A surrogate optimization approach for inverse problems: Application to turbulent mixed-convection flows

M Oulghelou, Claudine Béghein, Cyrille Allery

A Surrogate Optimization Approach for Inverse Problems : Application to Turbulent Mixed-Convection Flows

M. Oulghelou[1], C. Beghein[2], C. Allery[3]

LaSIE, UMR-7356-CNRS, Université de La Rochelle Pôle Science et Technologie, Avenue Michel Crépeau, 17042 La Rochelle Cedex 1, France.

Abstract

Optimal control of turbulent mixed-convection flows has attracted considerable attention from researchers. Numerical algorithms such as Genetic Algorithms (GAs) are powerful tools that allow to perform global optimization. These algorithms are particularly of great interest in complex optimization problems where cost functionals may lack smoothness and regularity. In turbulent flow optimization, the hybridization of GA with high fidelity Computational Fluid Dynamics (CFD) is extremely demanding in terms of computational time and memory storage. Thus, alternative approaches aiming to alleviate these requirements are of great interest. Nowadays, surrogate approaches gained attention due to their potential in predicting flow solutions based only on preexisting data. In the present paper, we propose a near-real time surrogate genetic algorithm for inverse parameter identification problems involving turbulent flows. In this optimization framework, the parametrized flow data are used in their reduced form obtained by the POD (Proper Orthogonal Decomposition) and solutions prediction is made by interpolating the temporal and the spatial POD subspaces through a recently developed Riemannian barycentric interpolation. The validation of the proposed optimization approach is carried out in the parameter identification problem of the turbulent mixed-convection flow in a cavity. The objective is to determine the inflow temperature corresponding to a given temperature distribution in a restricted area of the spatial domain. The results show that the proposed surrogate optimization framework is

[1]mourad.oulghelou@univ-lr.fr
[2]cbeghein@univ-lr.fr
[3]cyrille.allery@univ-lr.fr

able to deliver good approximations of the optimal solutions within less than two minutes.

Keywords: Flow inverse problem, optimal control, Surrogate optimization, indoor flows, heat problems, Genetic Algorithm, Proper Orthogonal Decomposition.

---

1. Introduction

Decreasing energy consumption of buildings is an important aspect of the reducing of global warming. However, the energy reduction has to be compromised with the quality of thermal comfort inside buildings. To achieve that, optimization applied to indoor airflows, which is aimed at determining optimal flow values for some well chosen parameters are of great interest. The optimization objective can be expressed in the whole or a part of the domain, in terms of field variables such as inlet velocity, wall temperature, heat source, etc. For flows in buildings, which are mostly mixed convection turbulent flows, high fidelity solvers are privileged for parameter identification problems. A usual class of flow optimization algorithms consists in standard gradient descent algorithms using high fidelity adjoint equations. The search direction is computed as the functional cost sensitivity over the design variables and the solution is moved along until an optimal solution is reached. This approach was used for instance by Liu et al. to find optimal thermo-fluid boundary conditions in a two-dimensional cavity [1] and to optimize the air supply location, size, and parameters in a two dimensional non isothermal ventilated cavity [2]. It was also used to optimize buoyancy-driven ventilation flows governed by Boussinesq equations [3, 4]. A major limitation of high fidelity adjoint-based algorithms is that they are more likely to stuck in local optima. To overcome this issue, a global optimization search can be carried out by Genetic Algorithms (GAs) [5]. In the context of mixed-convection flows, high fidelity solvers combined with GA have been investigated and validated in [6, 7]. Compared to high fidelity adjoint based optimization approach, high fidelity based GA is more efficient in terms of finding global optimal solutions, yet it requires a tremendous computing load, leading to turn the attention to techniques of model reduction.

Reduced-order models have been extensively used in fluid dynamics in order to reduce the computational burden in optimization and control applications. Recently, POD (Proper Orthogonal Decomposition) reduced order models based on Galerkin projection were successfully combined with optimization approaches allowing a drastic alleviation of the opti-

mization computational effort. A standard approach used by Tallet et al. [8] and Bergmann et al. [9] consists in using high fidelity simulations to extract a POD basis representing the main structures of a set of snapshots sampled at different parameter values. The temporal dynamics is afterwards calculated by solving an ordinary system of differential equations resulting from Galerkin projection of the governing equations onto the global POD basis. By considering the global POD/Galerkin reduced order model (ROM) as the state equations, a reduced scale optimization problem can be formulated and solved in near-real time. However, in many physical cases, the global POD/Galerkin ROM may experience issues of accuracy due to the overload of information in the global POD basis. Sophisticated subspace interpolation techniques such as the the Grassmann interpolation proposed by Amsallem et al. [10] is an efficient local method meant to restrict the ROM predictions to the wanted physical regime. In the context of the Galerkin/ROM adjoint-based optimal control, the Grassmann interpolation was successfully embedded in a suboptimal control strategy to achieve a near-real time optimal control of transfer phenomena [11].

In the last two decades, interest in surrogate models for flow optimization problems is increasingly growing. Interestingly, the power of these methods is their dispense on the underlying mathematical model. Surrogate Genetic Algorithms are intensively used in the context of airfoil and aerodynamics design. In these cases, in order to decrease the computational time, the fitness is evaluated with metamodels such as radial basis function network [12], feed-forward multi-layer perceptron [12, 13, 14, 15], support vector regression algorithm [16], Gaussian Process Regression [17], the Kriging model [18] or POD (Proper Orthogonal Decomposition) for separating the space and parameters [19], etc. For more details the reader can refer to the review article [18]. It is worth noting that in these applications, the performed studies involve only steady flows, and the aim of the present article is to study unsteady flows. Few attempts have been carried out in this subject, namely by using POD based reduced order models and interpolation in the Grassmann manifold [20] and Artificial

Neural Networks (ANN) [21, 22, 23, 24, 25, 26].

In this paper, we formulate a new surrogate optimization technique based on the Riemannian Barycentric interpolation of subspaces. This interpolation method is based upon the geometry of the manifold of fixed rank matrices studied in detail in [27]. It was initially used to interpolate low-rank solutions of the Luyapunov equations resulting from parametric linear input-output reduced order system [28], and recently adapted to interpolate the parametric Navier-Stokes Galerkin/ROM [29]. In contrast to a similar approach needing a calibration phase for the interpolated POD subspaces [30, 20], the barycentric interpolation naturally results in modes that are arranged according to the POD energetic content. This property allows to interpolate the time and space matrices separately and simply multiply them to form an approximation of the sought untrained solutions. The aim of the following study is to use preexisting time dependent flow solutions to solve the inverse parameter identification problem involving the turbulent mixed-convection flow in a cavity. The optimization objective is to determine the inlet temperature that optimizes the cost functional related to maintaining a desired temperature distribution in the occupied zone of the cavity.

The remainder of this article is organized as follows. First, the studied Mixed convection inverse problem is presented in section 2. In section 3, the barycentric interpolation used for parametrized time dependent solutions prediction is detailed. Next, the surrogate Genetic Algorithm is outlined in section 4. In section 5, numerical experiments assessing the potential of this approach are carried out on the inverse problem involving the turbulent mixed-convection flow in a cavity. Finally, conclusions are drawn in section 6.

2. Mixed convection inverse problem

2.1. Optimization problem settings

This study focuses on the inverse problem of temperature distribution in a two-dimensional ventilated cavity, whose dimensions are $1.04m \times 1.04m$, and which is shown in figure 1. The

temperature θ_{hot} of the bottom wall of the cavity is higher than the temperature θ_{cold} of the other walls:

$$\theta_{hot} = 35.5^o C \text{ and } \theta_{cold} = 15^o C \tag{1}$$

The air inlet (resp. outlet) is located at the top left (resp. bottom right) corner of the

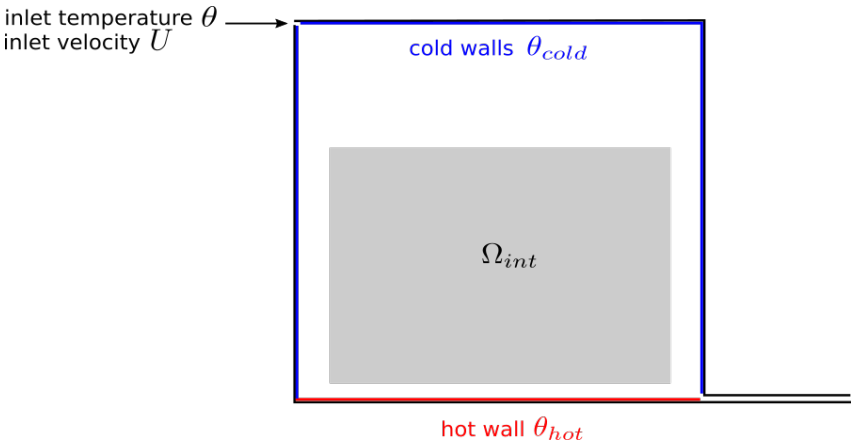


Figure 1: Description of the studied mixed-convection flow

cavity. The vertical dimension of the air inlet (resp. outlet) is 0.018 m (resp. 0.024 m). The inlet air temperature θ is meant to be variable and the inlet velocity U is set to a constant. The turbulent air flow in the cavity is governed by the equations of mass conservation, momentum conservation, energy conservation, of an incompressible Newtonian fluid with Boussinesq's assumption

$$\begin{cases} \nabla \cdot \boldsymbol{v} = 0 \\ \rho \partial_t \boldsymbol{v} + \rho \boldsymbol{v} \cdot \nabla \boldsymbol{v} = -\nabla p + \mu \Delta \boldsymbol{v} + \rho\, \mathbf{g}\, \beta(\Theta - \Theta_0) \boldsymbol{e}_y + \nabla \boldsymbol{\sigma_t} \\ \rho\, c_p\, \partial_t \Theta + \rho\, c_p\, \boldsymbol{v} \cdot \nabla \Theta = \lambda\, \Delta \Theta + \nabla \mathbf{q_t} \end{cases} \tag{2}$$

where $\boldsymbol{v}$, Θ, p are the time averaged velocity[4], temperature and pressure obtained with an Unsteady Reynolds Averaged Navier-Stokes (URANS) turbulence model. ρ, μ, C_p, λ are the density, dynamic viscosity, heat capacity and heat conductivity of the fluid at the reference

