DOI: 10.13009/EUCASS2019-403

Design of Experiment Method in Objective Space for Machine Learning of Flow Structures

Runze LI, Yufei ZHANG** and Haixin CHEN****
** School of Aerospace Engineering, Tsinghua University, Beijing, China*
lirz16@mails.tsinghua.edu.cn
*** School of Aerospace Engineering, Tsinghua University, Beijing, China*
zhangyufei@tsinghua.edu.cn
**** School of Aerospace Engineering, Tsinghua University, Beijing, China*
chenhaixin@tsinghua.edu.cn

Abstract

In the field of analyzing the response or relation of objectives (e.g. flow structures and performances), sufficient and various information of objective space should be provided. A design of experiment method in order to generate samples with various objectives based on adaptive sampling strategy is proposed, and the space-filling property in the objective space is compared with other methods in several test functions. Then the proposed method is utilized to generate airfoils with different pressure coefficient distribution features, and the relation among these features and drag is presented.

1. Introduction

Design of experiment (DoE) methods, or sampling methods, are usually utilized for generating sample inputs (i.e. design variables, $\mathbf{x} \in \mathbb{D}^x \subseteq \mathbb{R}^{n_x}$) in order to achieve certain space-filling property or to provide as much information as possible. [1] The importance of sample generation to machine learning has been stated by several researchers [1][2], and both the size and quality of the sample set are extremely important. There are two types of DoE methods, i.e. stationary methods and adaptive methods. [2] Stationary methods are designed to achieve the most favorable exploration by evenly spreading samples in the input space, and they only take advantage of the topology information of input space. Whereas the adaptive methods are designed for both exploration and exploitation, and the response of outputs (i.e. objectives, $\mathbf{p} \in \mathbb{D}^p \subseteq \mathbb{R}^{n_p}$) are also considered in these methods. Surrogate models in adaptive methods are implemented to approximate the relations between inputs and outputs, and the samples are sequentially generated based on several criteria to most effectively benefit the surrogate model. [3]

Surrogate model is an important tool of machine learning technologies for studying or approximating relations between inputs and outputs, they are quite effective for approximating high-dimensional non-linear relations $\mathbf{x} \rightarrow \mathbf{p}$, and they have been applied to construct aircraft forces and moments model or used in multi-disciplinary designs. [3,4] However, the challenge can be quite different when a new type of relations needs to be studied.

In the process of civil aircraft aerodynamic optimization design, flow structures or flow filed features are as important as flight performances. The cruise performances, such as lift, drag and pitching moment, directly depend on flow structures like spanwise load distribution, shock wave strength, aft loading, etc. And the off-design robustness or flight boundary are essentially the outcomes of flow structure evolution. Therefore, the optimization methods or engineers need the knowledge of the relations between flight performances and flow structures, in order to generate results with better off-design characteristics and robustness. For example, pressure distribution guided (PDG) optimization design method [5] utilizes flow features like foil/wing section pressure distribution features to implicate foil/wing performances, and the relation between flow structures and performances are used in these optimizations. Korn's equation [6] states the relation among thickness, swept angle, drag divergence Mach number and lift coefficient. And Oswatitsch's theorem shows that Mach number in front of shock wave can be used to estimate wave drag [7]. Unfortunately, the number of relations obtained by engineering experiences or aerodynamic theories are very limited, but with the help of machine learning technologies, new relations could be discovered and applied in the aircraft designs.

The nature of relations between flow structures and performances is the relation between objectives. For aircraft aerodynamic optimizations, the geometries are design variables, and then flow field is evaluated based on a specific geometry. Flow structures and performance coefficients are extracted from the flow field afterwards, therefore they

Copyright © 2019 by First Author and Second Author. Published by the EUCASS association with permission.

DOI: 10.13009/EUCASS2019-403

are both objectives. On the other hand, the flow structures are quantitative descriptions of flow field, thus they can also be regarded as the intrinsic variables ($\mathbf{y} \in \mathbb{D}^y \subseteq \mathbb{R}^{n_y}$) in the geometry-performance relations, i.e. $\mathbf{x} \to \mathbf{y} \to \mathbf{p}$. Therefore, in order to properly study the relation of $\mathbf{y} \to \mathbf{p}$, DoE methods should be applied to the objective space of $\mathbf{y}$ in $\mathbb{D}^y$ to provide sufficient information of $\mathbf{y}$ and $\mathbf{p}$.

