function to approximate. Given the vector of n_s ($n_s \geq n_c$) exact values $\mathcal{J}_s$, the minimization of the (MSE) leads to $A = (X^T X)^{-1} X^T \mathcal{J}_s$ where X is the following $n_s \times n_c$ matrix

$$X = \begin{bmatrix} \hat{f}_1(\alpha_1^1, \alpha_2^1) & \ldots & \hat{f}_{n_c}(\alpha_1^1, \alpha_2^1) \\ \vdots & \vdots & \vdots \\ \hat{f}_1(\alpha_1^{n_s}, \alpha_2^{n_s}) & \ldots & \hat{f}_{n_c}(\alpha_1^{n_s}, \alpha_2^{n_s}) \end{bmatrix}$$

Taking the assumption that the number of centers has been fixed, the RBF approximation model is fully determined once the center and radius of every function is chosen. In this article, every function has the same radius

$$r = \frac{1}{n_c} \left(\max_{1 \leq i,j \leq n_s} \sqrt{(\alpha_1^i - \alpha_1^j)^2 + (\alpha_2^i - \alpha_2^j)^2} \right)$$

Moreover, we have picked the radial function centers to coincide with the exact function evaluations points, so that $A = X^{-1} \mathcal{J}_s$

3.4 Kriging

Considering that the value of $\mathcal{J}$ at the center of the design space is a good approximation of its mean value m, simple Kriging is considered [8]. The statistical basis of the method cannot be described in the limited space of this article. Based on n_s sampling points, the formula of the simple Kriging is a linear interpolation of the known values (applied to $\mathcal{Z}(\alpha) = \mathcal{J}(\alpha) - m$)

$$\overline{\mathcal{Z}}(\alpha) = {K_\alpha}^T C^{-1} \mathcal{Z}_s = \mathcal{Z}_s^T C^{-1} K_\alpha$$

(as matrix C is symmetric) with

$$K_\alpha = [Cv(\mathcal{Z}(\alpha), \mathcal{Z}(\alpha^1)), ..., Cv(\mathcal{Z}(\alpha), \mathcal{Z}(\alpha^{n_s})]^T$$
$$\mathcal{Z}_s = [\mathcal{Z}(\alpha^1), \ldots, \mathcal{Z}(\alpha^{n_s})]^T$$

$$C = \begin{bmatrix} Cv(\mathcal{Z}(\alpha^1), \mathcal{Z}(\alpha^1)) & \ldots & Cv(\mathcal{Z}(\alpha^{n_s}), \mathcal{Z}(\alpha^1)) \\ \vdots & \vdots & \vdots \\ Cv(\mathcal{Z}(\alpha^1), \mathcal{Z}(\alpha^{n_s})) & \ldots & Cv(\mathcal{Z}(\alpha^{n_s}), \mathcal{Z}(\alpha^{n_s})) \end{bmatrix}$$

The method is fully defined when the function Cv is selected (it is the covariance of the function $\mathcal{Z}$ in the statistical framework of the method description). Most often the following function is casted $Cv(\mathcal{Z}(\alpha^a), \mathcal{Z}(\alpha^b)) = \sigma^2 exp(-\theta ||\alpha^a - \alpha^b||)$ The parameters (σ, θ) are defined in this study as proposed by Jouhaud et al. [11].

4 Evaluation of the surrogate model for design optimization

The goal of this section is to discuss the efficiency of the surrogate models for the sake of optimization. For the studied two-parameter problem, the computation of the surrogate models coefficients is neglectible compared to the cost of one CFD computation. For this reason the efficiency of a surrogate approximations for the optimization problem can be measured by the requested number of exact evaluations.
For all surrogate functions the strategy of sampling enrichment is the same :
a- start with a large enough sampling to determine all coefficients. This initial sampling is built on latin hypercubes.
b- add points if criterion (C*) -see below- is not achieved
b1- if the min and max locations of the surrogate model are not all in the sampling add the missing one
b2- else add to the sampling the four points with maximum distance to the points of the sampling

4.1 Definition of the evaluation criterion. Main results

The function of interest exhibits one global maximum (-0.4,0.4), one local maximum (0.04, 0.4), two local minima (-0.2,-0.4), (0.4,0.08) and one global minima (0.4,-0.4) on the 21×21 sample. A surrogate reconstruction will be tested against the following criterion:
(C1) abilitiy to build an approximation with mean error E on the 21×21 sampling lower than 2.10^{-3}. The mean error E being adimensioned by the variation of exit static pressure on the design space

$$E = \frac{\sqrt{\sum_{i,j=1}^{21} \left(p_i(\alpha_1^{ij}, \alpha_2^{ij}) - \overline{\mathcal{J}}(\alpha_1^{ij}, \alpha_2^{ij}) \right)^2}}{21^2 \times (p_{i\,max} - p_{i\,min})}$$

(C2) find the two local maxima, their location being exact or in a neighboring point of the exact place on the 21×21 sampling.
(C3) find the global and local maxima at the right place on the 21×21 sampling.
The reason for (C2) is that the second order derivatives values are rather low near the maxima so that (C3) is difficult to reach. The results are summarized in table 1. (IT) indicates the number of exact evaluations needed to satisfy a criterion. Indicated is also the value of E after 100 CFD computations for all four surrogate models.