[4]In these equations, $\boldsymbol{v}$, Θ and p should have been written $\overline{\boldsymbol{v}}$, $\overline{\Theta}$ and $\overline{p}$. To alleviate the notations in the reminder of the paper, the time averaged notation will not be used.

temperature Θ_0, $\mathbf{g}$ is the gravitational acceleration, β is the thermal expansion coefficient. $\boldsymbol{\sigma_t}$ and $\mathbf{q_t}$ are the turbulent Reynolds stress and the turbulent heat flux given by

$$\boldsymbol{\sigma_{t}}_{ij} = -\rho\,\overline{\boldsymbol{v}_i'\boldsymbol{v}_j'} \qquad \mathbf{q_{t_i}} = -\rho\,c_p\,\overline{\boldsymbol{v}_i'\Theta'}$$

where $\overline{\boldsymbol{v}'}$ and $\overline{\Theta'}$ stand for the temporal mean values of the fluctuating velocity and temperature. The aim of the following study is to solve the constrained nonlinear optimization problem

$$\min_{\theta} \;\; \mathcal{J}\,(y) \qquad \text{subject to} \qquad \mathcal{N}\,(y,\theta) = 0 \tag{3}$$

where $\mathcal{J}$ is the functional describing the cost to minimize, $\mathcal{N}$ the non-isothermal Navier-Stokes equations (2) and $y(\theta)$ the state variable which might be represented for example by the velocity field $\boldsymbol{v}$ or the temperature Θ. In the present article, since the turbulent mixed-convection flow is strongly influenced by the inlet temperature θ, we use this quantity as optimization variable. For a given temperature distribution $\hat{\Theta}$, the goal is to recover the inlet temperature θ that minimizes the objective functional

$$\mathcal{J}(\Theta) = \int_0^{t_f}\int_{\Omega_{int}} (\Theta - \hat{\Theta})^2\,dx\,dt \tag{4}$$

where $[0, t_f]$ is the time frame of simulation and Ω_{int} the restricted occupied zone of the spatial domain depicted in figure 1. It is worth mentioning that one could also think about optimizing on different parameters, such as inlet air velocity and the coordinates or the intensity of a heat source in the domain Ω. Since a GA strategy is to be used, these parameters can easily be incorporated into the cost functional without inducing any modification in the optimization process.

2.2. Standard GA approach

The general idea of GA is illustrated in the flowchart 2. GA consists in starting from a randomly generated set (of size N_{chrom}) of chromosomes $\theta_1, \theta_2, \ldots, \theta_{N_{\text{chrom}}}$, forming a population. The size of populations is unchanged and fixed to N_{chrom} along the generations. In each population, a fitness value [31] is assigned to each chromosome θ_j. Virtually, any

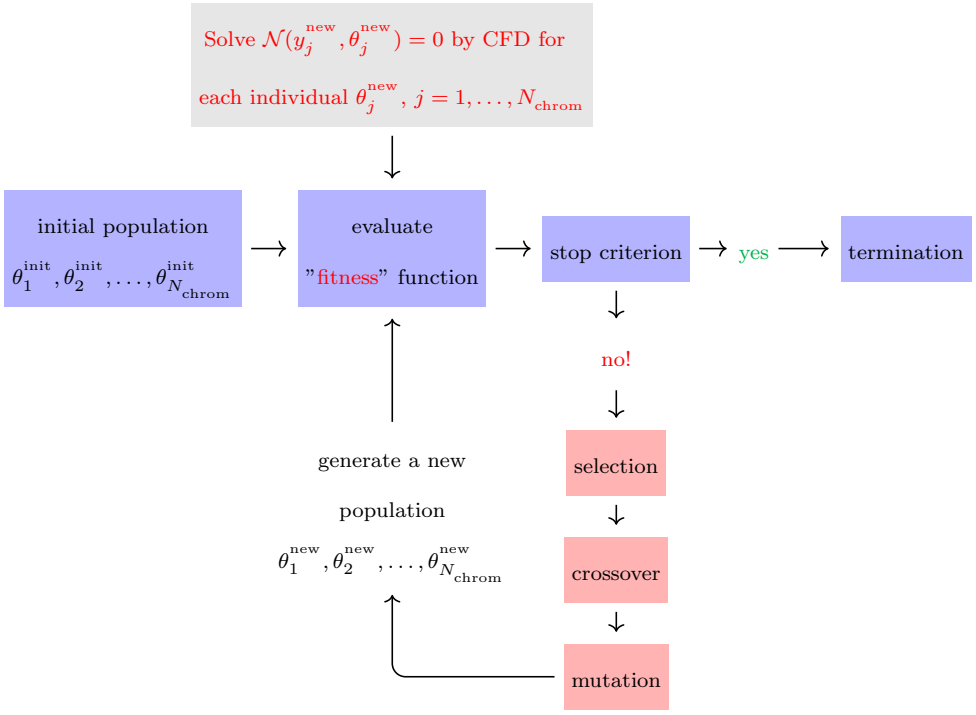


Figure 2: Outline of the generic Genetic Algorithm.

fitness function can be chosen given that no requirement for continuity in the derivatives is needed. Some examples of the choice of fitness functions can be found in [32, 33, 34, 35]. In the present paper, the fitness function f is chosen as the inverse of the objective function, i.e, the fitness of the j^{th} chromosome is calculated as follows

$$f(y_j) = \frac{1}{\mathcal{J}(y_j, \theta_j)}$$

where y_j is obtained by solving the constraint problem $\mathcal{N}(y_j, \theta_j) = 0$. In order to evolve populations, three main genetic operators [36], modeled on the Darwinian concepts of natural selection and evolution are used. These are :

Selection. There are several methods for selecting the best chromosomes and their transfer to the next generation. In general, a new population of chromosomes is chosen to survive based on their fitness values. That means that a chromosome θ_j with a large fitness value has higher probability of being reproduced and passed down into the next generation. The

probability of reproduction can be calculated as follows

$$P_s^j = f(y_j) \Big/ \sum_{i=1}^{N_{\text{chrom}}} f(y_i)$$

Using this reproduction probability, N_{chrom} solutions from the current generation are selected by the roulette rule [37] to survive for the next generation. These reproduced solutions are afterwards modulated by the crossover and mutation operators [36] .

Crossover. The crossover is the operation wherein genes are exchanged between two chromosomes. In particular, all the surviving chromosomes by the roulette selection rule are randomly paired. Precisely, two individuals are randomly selected as parent individuals, then arbitrary positions on both individuals are chosen for crossing locations where exchange of genes takes place. In practice, a random number ranging from 0 to 1 is generated. If the random number is greater than a probability P_c, the two chromosomes in the original pair remain into the next generation. Otherwise, the crossover takes place, and two new chromosomes are created to replace the parent chromosomes.

Mutation. The mutation operator is responsible for bringing new information to the population. With a probability P_m ranging from 0 to 1, the mutation operator accidentally changes one of the resulted genes.

The above genetic operations are repeated for a predetermined number of generations arbitrarily set by the user. The best chromosome of the final generation is declared as the global optimized solution. Despite their superiority with respect to other optimization approaches, a serious weakness of high fidelity based GAs is their considerable requirements in computational effort and memory storage [38]. In fact, GA needs to perform high fidelity simulations many times for each evolved population (iteration). With the increase in the number of generations, the populations and their required crossovers and mutations will increase. These, in turn increase the time complexity of GA, making unfeasible their application in near-real time. In order to tackle this issue, an interpolation strategy suited

for non-linear parameterized data and intended to replace the high fidelity solver in the GA is proposed in the next section.