In order to improve the quality of $\mathbf{y} \to \mathbf{p}$ approximation, a new DoE method is proposed to generate samples providing as many different combinations of objectives $\mathbf{y}$ as possible. In other word, the new method tries to obtain a good space-filling property in objective space $\mathcal{D}^y$, rather than being used as a DoE method in input space. The DoE method in objective space is based on an adaptive sampling method using radial basis function (RBF) response surface, the new samples are sequentially generated by RBF response surface (RBFRS) based optimizations to explore the objective space. The method is tested and compared with some other DoE methods on several test functions to demonstrate its ability of objective space exploration. Then the method is used to generate airfoils with various flow structures, some relations among flow structures and performances are studied.

2. Methods

Latin Hypercube Sampling (LHS) method [8] is a commonly used stationary sampling method, but its space filling behavior is not always guaranteed. There have been several modifications developed to improve its performance, and the improved LHC method can achieve good enough input space filling property for most situations [9]. Adaptive sampling methods are mostly developed with surrogate models, and several criteria are used to improve approximation accuracy. RBFRS is one commonly used surrogate model. [1] And a relatively cheap criterion combining power function and curvature for RBFRS can be used to achieve good exploration and exploitation. [10]

In order to generate samples with different objectives, a diversity criterion is proposed to quantify the variety of sample objectives or the space-filling property in the objective space. Then the diversity criterion is used in the RBFRS-based optimization for each DoE iteration. New samples are generated for both increasing objective diversity and RBFRS quality.

2.1 Diversity criterion

Diversity of a sample set measures the state of samples being different or diverse, it is usually called space-filling quality when considering the input space diversity. In order to quantify the diversity of samples in objective space $\mathbb{D}^y$, a criterion θ for diversity should meet 3 requirements:

(1) $\theta \in [0,1]$, and θ should increase with increasing diversity;

(2) θ must not decrease when a new sample is added, and it should have a smaller increment when the new sample is more similar to existed ones;

(3) $\theta = 1$ stands for a good enough space filling property for the entire space $\mathbb{D}^y$, and $\theta = 0$ means there are less than 2 samples in the sample set.

A higher diversity means more different combinations of objectives are achieved for the same sample amount N_S. But it needs to be stated that the objective space $\mathbb{D}^y$ is assumed to be a multi-dimensional hyper-cube, but in reality, there usually are constraints among objectives, and the actual objective space is only a sub-set of $\mathbb{D}^y$. And usually the upper bound of diversity is an unknown value less than 1.

In thios section, a diversity criterion based on 1D string energy is proposed. When there are n samples located in the 1D space of length D, and let d_{max} be the largest distance between two samples in the 1D space, then the diversity criterion is used to quantify the space-filling property of the $n-2$ samples between these 2 samples. Assume there are one spring between each two adjacent samples, and the $n-1$ springs have the same original length d_0 and elastic modulus $k = 0.5$. The springs are not connected to samples, and thus they can only be compressed but not stretched. Then the total elastic potential energy E is:

$$E = \sum_{j=1}^{n-1} \lfloor d_0 - d_j \rfloor^2 \tag{1}$$

Where the symbol $\lfloor\cdot\rfloor$ means:

DOI: 10.13009/EUCASS2019-403

$$\lfloor x \rfloor = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2}$$

d_j is the distance between the two adjacent sample j and $j+1$. If $d_0 = t \cdot \frac{D}{N_{S,max}-1}$, and coefficient t is chosen to make $d_0 \lesssim \frac{d_{max}}{n-1}$, then the upper and lower bound of total elastic potential energy are:

$$\mathrm{E}_{max} = (n-1)d_0^2 \tag{3}$$

$$\mathrm{E}_{min} = 0 \tag{4}$$

And set:

$$\varphi_E = \frac{\mathrm{E}_{max} - E}{\mathrm{E}_{max} - \mathrm{E}_{min}} \cdot \frac{n-1}{N_{S,max}-1} = \frac{(n-1) - \sum_j^{n-1} \lfloor 1 - d_j/d_0 \rfloor^2}{N_{S,max}-1} \tag{5}$$