Sur. Mod	C1	IT	C2	IT	C3	IT	E(100)
Pol. 2	KO	-	KO	-	KO	-	$2.3e^{-3}$
Pol. 4	KO	-	KO	-	KO	-	$2.1e^{-3}$
Pol. 6	OK	56	KO	-	KO	-	$1.2e^{-3}$
Pol. 8	OK	73	OK	73	OK	73	$4.9e^{-4}$
Mu.Pe.	OK	300	KO	-	KO	-	$2.5e^{-3}$
RBF	OK	37	OK	38	OK	46	$1.6e^{-3}$
Sim Kri	OK	26	OK	62	OK	170	$4.0e^{-4}$

Table 1: Summary of surrogate model performances

4.2 Discussion of the results

From a general point of view, it is clear that RBF and Simple Kriging lead to the best results. Both of them satisfy the (C2) criterion with a reasonable number of exact evaluations. As concerning (C3) only (RBF) and degree-8 polynom satisfy it with an acceptable number of exact evaluations. The surrogate function surfaces satisfying (C3) are presented in figures 5 and 6. More details are given below concerning the different surrogate models.

• Least square polynomials

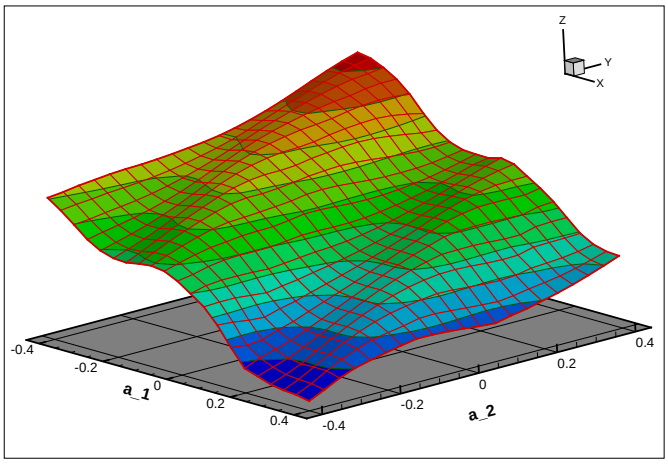

Figure 5: Kriging surface satisfying (C3)

Obviously the exact surface cross-sections (α_1 =const, or α_2=const on Fig.3) are much more complicated than parabols, which means that the exact function can not be well fitted with a second order polynom. Considering the plots obtained with degree-4 and 6 polynoms, this seems also to be the case. As concerning the degree 8 polynom, construction algorithm is started with a 45 sampling (number of coefficients). At step (b1) 28 points are added leading to 73 points

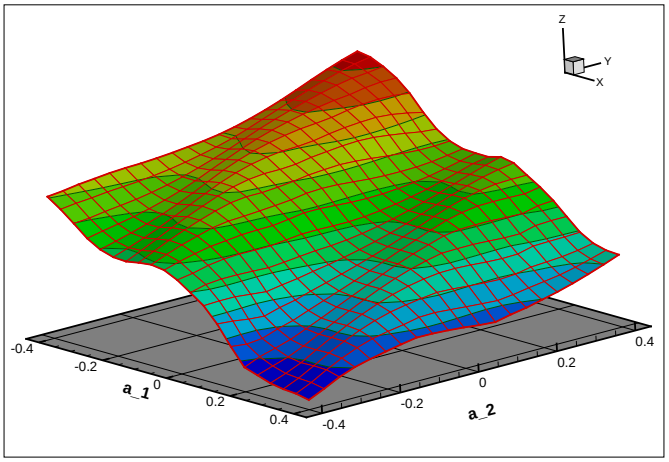

Figure 6: RBF surface satisfying (C3)

for the second sampling. A more sophisticated strategy of sampling enrichment could have led to lower numbers of exact evaluations to satisfy the (C*) criterions. This surrogate function obtained based on the 73 exact evaluations has almost the same aspect as the exact one.