3. Barycentric interpolation for nonlinear parametrized data

3.1. Data compression strategy by using the POD

Consider a set of parametrized matrices $\{\boldsymbol{\mathcal{Y}}_k \in \mathbb{R}^{N_x \times N_s}, k = 1, \ldots, N_p\}$ formed from the discrete solutions[5] $y(\theta_k)$ of a transient non-linear flow problem. i.e,

$$\boldsymbol{\mathcal{Y}}_k = \begin{bmatrix} y(t_1, x_1, \theta_k) & y(t_2, x_1, \theta_k) & \\ \vdots & \ddots & \\ y(t_1, x_{N_x}, \theta_k) & & y(t_{N_s}, x_{N_x}, \theta_k) \end{bmatrix}$$

In practice, θ refers to a parameter of the flow problem, N_x the number of spatial degrees of freedom and N_s the number of time steps, where it is assumed that N_x exceeds N_s by several orders of magnitude. The aim of the following is to extract a set of reduced matrices $\mathcal{Y}_k$ that describes the dynamics of the full order matrices $\boldsymbol{\mathcal{Y}}_k$. To this end, assume that each matrix $\boldsymbol{\mathcal{Y}}_k$ is approximated in a POD basis [6] of dimension q as follows

$$\boldsymbol{\mathcal{Y}}_k \approx \Phi_k {\Lambda_k}^T \tag{5}$$

where $\Phi_i \in \mathbb{R}^{N_x \times q}$ and $\Lambda_i \in \mathbb{R}^{N_s \times q}$ are respectively the spatial and temporal bases. Now, consider the POD respectively of orders r and s, $r, s \leq qN_p$, of the column block matrices

$$\begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{N_p} \end{bmatrix} = \boldsymbol{\Phi}\boldsymbol{\varphi}^T \quad \text{and} \quad \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_{N_p} \end{bmatrix} = \boldsymbol{\Lambda}\boldsymbol{\alpha}^T$$

[5] In our case, the solutions y correspond to the turbulent mixed convection temperature distribution Θ and θ the inlet temperature.

[6] The POD bases are constructed such that they verify optimality with respect to the Euclidean inner product. In this case, the POD is nothing but the Singular Value Decomposition (SVD). However, other inner products such as L^2 or H^1 can be used. More details about the POD approach can be found in [39].

where $\boldsymbol{\Phi} \in \mathbb{R}^{N_x \times r}$, $\boldsymbol{\varphi} \in \mathbb{R}^{qN_p \times r}$, $\boldsymbol{\Lambda} \in \mathbb{R}^{N_s \times s}$, $\boldsymbol{\alpha} \in \mathbb{R}^{qN_p \times s}$. Let $\varphi_i \in \mathbb{R}^{q \times r}$ and $\alpha_i \in \mathbb{R}^{q \times s}$, $i = 1, \ldots, N_p$, be the column block matrices of $\boldsymbol{\varphi}^T$ and $\boldsymbol{\alpha}^T$ such as

$$\boldsymbol{\varphi}^T = \begin{bmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_{N_p} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\alpha}^T = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{N_p} \end{bmatrix}$$

It yields that the full order snapshots matrix associated to the parameter θ_k can be written as

$$\boldsymbol{\mathcal{Y}}_k \approx \boldsymbol{\Phi} \varphi_k \alpha_k^T \boldsymbol{\Lambda}^T \tag{6}$$

It is important to note that in the above expression is very economic in terms of cost and memory load. In fact, the global matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Lambda}$ are calculated and stored for once. Only the nested matrices $\mathcal{Y}_k = \varphi_k \alpha_k^T$ that change with respect to parameter variation. In parametric applications such as parameter identification problems, rather than using the full order matrices $\boldsymbol{\mathcal{Y}}_k$, it is more convenient to manipulate the corresponding nested reduced matrices $\mathcal{Y}_k$ of significantly reduced size $r \times s$ in order to achieve low cost calculations. The interpolation strategy of the matrices $\mathcal{Y}_k$ is detailed in the next subsection.

3.2. Data interpolation

In the following, the interpolation approach is first presented for two data samples. The generalization to an arbitrary number of data samples is given afterwards. Let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be two parametrized compressed matrices associated respectively to θ_1 and θ_2, such that

$$\mathcal{Y}_1 = \varphi_1 \alpha_1^T \qquad \mathcal{Y}_2 = \varphi_2 \alpha_2^T$$

where φ_k and α_k are rank-q parameterized matrices resulted from the data compression procedure. By using the above representations, the goal is to predict the matrix $\tilde{\mathcal{Y}}$ associated to a new parameter value $\tilde{\theta}$ different from θ_1 and θ_2. To this end, the barycentric interpolation proposed in [29] for subspaces interpolation is used. For the sake of simplicity, we restrict ourselves to univariate Lagrange functions to generate interpolation weights. In the general, other interpolation functions such as Radial Basis Functions (RBF), Kriging,

etc, can be used for univariate or multivariate interpolation parameters. The Lagrange functions constructed by using two points θ_1 and θ_2 are given by

$$\omega_1(\tilde{\theta}) = \frac{\tilde{\theta} - \theta_2}{\theta_1 - \theta_2} \qquad \omega_2(\tilde{\theta}) = \frac{\tilde{\theta} - \theta_1}{\theta_2 - \theta_1}$$

During the interpolation process, two sorts of subspaces have to be distinguished. The spatial subspaces $span(\varphi_1)$ and $span(\varphi_2)$, and the temporal subspaces $span(\alpha_1)$ and $span(\alpha_2)$. The proposed data interpolation technique suggests to predict the new matrix $\tilde{\mathcal{Y}}$ by applying the barycentric interpolation strategy to the spatial and temporal subspaces separately, i.e, it consists in solving the fixed point problems

$$(\mathcal{P}_x) \quad \begin{cases} \text{Find } \tilde{\varphi} \text{ such that :} \\ \tilde{\varphi}^T \varphi_1 \overset{\text{SVD}}{=} \xi_1 \Sigma_1 \eta_1^T \quad \text{and} \quad \tilde{\varphi}^T \varphi_2 \overset{\text{SVD}}{=} \xi_2 \Sigma_2 \eta_2^T \\ \tilde{\varphi} = \dfrac{\tilde{\theta} - \theta_2}{\theta_1 - \theta_2} \varphi_1 \tilde{Q}_1 + \dfrac{\tilde{\theta} - \theta_1}{\theta_2 - \theta_1} \varphi_2 \tilde{Q}_2 \quad \text{where} \quad \tilde{Q}_1 = \eta_1 \xi_1^T \quad \text{and} \quad \tilde{Q}_2 = \eta_2 \xi_2^T \end{cases}$$

$$(\mathcal{P}_t) \quad \begin{cases} \text{Find } \tilde{\alpha} \text{ such that :} \\ \tilde{\alpha}^T \alpha_1 \overset{\text{SVD}}{=} \zeta_1 \Upsilon_1 \tau_1^T \quad \text{and} \quad \tilde{\alpha}^T \alpha_2 \overset{\text{SVD}}{=} \zeta_2 \Upsilon_2 \tau_2^T \\ \tilde{\alpha} = \dfrac{\tilde{\theta} - \theta_2}{\theta_1 - \theta_2} \alpha_1 \tilde{K}_1 + \dfrac{\tilde{\theta} - \theta_1}{\theta_2 - \theta_1} \alpha_2 \tilde{K}_2 \quad \text{where} \quad \tilde{K}_1 = \tau_1 \zeta_1^T \quad \text{and} \quad \tilde{K}_2 = \tau_2 \zeta_2^T \end{cases}$$

where the notation $\overset{\text{SVD}}{=}$ stands for the Singular Value Decomposition of the left hand side. The iterative process to solve the problem $(\mathcal{P}_x)$ is described by the following fixed point sequence

$$(\mathcal{P}_x) \quad \begin{cases} \tilde{\varphi}^{(0)} \text{ given, for } n \geq 0 \\ \text{Perform the SVD of } \tilde{\varphi}^{(n)^T} \varphi_1 \overset{\text{SVD}}{=} \xi_1^{(n)} \Sigma_1^{(n)} \eta_1^{(n)^T} \quad \text{then set} \quad \tilde{Q}_1^{(n)} = \eta_1^{(n)} \xi_1^{(n)^T} \\ \text{Perform the SVD of } \tilde{\varphi}^{(n)^T} \varphi_2 \overset{\text{SVD}}{=} \xi_2^{(n)} \Sigma_2^{(n)} \eta_2^{(n)^T} \quad \text{then set} \quad \tilde{Q}_2^{(n)} = \eta_2^{(n)} \xi_2^{(n)^T} \\ \text{Update the interpolant as } \tilde{\varphi}^{(n+1)} = \dfrac{\tilde{\theta} - \theta_2}{\theta_1 - \theta_2} \varphi_1 \tilde{Q}_1^{(n)} + \dfrac{\tilde{\theta} - \theta_1}{\theta_2 - \theta_1} \varphi_2 \tilde{Q}_2^{(n)} \end{cases}$$

The same strategy applies for the resolution of problem $(\mathcal{P}_t)$. Now, once the solutions $\tilde{\varphi}$ and $\tilde{\alpha}$ respectively, of the fixed points problems $(\mathcal{P}_x)$ and $(\mathcal{P}_t)$ are found, the reduced snapshot

matrix $\tilde{\mathcal{Y}}$ can be formed as

$$\tilde{\mathcal{Y}} = \tilde{\varphi}\,\tilde{\alpha}^T = \left(\frac{\tilde{\theta}-\theta_2}{\theta_1-\theta_2}\right)^2 \varphi_1\tilde{Q}_1\tilde{K}_1^T\alpha_1^T + \left(\frac{\tilde{\theta}-\theta_1}{\theta_2-\theta_1}\right)^2 \varphi_2\tilde{Q}_2\tilde{K}_2^T\alpha_2^T + \frac{(\tilde{\theta}-\theta_1)(\theta_2-\tilde{\theta})}{(\theta_1-\theta_2)^2}\left(\varphi_1\tilde{Q}_1\tilde{K}_2^T\alpha_2^T + \varphi_2\tilde{Q}_2\tilde{K}_1^T\alpha_1^T\right)$$

A very interesting property of the above formula is that even though space and time reduced bases $\{\varphi_1, \varphi_2\}$ and $\{\alpha_1, \alpha_2\}$ are separately interpolated, the calibration between the columns of $\tilde{\varphi}$ and $\tilde{\alpha}$ is naturally ensured by the barycentric interpolation, unlike the method in [20] where the calibration is lost by the Grassmannian interpolation.