Then the diversity criterion can be descripted by the following formula.

$$\theta = \varphi_E \left(\frac{n \cdot d_{max}}{N_{S,max} \cdot D} \right)^{\mu} \tag{6}$$

Coefficient μ is used to balance the weight between exploring region beyond the current far most two samples (i.e. the distance between them is d_{max}) and exploitation within them, usually $\mu = 0.5$. It can be easily proved that the diversity criterion θ can meet the previous requirements (1) and (3), and it can be proved that it can also meet the 2nd requirement by looking at the following two situations. Let φ_E^+ stands for the new φ_E when a new sample is added.

When a new sample is added beyond the current far most two samples, the φ_E increment can be written in:

$$\varphi_E^+ - \varphi_E = \frac{1}{N-1} \left[1 - \left\lfloor 1 - \frac{d_n}{d_0} \right\rfloor^2 \right] \tag{7}$$

When a new sample is added with the current far most two samples, one spring length is changed, let's say it is d_1. Then the φ_E increment can be written in:

$$\varphi_E^+ - \varphi_E = \frac{1}{N-1} \left[1 - \left\lfloor 1 - \frac{d_n}{d_0} \right\rfloor^2 - \left\lfloor 1 - \frac{d_1}{d_0} - \frac{d_n}{d_0} \right\rfloor^2 + \left\lfloor 1 - \frac{d_1}{d_0} \right\rfloor^2 \right] \tag{8}$$

And it can be easily proved for both cases that $\varphi_E^+ - \varphi_E \geq 0$, and $\frac{\partial(\varphi_E^+ - \varphi_E)}{\partial d_n} \geq 0$. Since $\frac{n \cdot d_{max}}{N \cdot D}$ also meets the 2nd requirement, so the diversity criterion θ can meet all previous requirements.

As for multi-dimensional space, the diversity criterion is trivially extended to high-dimensional problems by defining D as the maximum distance of the objective space $\mathcal{D}^y$. And the spring length d_j of sample j is defined as the minimum distance to other existed samples.

$$D = \left[\sum_k^{n_y} \left(y_{upper\ bound}[k] - y_{lower\ bound}[k] \right)^2 \right]^{1/2} \tag{9}$$

Where $y_{upper\ bound}[k]$ and $y_{lower\ bound}[k]$ are the upper and lower bound of kth component of objective vector $\mathbf{y}$.

2.2 DoE method in objective space

After an initial sample set generated by stationary sampling methods like improved LHC, an adaptive sampling procedure is conducted. The adaptive sampling is achieved by optimizations on the surrogate model to generate new sample locations $\{\mathbf{x}_i\}$ $(i = 1, \dots, N_{Add})$ in each iteration, and then update the sample set and surrogate model after the evaluation of these samples. It should be noticed that constraints ($\mathbf{x} \in \mathbb{D}^x$, $\mathbf{y} \in \mathbb{D}^y$) should always be met when searching for new samples.

```
> Generate initial sample set $\{(\mathbf{x}_i, \mathbf{y}_i)\}(i = 1, \dots, N_{S,0})$;
> $N_S = N_{S,0}$;
> for $k = 1, \dots, (N_{S,max} - N_{S,0})/N_{Add}$ {
>         multi-objective optimization on RBFRS to increase diversity;
```

DOI: 10.13009/EUCASS2019-403

```
>                max$\{C, \theta\}$; ($C$ is the criterion in [10])
>                subject to: $\mathbf{x} \in \mathbb{D}_x$, $\mathbf{y} \in \mathbb{D}_y$;
>        Evaluation of new samples, i.e. $\{(\mathbf{x}_i, \mathbf{y}_i)\}$ ($i = 1, \dots, N_{Add}$);
>        Update sample set, $N_S = N_S + N_{Add}$;
>        Update RBFRS;
> }
> end
```

3. Application on test functions

3.1 One dimensional test function

The DoE method in objective space is firstly tested on a one-dimensional function descripted as following formula. Then the method is compared with improved LHC method and an adaptive method using criterion in [10].