• Radial basis function network

It has also been checked for (RBF) network that the results depend only slightly on the initial six-point sampling.

• Multi-layer perceptron

Two multi-layer perceptrons have been tested, respectively with 5 and 10 units in the hidden layer. The first requires almost the whole exact CFD evaluations to satisfy C_1, as for the second, up to 200 evaluations are actually needed. But in spite of the global optima being located from 42 evaluations, none of these perceptrons succeeds in locating the local optima. Besides, from a sampling of 50 evaluations, the computed error does not vary much from $2.4\ 10^{-3}$.

• Simple Kriging

This method leads to the lower error (E) after a definite number of iterations. Nevertheless it is not the most efficient for the accurate detection of the two maxima.

5 Conclusion

This article presented how four types of surrogate models can be used in an industrial context to design a stator blade so as to optimize the total pressure at the exit. Among all the models that have been tested, the simple Kriging model and the radial basis

function network appear to give the best results in terms of approximation of the exact function. However, the Kriging model used in this study has not taken advantage of its error estimation, which could have improved the approximation results. Besides, including the available gradient information could also be used as a way to enhance the level of approximation reached by the best models. Only a 2D configuration has been considered, we plan to extend the framework depicted here to 3D cases.

Acknowledgements: This work was supported by the project NODESIM-CFD "Non-Deterministic Simulation for CFD-based Design Methodologies" funded by the European Community represented by the CEC, Research Directorate-General, in the 6th Framework Programme, under Contract No. AST5-CT-2006-030959.

References:

[1] J. Peter, *Discrete adjoint method in $elsA$ (part I) : method/theory.* Proceedings of 7^{th} ONERA-DLR Aerospace Symposium. 2006.

[2] I. Salah el Din, G. Carrier, S. Mouton *Discrete adjoint method in $elsA$ (part II) : application to aerodynamic design optimisation.* Proceedings of 7^{th} ONERA-DLR Aerospace Symposium. 2006.

[3] J. Peter, F. Drullion, Large stencil viscous flux linearization for the simulation of 3D turbulent flows with backward Euler schemes. *Computers and Fluids* 36, 2007, pp. 1007–1027.

[4] A. Jameson, Aerodynamic design via control theory *Journal of Scientific Computing* 3(3), 1988, pp. 233–260

[5] O. Baysal, M. Eleshaky, Aerodynamic design sensitivity analysis method for the compressible Euler equations *Journal of Fluids Engineering* 113(4), 1991, pp. 681–688

[6] D. E. Goldberg, *Genetic algorithms in search optimization and machine learning.* Addison Weasley. 1989.

[7] G. N. Vanderplaats, *Numerical optimization for engineering design* 3rd ed., VR and D. 1999.

[8] J. Sachs, S. Schiller, W. Welch, Designs for computer experiments *Technometrics* 31, 1989, pp. 41–47

[9] C. Poloni, P. Loris, L. Larussi, S. Pieri, V. Pediroda, Robust Design of Aircraft Components : a multi-objective optimisation problem. *VKI Lecture series 2004-07*, 2004

[10] K.C. Giannacoglou, D.I. Papadimitriou,I.C. Kampolis, Coupling evolutionary algorithms, surrogate models and adjoint methods in inverse design and optimization problems *VKI Lecture series 2004-07*, 2004

[11] J.-C. Jouhaud, P. Sagaut, M. Montagnac, J. Laurenceau, A surrogate-model based multidisciplinary shape optimization method application to a 2D subsonic airfoil. *Computers and Fluids* 36, 2007, pp. 520–529

[12] G. VanderPlaats, R. Hicks, *Numerical airfoil optimization using a reduced number of design coordinates.* TMX 73151, NASA, 1976.

[13] J. Délery, R. Gaillard, G. Losfeld, C. Pendria, *Study of the iso cascade in the ONERA S5Ch wind tunel. Pressure probe and LDV measurements.* ONERA RT 192/1865 DAFE/Y. 1999.

[14] F. Renac, C.-T. Pham, J. Peter, *Sensitivity Analysis for the RANS equations coupled with linearized turbulence model* AIAA Paper 2007-76839. 2007.

[15] L. Cambier, M. Gazaix, *elsA: an efficient object-oriented solution to CFD complexity* AIAA Paper 2002-0108, 2002.

[16] K. Hornik, M. Stinchcombe and H. White *Multilayer Feedforward Networks are Universal Approximators* Neural Networks, Vol. 2, No. 5, pp. 359-366, 1989.

[17] M. Orr *An Introduction to Radial Basis Function Networks* Center for Cognitive Science, University of Edinburg, Scotland, June 1999.