Let us now state the general framework of the data interpolation approach. To do so, consider a set of parametrized data matrices $\mathcal{Y}_1, \cdots, \mathcal{Y}_{N_p}$ associated to the parameter values $\theta_1, \theta_2, \ldots, \theta_{N_p}$, such that

$$\mathcal{Y}_k = \varphi_k\alpha_k^T, \quad k = 1\ldots, N_p$$

The approximate matrix $\tilde{\mathcal{Y}}$ for a new untrained value $\tilde{\theta} \neq \theta_k$ obtained by solving the following fixed point problems

$$(\mathcal{P}_x) \quad \begin{cases} \text{Find } \tilde{\varphi} \text{ such that :} \\ \tilde{\varphi}^T\varphi_k \overset{\text{SVD}}{=} \xi_k\Sigma_k\eta_k^T, \quad k = 1, \ldots, N_p \\ \tilde{\varphi} = \displaystyle\sum_{k=1}^{N_p} \omega_k(\tilde{\theta})\,\varphi_k\tilde{Q}_k \quad \text{where} \quad \tilde{Q}_k = \eta_k\xi_k^T \end{cases} \qquad (\mathcal{P}_t) \quad \begin{cases} \text{Find } \tilde{\alpha} \text{ such that :} \\ \tilde{\alpha}^T\alpha_k \overset{\text{SVD}}{=} \zeta_k\Upsilon_k\tau_k^T \quad k = 1, \ldots, N_p \\ \tilde{\alpha} = \displaystyle\sum_{k=1}^{N_p} \kappa_k(\tilde{\theta})\,\alpha_k\,\tilde{K}_k \quad \text{where} \quad \tilde{K}_k = \tau_k\zeta_k^T \end{cases}$$

The solution is then constructed as follows

$$\tilde{\mathcal{Y}} = \sum_{k,h=1}^{N_p} \omega_k(\tilde{\theta})\,\kappa_h(\tilde{\theta})\,\varphi_k\tilde{Q}_k\tilde{K}_h^T\alpha_h^T$$

where $\tilde{Q}_k$ and $\tilde{K}_k$ are orthogonal matrices and ω_k and κ_k are some interpolation functions of sum equal to 1, verifying $\omega_k(\theta_i) = \kappa_k(\theta_i) = \boldsymbol{\delta^{ki}}$, with $\boldsymbol{\delta^{ki}}$ the piecewise Kronecker delta function which value is 1 if k equals i and 0 otherwise. The interpolation procedure of nonlinear parametrized data is summarized in algorithm 1.

Algorithm 1: Non-linear data interpolation strategy

Offline :

Use the POD to compress the trained parametrized data matrices $\boldsymbol{\mathcal{Y}}_k$ such that

$$\boldsymbol{\mathcal{Y}}_k \approx \boldsymbol{\Phi}\mathcal{Y}_k\boldsymbol{\Lambda}^T \qquad \text{where} \quad \mathcal{Y}_k = \varphi_k\alpha_k^T$$

Online :

Give a value of $\tilde{\theta}$ (chosen by the user) and calculate the weights $\omega_k(\tilde{\theta})$ and $\kappa_h(\tilde{\theta})$

Set $\tilde{\mathcal{Y}}^{(0)} = \tilde{\varphi}^{(0)}\tilde{\alpha}^{(0)^T}$ arbitrary, for example choose a point $\mathcal{Y}_k$ from the sampling

while $Error > \varepsilon$ **do**

  for $k \in \{1, \ldots, N_p\}$ **do**

    Calculate the matrix $\tilde{Q}_k^{(n)} = \eta_k^{(n)}\xi_k^{(n)^T}$ where $\tilde{\varphi}^{(n)^T}\varphi_k \overset{\text{SVD}}{=} \xi_k^{(n)}\Sigma_k^{(n)}\eta_k^{(n)^T}$

    Calculate the matrix $\tilde{K}_k^{(n)} = \tau_k^{(n)}\zeta_k^{(n)^T}$ where $\tilde{\alpha}^{(n)^T}\alpha_k \overset{\text{SVD}}{=} \zeta_k^{(n)}\Upsilon_k^{(n)}\tau_k^{(n)^T}$

  Update : $\tilde{\varphi}^{(n+1)} = \sum_{k=1}^{N_p}\omega_k(\tilde{\theta})\,\varphi_k\,\tilde{Q}_k^{(n)}, \qquad \tilde{\alpha}^{(n+1)} = \sum_{h=1}^{N_p}\kappa_h(\tilde{\theta})\,\alpha_h\,\tilde{K}_h^{(n)}$

  Evaluate the error [a] : $Error = \sum_{k=1}^{N_p}\sum_{h=1}^{N_p}||\tilde{Q}_k^{(n)}\tilde{K}_h^{(n)^T} - \tilde{Q}_k^{(n-1)}\tilde{K}_h^{(n-1)^T}||_F.$

  $n \longleftarrow n+1$

Form the interpolated matrix : $\tilde{\mathcal{Y}} = \tilde{\varphi}^{(n+1)}\,\tilde{\alpha}^{(n+1)}$

[a] The error is evaluated by the the Frobenius norm denoted $||\cdot||_F$.

In terms of computational complexity, the operations in Algorithm 1 involve matrix-matrix products and Singular Value Decompositions of matrices of size $N_p q \times q$ where N_p (number of training points) and q (POD order of truncation) are in general of order of dozens or hundreds. Thus, the total computational cost for a single prediction by Algorithm 1 is proportional to $\mathcal{O}(N_p q^3)$, which makes it a very computationally efficient procedure. As a result, and in order to tackle the severe computational effort of Genetic algorithms, a surrogate optimization procedure based on the barycentric interpolation is proposed in the next section. In this procedure, Algorithm 1 is used as solution predictor instead of the Navier-Stokes turbulent high fidelity solver.

4. Surrogate Genetic Algorithm

Basically, the proposed surrogate GA is an optimization strategy to solve inverse problems by means of available precomputed parametrized flow data. The major advantage of this approach is that the relationship between the state variable y and the optimization variable θ, earlier established through the mapping $\mathcal{N}$, is now replaced by the cheap explicit formula of the barycentric interpolation

$$y(t_l, x_j, \tilde{\theta}) \approx \mathbf{\Phi}(x_j)\tilde{\mathcal{Y}}\mathbf{\Lambda}^T(t_l) \tag{7}$$

where $\mathbf{\Phi}(x_j)$ and $\mathbf{\Lambda}(t_l)$ denote respectively the j^{th} and l^{th} rows of the matrices $\mathbf{\Phi}$ and $\mathbf{\Lambda}$, and $\tilde{\mathcal{Y}}$ the reduced snapshots matrix to be found by algorithm 1.

In order to make sure that the surrogate GA performs in an optimal manner, the chromosomes are enriched by virtual genes. These genes are the order of POD truncation q and the number of spatial and temporal interpolation neighbors denoted respectively ne_x and ne_t. To illustrate this, let $\mathcal{Y}_1, \ldots, \mathcal{Y}_4$ be four reduced matrices associated to the parameter values $\theta_1 < \theta_2 < \theta_3 < \theta_4$ respectively such that

$$\mathcal{Y}_k = \varphi_k \alpha_k^T, \quad k = 1, \ldots, 4$$

where φ_k and α_k are rank-q matrices. Suppose that we want to find an approximation of the reduced matrix $\tilde{\mathcal{Y}}$ for an untrained value $\tilde{\theta} \in]\theta_1, \theta_2[$ by using an order of POD truncation $m < q$, three neighbors for spatial interpolation ($ne_x = 3$) and two neighbors for temporal interpolation ($ne_t = 2$). Then the untrained reduced matrix is approximated as

$$\tilde{\mathcal{Y}} = \sum_{k=1}^{3}\sum_{h=1}^{2} \omega_k(\tilde{\theta})\, \kappa_h(\tilde{\theta})\, \varphi_k \tilde{Q}_k \tilde{K}_h^T \beta_h^T$$

where the columns of φ_k and β_h are truncated up to the order m. If Lagrange functions are used, the weights write

$$\omega_k(\tilde{\theta}) = \prod_{\substack{i=1 \\ i \neq k}}^{3} \frac{\tilde{\theta} - \theta_i}{\theta_k - \theta_i} \qquad \text{and} \qquad \kappa_h(\tilde{\theta}) = \prod_{\substack{i=1 \\ i \neq h}}^{2} \frac{\tilde{\theta} - \theta_i}{\theta_h - \theta_i}$$

In the proposed genetic algorithm strategy, the j^{th} chromosome is then the candidate $\bar{\theta}_j = \{\theta_j, ne_t, ne_x, q\}$ where θ_j, ne_t, ne_x and q are its genes. Accordingly, the original optimization problem (3) is modified yielding to

$$\min_{\bar{\theta}} \ \mathcal{J}(\tilde{\mathcal{Y}}, \bar{\theta}) \quad \text{such that } \tilde{\mathcal{Y}} \text{ is the output of algorithm 1}$$

In the next section, the potential of this approach is assessed on the inverse parameter identification problem involving a turbulent mixed convection flow.