$$y = \frac{10 \cdot sin(\pi x/5)}{(\pi x/5)(|1 - x/32.5| + 0.02)}, \quad x \in [-10,50] \tag{10}$$

Figure 1 shows the samples generated by the three methods, the number of samples is the same of 40 for all 3 methods. It can be seen that the improved LHC method and adaptive DoE method have similar results in this problem, which is mainly because that the test function is relatively simple. However, the DoE method in objective space (labelled as Objective Space DoE) can achieve the intentional exploration in the objective space. Larger range of the objective is explored and more samples with different objectives are generated.

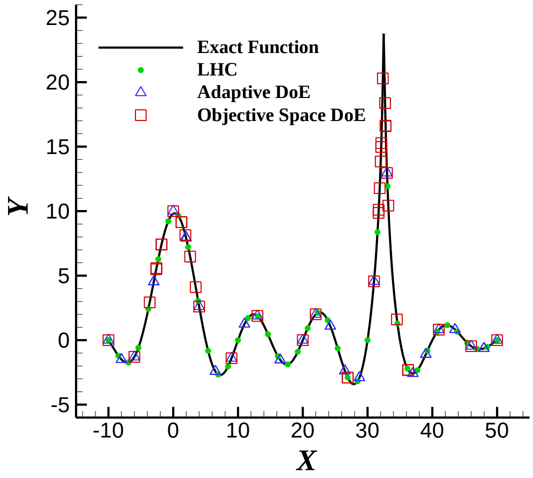


Figure 1: Samples generated by different DoE methods

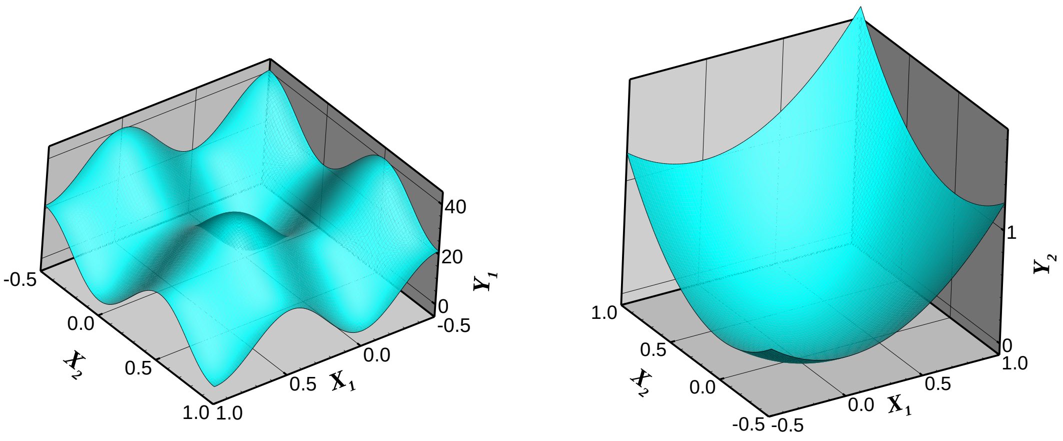
3.2 Rastrigin-Sphere function

Relatively more complex functions are further used to test the effect of DoE method in objective space. The test function is a dual-objective function, and the objectives are from Rastrigin function and Sphere function, respectively. The detailed expression can be found in the following formula. This function is used in both two-dimensional and 10-dimensional test cases (i.e. $n_x = 2, 10$).

$$\begin{aligned} &y_1 = 10n_x + \sum_{k=1}^{n_x}[x_k^2 - 10\cos(2\pi x_k)] \\ &y_2 = \sum_{k=1}^{n_x} x_k^2 \\ &\mathbf{x} \in [-0.5,1]^{n_x} \end{aligned} \tag{11}$$

In the two-dimensional case, Rastrigin and Sphere functions are relatively simple, as shown in Figure 2. However, the objective space can be complex. 400 samples are generated by the 3 methods, and their distributions in input space and objective space are plotted in Figure 3. And the cyan shade in these figures shows the exact distribution of objective space. It can be seen that all three methods can achieve a rather good diversity, but still the proposed method has a better objective space filling property than the other two.