5. Numerical experiments

5.1. Validation of the high fidelity computations

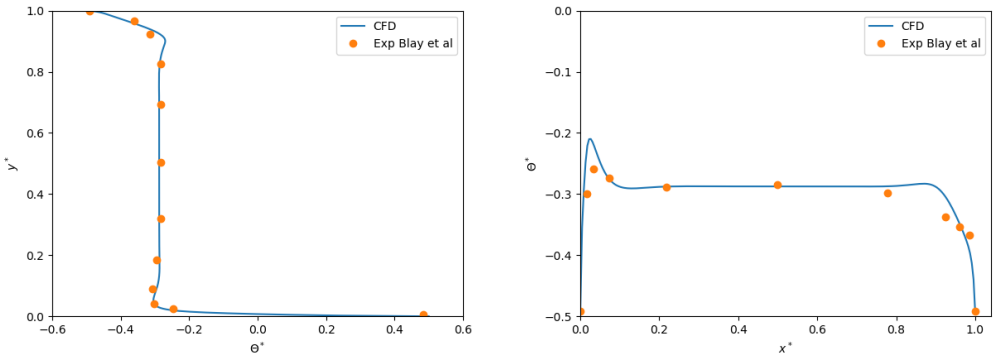
In this section, the CFD model used to solve the mixed-convection problem is first validated.This validation is performed by comparing the numerical results to the experimental results obtained by Blay et al. [40], where a turbulent mixed convection flow was generated in a ventilated cavity with dimensions $1.04 \times 1.04 \times 0.7\ m^3$. In this experiment, a two-dimensional flow was generated in the enclosure shown in figure 1, which was surrounded by two guard cavities. The reference temperature Θ_0 was the average temperature in the cavity. The Rayleigh number of this configuration, based on the cavity height and on the temperature difference between the heated floor ($\theta_{hot} = 35.5^oC$) and the other walls and the inlet ($\theta_{cold} = \theta = 15^oC$), was 2.13×10^9. The Reynolds number based on the air velocity at inlet $U = 0.57m/s$ and on the inlet height was 654.

To compute this flow, and to generate all input data necessary for the study presented in this paper, the finite volume code OpenFOAM [41] was used. The computational domain was discretized into a non uniform grid made of 26680 hexaedral cells, which was very tight close to the walls, in order to properly discretized the boundary layer. The two-dimensional turbulent flow was modeled with the RNG k-epsilon model [42]. The non-isothermal flow described by equations (2) was calculated with the buoyantBoussinesqPimpleFoam solver. At the inlet, the velocity boundary conditions were $u = 0.57$ m/s and $v = 0$ m/s, and the turbulent boundary conditions were $k = 1.25 \times 10^{-3} m^2/s^2$ and $\epsilon = 5.76 \times 10^{-3} m^2/s^3$. In

what follows, the inlet temperature will be varied between 2^oC and 27^oC. On the walls, no-slip boundary conditions were applied for the velocity components, the temperature was equal to 35.5^oC on the floor, and to 15^oC on the other walls. At the outlet, zero gradient boundary conditions were applied for the temperature, the velocity components and the turbulent variables. The convection terms were discretized with the Gauss linear Upwind scheme, and the laplacian terms were approximated with the Gauss linear corrected scheme.

At $t = 0$ s, the temperature in the cavity is equal to θ_{cold}, and the velocity to 0. The final time instant t_f was chosen in such a way that the temporal evolution of the temperature in the center of the cavity did not vary according to time. For all simulations, $t_f = 1250$ s was a sufficiently a long time interval to achieve the established regime of the flow.

The CFD computations were validated with the experiments carried out by Blay et al. for an inlet temperature equal to 15^oC. With the used non uniform mesh, the average y^+ value was equal to 0.55, and the maximum value was 3.09. In figure 3, the temperature profiles for the inlet temperature $\theta = 15^oC$, at $x = 0.52$ and at $y = 0.52$ are shown such that, $\Theta^* = \frac{\Theta - \Theta_0}{\theta_{hot} - \theta_{cold}}$, $x^* = x/H$ and $y^* = y/H$ where H is the cavity height. A satisfactory agreement can be noticed.



(a) Θ^* at $x^* = 0.5$ (b) Θ^* at $x^* = 0.5$

Figure 3: Comparison between the numerical results (CFD) and the experimental results (Exp Blay et al.) of Θ^* at $x^* = 0.5$ (a) and $y^* = 0.5$ (b) at the established regime of the flow.

5.2. Dynamics of the mixed convection flow for variable inlet temperatures

The isovalues of temperature obtained for three inlet temperatures 2^oC, 14^oC and 26^oC are represented in Figure 4. It can be seen that these flows represent three different complex regimes. For the small inlet temperature $\theta = 2^oC$, the air in the upper left part of the cavity, which is too cold, falls along the left wall, warmed by the hot floor, and finally lifted by natural convection with a counterclockwise motion along the hot floor. For the higher inlet temperature $\theta = 14^oC$, the air in the upper part of the cavity is warm and the clockwise motion of a large recirculation region induced by the combined effects of the forced convection phenomenon and the natural convection phenomenon along the hot floor can be seen. For the highest inlet temperature $\theta = 26^oC$, the injected air is hot and remains in a large region along the ceiling, it falls afterwards along the left and right cold walls, and lift up along the heated floor, inducing two recirculation regions, a clockwise one in the right part of the cavity, and a counterclockwise one in the left part of the cavity. In the following, the surrogate GA will be applied for this complex unsteady flow involving different regimes.

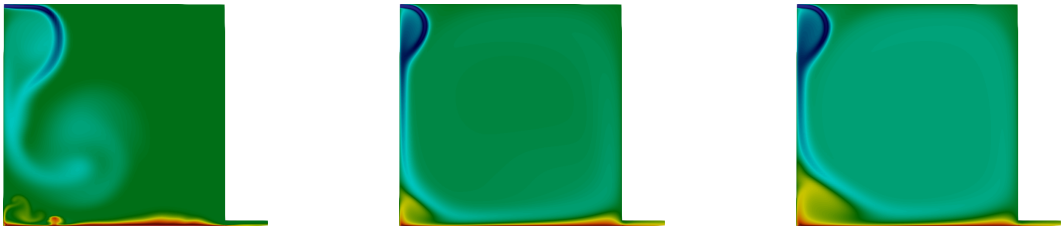
5.3. Surrogate optimization problem

In what follows, the earlier proposed surrogate GA approach is used in order to tackle the severe computational effort due to high fidelity simulations. Thereby, the time required for evaluating the fitness of one chromosome passes from several hours[7] to real time , and thus, drastically reducing the time needed for optimization. The goal of the following experiment is to act on the inlet temperature θ in order to minimize the discrete cost functional

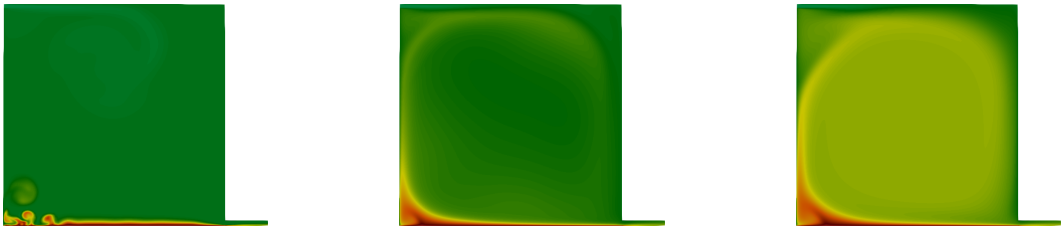
$$\mathcal{J}(\Theta) = \frac{1}{N_s}\sum_{n=1}^{N_s}\int_{\Omega_{int}} (\Theta^n - \hat{\Theta}^n)^2\, dx \tag{8}$$

where the interior subdomain represented in figure 1 is considered such that $\Omega_{int} = [0.1, 0.9]\times[0.15, 0.7]$. The superscript n refers to the time instant, Θ^n the calculated temperature and

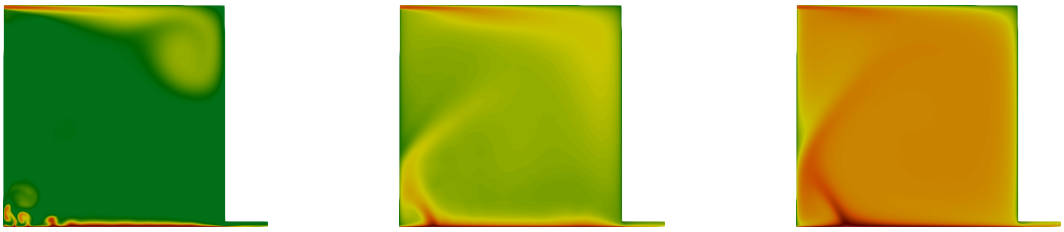
[7] In our case, the full parallel simulation on 15 processors of the mixed convection flow problem on a mesh of 26680 cells, needed about 6 hours.



(a) Temperature distributions associated to the inlet temperature $\theta = 2^oC$



(b) Temperature distributions associated to the inlet temperature $\theta = 14^oC$



(c) Temperature distributions associated to the inlet temperature $\theta = 26^oC$

Figure 4: Illustration of the rich dynamics of the training sampling solutions. The temperature distributions are represented in at three time instants : $t = 8.75s$ (left), $t = 55s$ (middle) and $t = 1250s$ (right)

$\hat{\Theta}^n$ the target temperature.

The training parametrized solution data are calculated by OpenFOAM for different values of inlet temperature θ over the time interval $[0, t_f]$. The corresponding POD decompositions are constructed by using 1000 snapshots uniformly spaced in the time interval $[0, t_f]$ and the maximal POD truncation order is set to 50.

Let $\hat{\Theta}(\theta)$ be a CFD solution for the inlet temperature θ at the time instant t_n and $\widetilde{\Theta}^n(\theta)$ its prediction by Algorithm 1. The percentage of error between the CFD and predicted

temperatures is given by the expression

$$\epsilon_{\text{mean}}(\theta) = 100 \frac{\sqrt{\frac{1}{N_s}\sum_{n=1}^{N_s}\int_{\Omega}\left(\widetilde{\Theta}^n(\theta) - \hat{\Theta}^n(\theta)\right)^2 dx}}{\sqrt{\frac{1}{N_s}\sum_{n=1}^{N_s}\int_{\Omega}\hat{\Theta}^n(\theta)^2 dx}} \tag{9}$$

Before applying the new surrogate GA to the turbulent mixed convection flow problem, a study to assess the accuracy and the sensibility of the barycentric interpolation with respect to the considered number of samples (i.e distance between parameters $\Delta\theta$), number of interpolation neighbors ne_t and ne_x for temporal and spatial POD bases respectively, and the order of the POD truncation q is performed in the next subsection.

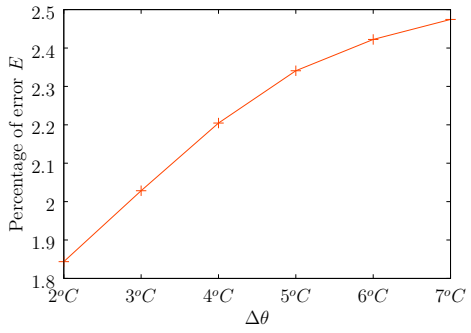
5.4. Sensibility of the barycentric interpolation with respect to $\Delta\theta$, ne_x, ne_t and q

A thorough analysis of the barycentric interpolation is performed herein for the mixed-convection flow problem when the inlet temperature θ varies. To this end, we consider the ensemble of snapshots data generated by considering the inlet temperature values $2^oC, 3^oC, 4^oC, \ldots, 27^oC$. For a fixed value of $\Delta\theta$ the training set of parameters is $I_{\text{tr}}(\Delta\theta) = \{2^o + k\Delta\theta\}_{k\geq 0}$ such that the last value does not exceed 27^oC and the set of test parameters are $I_{\text{test}}(\Delta\theta) = \{2^oC, 3^oC, \ldots, 27^oC\} \setminus \{2^oC + k\Delta\theta\}_{k\geq 0}$. For example if $\Delta\theta = 2^oC$, then $I_{\text{tr}}(2^oC) = \{2, 4, \ldots, 26\}$ and the $I_{\text{test}}(2^oC) = \{3^oC, 5^oC, \ldots, 27^oC\}$. For all values $\Delta\theta$ varying in $\{2^oC, 3^oC, \ldots, 8^oC\}$, $ne_t \geq 2$, $ne_x \geq 2$ and $q \in \{1, 2, \ldots, 50\}$, we run all the possible predictions by the barycentric interpolation and evaluate the following error

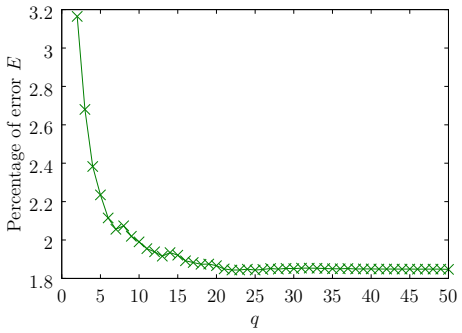
$$E(\Delta\theta, ne_t, ne_x, q) = \max_{k \in I_{\text{test}}(\Delta\theta)} \epsilon_t(\theta_k) \tag{10}$$

$\epsilon_t(\theta_k)$ is the percentage of error (9) where the prediction of the temperature $\widetilde{\Theta}(\theta_k)$, by the barycentric interpolation, is performed on the training POD data (truncated to the order q) corresponding to the points $I_{\text{tr}}(\Delta\theta)$ and by considering ne_t and ne_x neighbors for spatial and temporal subspaces respectively. In Figure 5, we plot the lower hull of the errors in the four directions $\Delta\theta$, q, ne_x and ne_t. We observe that the barycentric interpolation is less sensible to $\Delta\theta$, ne_t and q, but most sensible to ne_x. Overall, an error E nearly less than

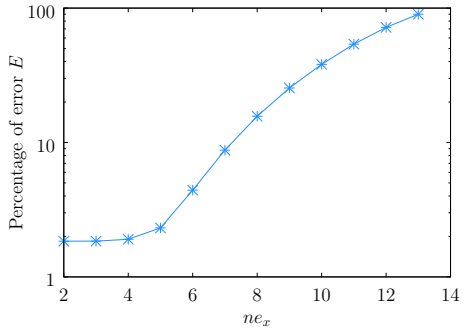
2% can be noticed within the ranges $\Delta\theta = \{2^oC, 3^oC\}$, $q \geq 7$, $ne_x \leq 4$ and $ne_t \leq 6$. In the next subsection, we chose $\Delta\theta = 3^oC$ for the numerical experiment of the surrogate GA, while the parameters ne_x, ne_t and q are allowed to freely vary in the search space.



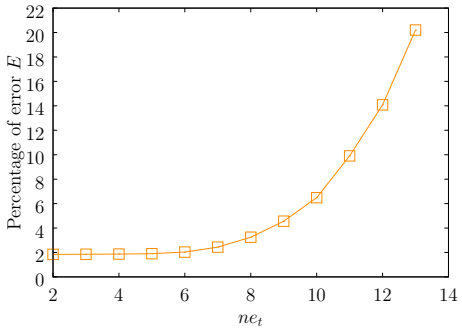
(a) Influence of the step $\Delta\theta$



(b) Influence of truncation q



(c) Influence of the number of neighbors ne_x



(d) Influence of the number of neighbors ne_t

Figure 5: Lower hull of the errors in the four directions $\Delta\theta$, q, ne_x and ne_t.

5.5. Application of the surrogate GA to the turbulent mixed-convection flow problem

The optimization is performed by acting on the inlet temperature θ. The considered training injection temperature values belong to the set $I_{\text{tr}}(3^oC) = \{2^oC, 5^oC, \ldots, 23^oC, 26^oC\}$. Given a temperature distribution $\hat{\Theta}$, the aim is to apply the surrogate GA to approximate the associated optimal inlet temperature. The space of search by surrogate GA is set to

$$K = \left\{(\theta, ne_t, ne_x, m) \in \mathbb{R}_+ \times \mathbb{N}^3, \quad 2^oC \leq \theta \leq 26^oC;\ 2 \leq ne_t, ne_x \leq 13 \ \text{ and } 4 \leq m \leq q\right\}$$

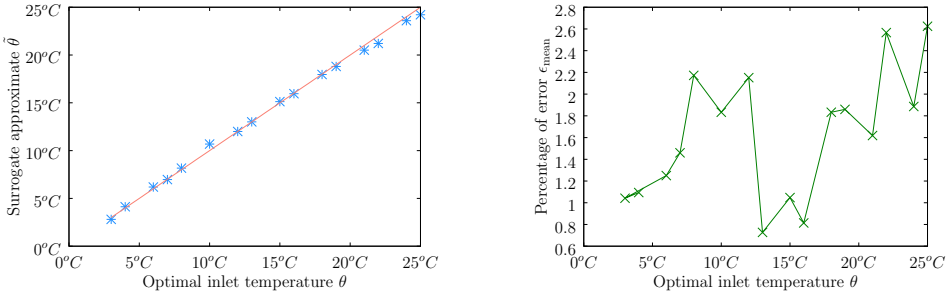
Different tests are performed for temperature distributions associated to 16 inlet values in the set $I_{\text{test}}(3^oC) = \{3^oC, 4^oC, 6^oC, 7^oC, \ldots, 24^oC, 25^oC\}$. In the following numerical experiments of surrogate GA, a population of 20 chromosomes formed by 4 genes randomly generated in K is used as initial guess to run the surrogate GA. The selection probability is set to 60%, the crossover probability to 75% and the mutation probability to 30%. The algorithm is allowed to run until a maximum number of iterations predetermined by the user is reached. The maximum number of iterations here is set to 20.