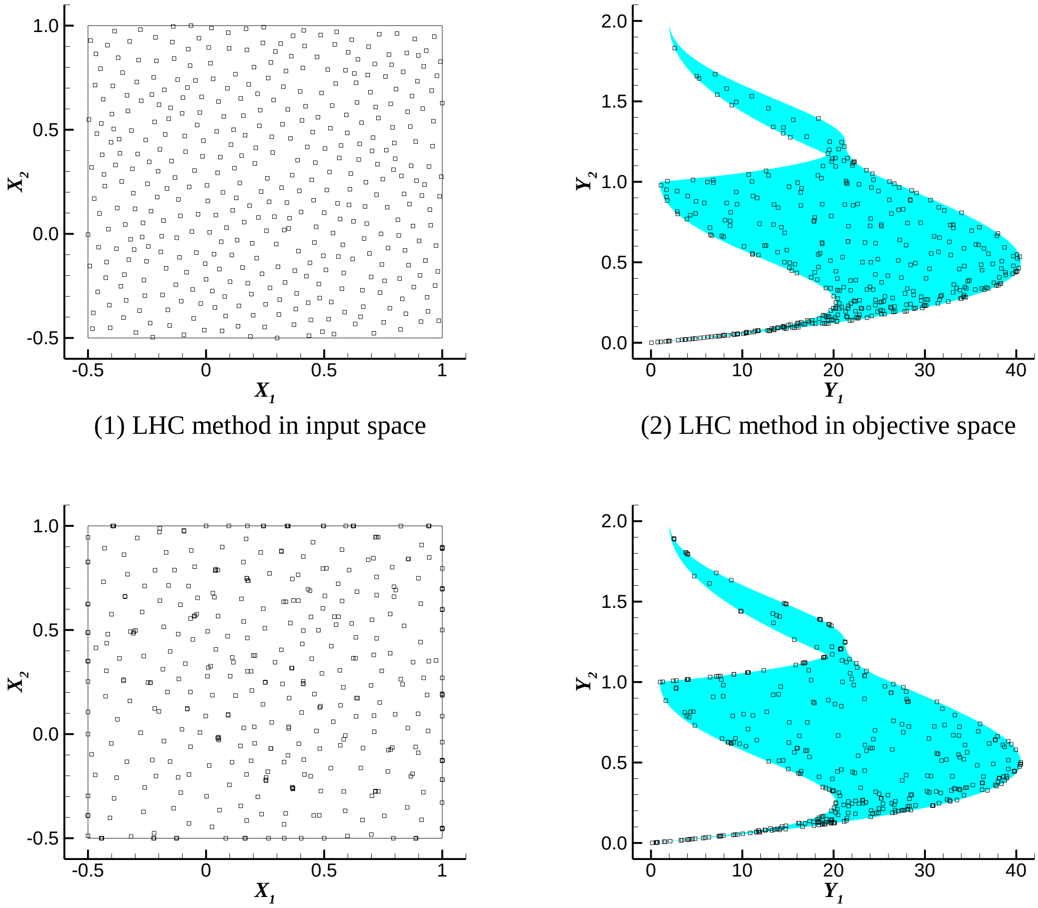
DOI: 10.13009/EUCASS2019-403



(1) Exact function of Rastrigin

(2) Exact function of Sphere

Figure 2: Exact function of test functions

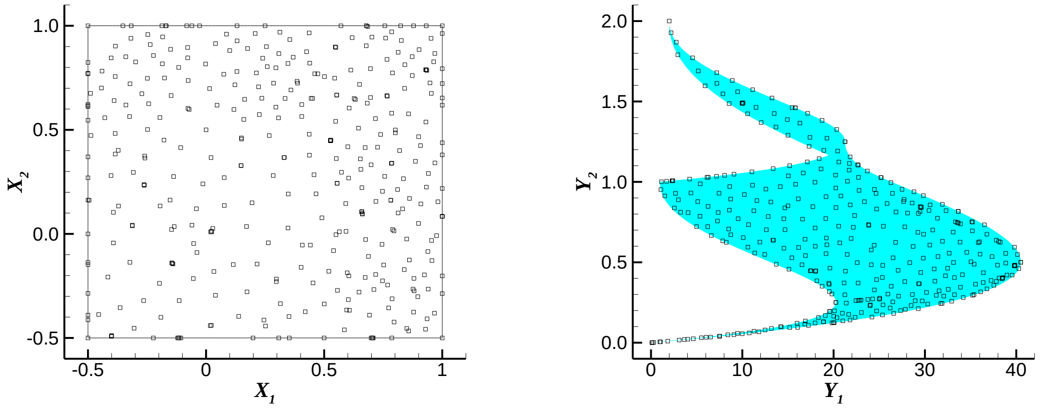


(1) LHC method in input space

(2) LHC method in objective space

(3) Adaptive DoE method in input space

(4) Adaptive DoE method in objective space



(5) Objective Space DoE in input space (6) Objective Space DoE in objective space

Figure 3: Sample distribution of different methods in input space and output space

On the other hand, the functions can be rather complex in the 10-dimensional case, and it is virtually impossible to get the exact objective space boundaries. The red dashed curves in Figure 4 are the estimated boundary of objective space obtained by a number of multi-objective optimizations, and they are plotted for better demonstration of the DoE methods. The three methods are applied to generate 400 samples to explore the objective space, and their sample distribution are plotted in Figure 4. The stationary method, i.e. LHC, can only explore a rather limited region of objective space, and the adaptive sampling method can have a relatively better performance. Whereas, the DoE method in objective space can achieve a rather good exploration of the objective space.

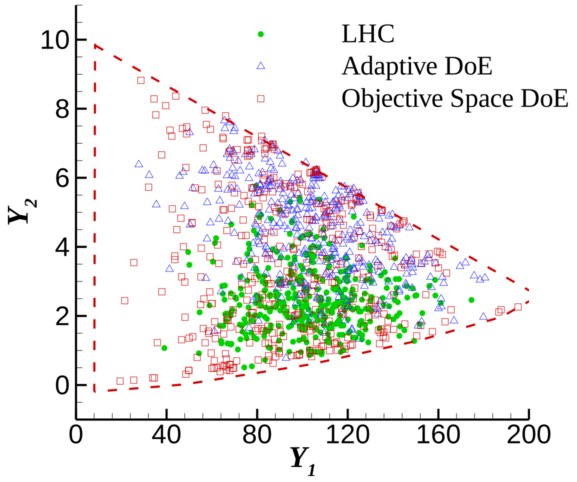


Figure 4: Sample distribution of different methods in objective space

4. Application on airfoil generation

The DoE method in objective space is applied to generated airfoils with various flow structures. The flow field is evaluated by Euler equation at a fixed lift coefficient of 0.7, and the free stream Mach number is 0.76, Reynolds number equals 5.0 million based on airfoil chord length (equals 1). The concerned objectives are shock wave location X_1, wall Mach number in front of shock wave M_{w1}, and wall Mach number of suction peak $M_{w,suction}$. The wall Mach number is defined as the equivalent Mach number of surface pressure coefficient calculated by isentropic relation based on free stream Mach number.

The airfoils are generated to obtain diverse objectives under the following constraints. Firstly, the leading edge radius must not be smaller than 0.007. Secondly, the drag coefficient should not exceed 0.1 in order to avoid impractical airfoils. And lastly, only the airfoils with single shock wave are considered. Therefore, the DoE method does not only search for samples with different objectives, but also needs to search within the bound of these pre-set constraints.

The initial sample set is 200 airfoils generated in the process of an airfoil drag reduction optimization, and 8 samples are generated by the proposed method in each iteration Total amount of 2000 airfoils are generated, and 1300 of them

DOI: 10.13009/EUCASS2019-403

are feasible under the previous constraints. Figure 5 shows the distribution of the initial sample set in the objective space.