For each test inlet temperature, the surrogate GA is run until the averaged cost functional stagnates, meaning by that the populations contain a chromosome of high recurrence. The results of these tests in terms of parameter approximation and percentage of error defined by equation (9) are plotted in Figure 6. It can be confirmed that the surrogate GA succeeds to recover a good approximation $\tilde{\theta}$ of the sought inlet temperature $\hat{\theta}$ with a good accuracy of the reconstructed temperature distribution of less than 3% of error. Moreover for all the optimization tests, the optimal values of q, ne_x and ne_t are within the ranges $q \geq 7$, $ne_x \leq 4$ and $ne_t \leq 6$. This matches the ranges of the previous sensibility study. In the following, we focus on the surrogate GA predictions for three inlet temperatures 4^oC, 13^oC and 24^oC that correspond to three different flow regimes. For these cases, the evolution of the cost functional and percentage of error of the surrogate GA solutions are given in Figure 7. It can be noticed that the cost functional has a good decay behavior and that the recorded percentage of error at the end of the surrogate GA is nearly less than 4% almost everywhere in the time interval, where the percentage of error at time t_n between the CFD temperature $\hat{\Theta}^n$ and the approximate temperature $\widetilde{\Theta}^n$ is given by the expression

$$\epsilon(t_n) = 100 \frac{\sqrt{\int_\Omega \left(\widetilde{\Theta}^n - \hat{\Theta}^n\right)^2 dx}}{\sqrt{\int_\Omega \hat{\Theta}^{n^2} dx}}$$

This is further inspected from a visual perspective in Figures 8, 9 and 10, where it can be seen that the surrogate GA succeeded to track the provided target temperature and catch most of the dynamics features present in the temperature.

In terms of computational effort, the proposed surrogate GA is very efficient and performs in near-real time. The overall computational time needed to perform 20 generations in a single cluster was less than two minutes. This represents a very important gain in CPU time compared to traditionally used high fidelity approaches in inverse problems of time dependent turbulent flows.



(a) Comparison of inlet temperatures (b) Percentage of error

Figure 6: Comparison of the predicted inlet temperatures obtained by the surrogate GA for the different tests, and the associated percentage of error with respect to the CFD simulation.

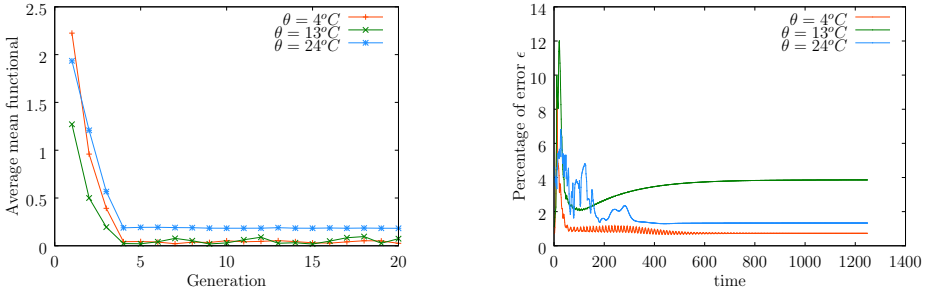


Figure 7: Evolution of the mean averaged functional by the surrogate GA and percentage of errors over the time interval of the converged temperature solution.

6. Conclusions

In this paper, we proposed a surrogate optimization approach by combining the Genetic Algorithm and the barycentric interpolation. The barycentric interpolation is presented here as an equation-free approach that allows to predict via interpolation the evolution of new untrained solutions without any knowledge of the physics hidden behind. The numerical

(a) High fidelity temperature associated to $\hat{\theta} = 4^oC$

(b) Approximate Temperature by the surrogate GA

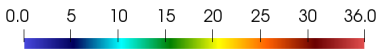
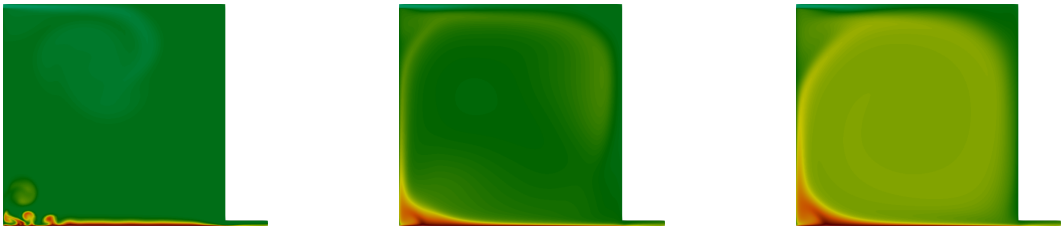
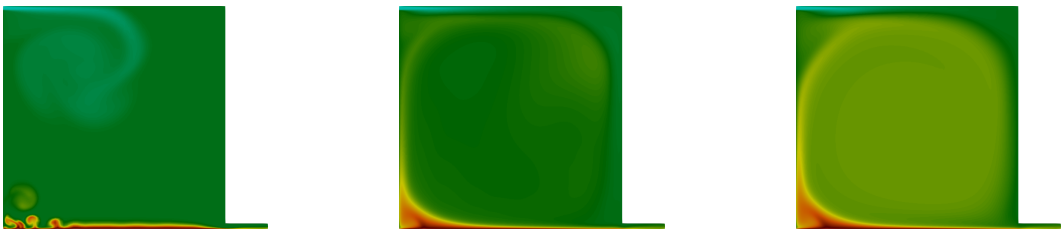


Figure 8: Comparison of the high fidelity and surrogate GA temperature solutions for $\hat{\theta} = 4^oC$ at three different instants of the flow. The first columns is the temperature at $t = 8.75s$ where the first thermal plumes appear, the middle is temperature at $t = 55s$ and the last one is the temperature at the established regime ($t = 1250s$).



(a) High fidelity temperature associated to $\hat{\theta} = 13^oC$



(b) Approximate Temperature by the surrogate GA

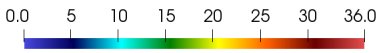
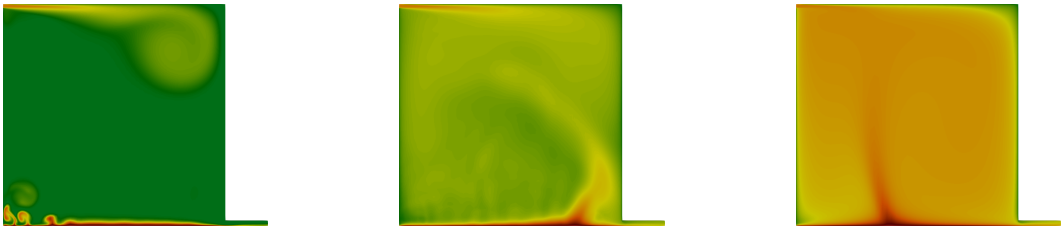
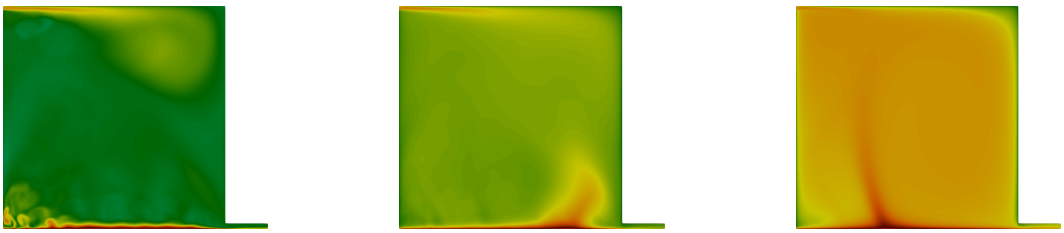


Figure 9: Comparison of the high fidelity and surrogate GA temperature solutions for $\hat{\theta} = 13^oC$ at three different instants of the flow. The first columns is the temperature at $t = 8.75s$ where the first thermal plumes appear, the middle is temperature at $t = 55s$ and the last one is the temperature at the established regime ($t = 1250s$).



(a) High fidelity temperature associated to $\hat{\theta} = 24^oC$



(b) Approximate Temperature by the surrogate GA

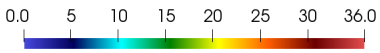


Figure 10: Comparison of the high fidelity and surrogate GA temperature solutions for $\hat{\theta} = 24^oC$ at three different instants of the flow. The first columns is the temperature at $t = 8.75s$ where the first thermal plumes appear, the middle is temperature at $t = 55s$ and the last one is the temperature at the established regime ($t = 1250s$).

assessments of the surrogate is performed on the inverse problem involving a turbulent mixed convection problem, where the variation is carried out on the inlet temperature. This strategy succeeded to track the optimal solutions and to deliver satisfying approximations in less than two minutes. This first successful attempt on turbulent dynamics endorses the potential of this approach and opens the path to the perspective of its application to multi-parametric turbulent data solutions.

Acknowledgement

This material is based upon work financially supported by CPER BATIMENT DURABLE - Axe 3 "Qualité des Environnement Intérieurs (QEI)" (P-2017-BAFE-102) and French Astrid ANR MODULO'PI (ANR-16-ASTR-0018 MODUL'O II).

References

[1] W. Liu and Q. Chen, "Optimal air distribution design in enclosed spaces using an adjoint method," *Inverse Problems in Science and Engineering*, vol. 23, no. 5, pp. 760–779, 2015.

[2] W. Liu, M. Jin, C. Chen, and Q. Chen, "Optimization of air supply location, size, and parameters in enclosed environments using a computational fluid dynamics-based adjoint method," *Journal of Building Performance Simulation*, vol. 9, no. 2, pp. 149–161, 2016.

[3] S. Nabi, P. Grover, and C. Caulfield, "Adjoint-based optimization of displacement ventilation flow," *Building and Environment*, vol. 124, pp. 342 – 356, 2017.

[4] S. Nabi, P. Grover, and C. Caulfield, "Nonlinear optimal control strategies for buoyancy-driven flows in the built environment," *Computers & Fluids*, vol. 194, p. 104313, 2019.

[5] J. H. Holland, *Adaptation in Natural and Artificial Systems.* Ann Arbor, MI: University of Michigan Press, 1975. second edition, 1992.

[6] Y. Xue, Z. J. Zhai, and Q. Chen, "Inverse prediction and optimization of flow control conditions for confined spaces using a CFD-based genetic algorithm," *Building and Environment*, vol. 64, pp. 77 – 84, 2013.

[7] T. Dias and L. F. Milanez, "Optimal location of heat sources on a vertical wall with natural convection through genetic algorithms," *International Journal of Heat and Mass Transfer*, vol. 49, no. 13, pp. 2090 – 2096, 2006.

[8] A. Tallet, C. Allery, and C. Leblond, "Optimal flow control using a POD based Reduced-Order Model," *Numerical Heat Transfer, Part B*, vol. 170, 2016.

[9] M. Bergmann, L. Cordier, and J.-P. Brancher, "Optimal rotary control of the cylinder wake using proper orthogonal decomposition reduced-order model," *Physics of Fluids*, vol. 17, no. 9, pp. 97–101, 2005.

[10] D. Amsallem and C. Farhat, "An interpolation method for adapting reduced-order models and application to aeroelasticity," *AIAA Journal*, pp. 1803–1813, 2008.

[11] M. Oulghelou and C. Allery, "A fast and robust sub-optimal control approach using reduced order model adaptation techniques," *Applied Mathematics and Computation*, vol. 333, pp. 416 – 434, 2018.

[12] S. Suresh, S. Omkar, V. Mani, and T. Guru Prakash, "Lift coefficient prediction at high angle of attack using recurrent neural network," *Aerospace Science and Technology*, vol. 7, no. 8, pp. 595–602, 2003.

[13] A. Shahrokhi and A. Jahangirian, "A surrogate assisted evolutionary optimization method with application to the transonic airfoil design," *Engineering Optimization*, vol. 42, no. 6, pp. 497–515, 2010.

[14] K. Giannakoglou, D. Papadimitriou, and I. Kampolis, "Aerodynamic shape design using evolutionary algorithms and new gradient-assisted metamodels," *Computer Methods in Applied Mechanics and Engineering*, vol. 195, no. 44, pp. 6312–6329, 2006.

[15] M. Ebrahimi and A. Jahangirian, "Accelerating global optimization of aerodynamic shapes using a new surrogate-assisted parallel genetic algorithm," *Engineering Optimization*, vol. 49, no. 12, pp. 2079–2094, 2017.

[16] E. Andrés, S. Salcedo-Sanz, F. Monge, and A. Pérez-Bellido, "Efficient aerodynamic design through evolutionary programming and support vector regression algorithms," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10700–10708, 2012.

[17] F. Tejero, D. G. MacManus, and C. Sheaf, "Surrogate-based aerodynamic optimisation of compact nacelle aero-engines," *Aerospace Science and Technology*, vol. 93, p. 105207, 2019.

[18] M. Yoshimura, K. Shimoyama, T. Misaka, and S. Obayashi, "Topology optimization of fluid problems using genetic algorithm assisted by the kriging model," *International Journal for Numerical Methods in Engineering*, vol. 109, no. 4, pp. 514–532, 2017.

[19] E. Iuliano and D. Quagliarella, "Proper orthogonal decomposition, surrogate modelling and evolutionary optimization in aerodynamic design," *Computers & Fluids*, vol. 84, pp. 327–350, 2013.

[20] M. Oulghelou and C. Allery, "Non-intrusive reduced genetic algorithm for near-real time flow optimal control," *International Journal for Numerical Methods in Fluids*, https://doi.org/10.1002/fld.4820, 2020.

[21] R. R. Madadi and C. Balaji, "Optimization of the location of multiple discrete heat sources in a ventilated cavity using artificial neural networks and micro genetic algorithm," *International Journal of Heat and Mass Transfer*, vol. 51, no. 9, pp. 2299 – 2312, 2008.

[22] M. Cheng, F. Fang, C. Pain, and I. Navon, "Data-driven modelling of nonlinear spatio-temporal fluid flows using a deep convolutional generative adversarial network," *Computer Methods in Applied Mechanics and Engineering*, vol. 365, 2020.

[23] D. Xiao, C. Heaney, L. Mottet, F. Fang, W. Lin, I. Navon, Y. Guo, O. Matar, A. Robins, and C. Pain, "A reduced order model for turbulent flows in the urban environment using machine learning," *Building and Environment*, vol. 148, pp. 323–337, 2019.

[24] O. San, R. Maulik, and M. Ahmed, "An artificial neural network framework for reduced order modeling of transient flows," *Communications in Nonlinear Science and Numerical Simulation*, vol. 77, pp. 271–287, 2019.

[25] J. Yu, C. Yan, and M. Guo, "Non-intrusive reduced-order modeling for fluid problems: A brief review," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 233, no. 16, pp. 5896–5912, 2019.

[26] S. Ahmed, S. Rahman, O. San, A. Rasheed, and I. Navon, "Memory embedded non-intrusive reduced order modeling of non-ergodic flows," *Physics of Fluids*, vol. 31, no. 12, 2019.

[27] E. Massart, P.-Y. Gousenbourger, N. Son, T. Stykel, and P.-A. Absil, "Interpolation on the manifold of fixed-rank positive-semidefinite matrices for parametric model order reduction: preliminary results," 08 2019.

[28] E. Massart and P.-A. Absil, "Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 41, no. 1, pp. 171–198, 2020.

[29] M. Oulghelou and C. Allery, "A Riemannian barycentric interpolation : Application to the parametric unsteady navier-stokes reduced order model," *arXiv:2009.11231*, 2020.

[30] M. Oulghelou and C. Allery, "Non intrusive method for parametric model order re-

duction using a bi-calibrated interpolation on the grassmann manifold," *Journal of Computational Physics*, vol. 426, p. 109924, 2021.

[31] V. Kozeny, "Genetic algorithms for credit scoring: Alternative fitness function performance comparison," *Expert Systems with Applications*, vol. 42, no. 6, pp. 2998 – 3004, 2015.

[32] S. T. Selvi, S. Baskar, and S. Rajasekar, "Chapter 17 - application of evolutionary algorithm for multiobjective transformer design optimization," in *Classical and Recent Aspects of Power System Optimization* (A. F. Zobaa, S. H. A. Aleem, and A. Y. Abdelaziz, eds.), pp. 463 – 504, Academic Press, 2018.

[33] J. S. Arora, "Chapter 17 - nature-inspired search methods," in *Introduction to Optimum Design (Fourth Edition)* (J. S. Arora, ed.), pp. 739 – 769, Boston: Academic Press, fourth edition ed., 2017.

[34] S. Ali, H. Lu, S. Wang, T. Yue, and M. Zhang, "Chapter two - uncertainty-wise testing of cyber-physical systems," vol. 107 of *Advances in Computers*, pp. 23 – 94, Elsevier, 2017.

[35] B. V. Kumar, G. Karpagam, and Y. Zhao, "Chapter 9 - evolutionary algorithm with memetic search capability for optic disc localization in retinal fundus images," in *Intelligent Data Analysis for Biomedical Applications* (D. J. Hemanth, D. Gupta, and V. E. Balas, eds.), Intelligent Data-Centric Systems, pp. 191 – 207, Academic Press, 2019.

[36] K. Khoo and P. Suganthan, "Evaluation of genetic operators and solution representations for shape recognition by genetic algorithms," *Pattern Recognition Letters*, vol. 23, no. 13, pp. 1589 – 1597, 2002.

[37] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1st ed., 1989.

[38] X. Yang, Z. Yang, G. hua Lu, and J. Li, "A gray-encoded, hybrid-accelerated, genetic algorithm for global optimizations in dynamical systems," *Communications in Nonlinear Science and Numerical Simulation*, vol. 10, no. 4, pp. 355 – 363, 2005.

[39] L. Sirovich, "Turbulence and the dynamics of coherent structures : Part I, II and III," *Quarterly of Applied Mathematics*, pp. 461–590, 1987.

[40] D. Blay, S. Mergui, and C. Niculae, "Confined turbulent mixed convection in the presence of a horizontal buoyant wall jet," *ASME Heat Transfer Division*, vol. 213, pp. 65–72, 1992.

[41] *OpenFOAM.* https://openfoam.org, 2017.

[42] V. Yakhot, V. Orszag, S. Thangam, T. B. Gatski, and C. G. Speziale, "Development of turbulence models for shear flows by a double expansion technique," *Physics of Fluids A: Fluid Dynamics*, vol. 4, no. 7, pp. 1510–1520, 1992.