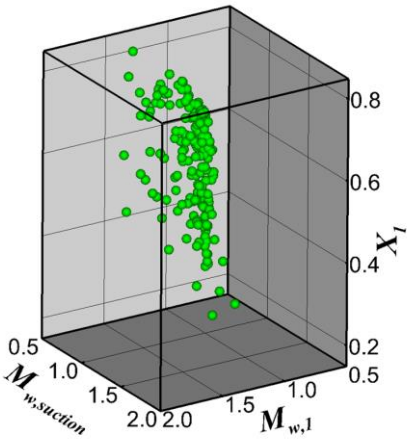
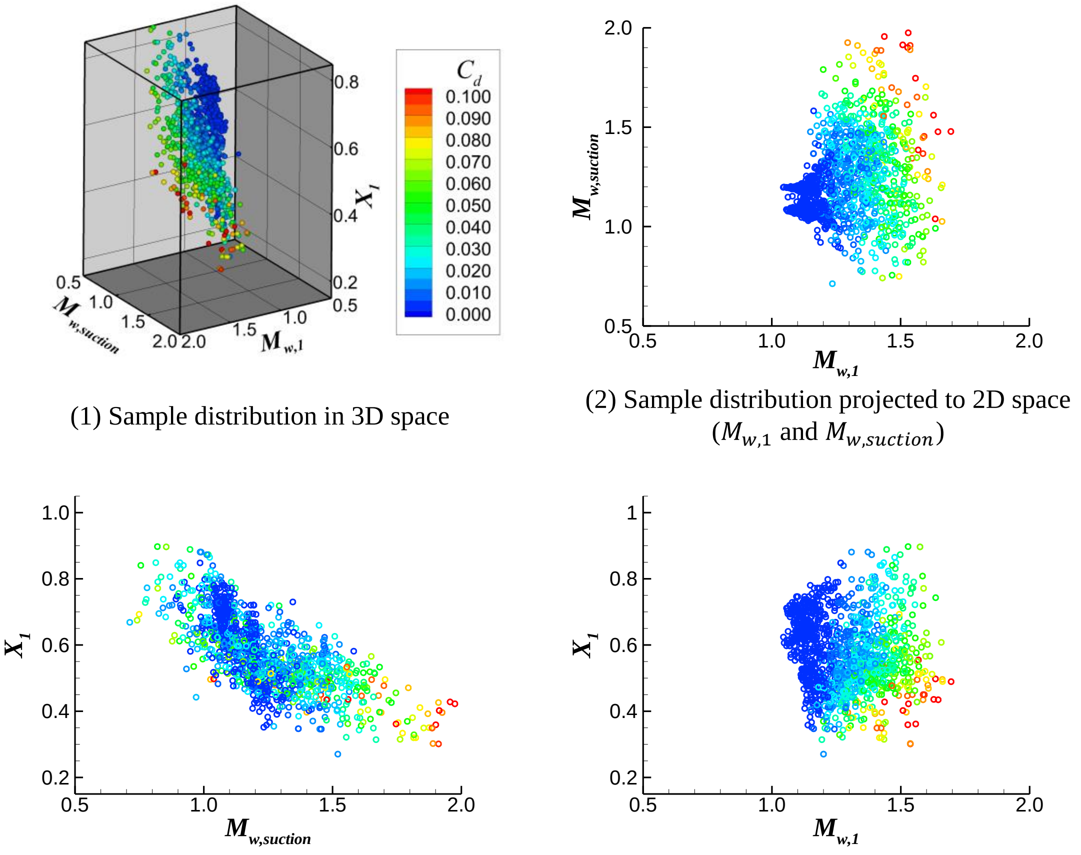


Figure 5: Initial sample set distribution in objective space

Figure 6(1) shows the distribution of feasible samples in objective space, and the samples are further coloured by their drag coefficient C_d. Figure 6(2-4) show the sample distribution projected in 2 dimensions. It can be seen that the DoE method in objective space can better fill the objective space than optimizations, therefore the analysis based on sample sets generated by the proposed method can be more convincing and close to the truth. It can be seen that the drag coefficient strongly depends on the Mach number in front of shock waves, and almost not affected by the shock wave position or suction peak. However, there is a strong correlation between X_1 and $M_{w,suction}$, whereas X_1 and $M_{w,1}$, or $M_{w,1}$ and $M_{w,suction}$ are basically independent.



(1) Sample distribution in 3D space

(2) Sample distribution projected to 2D space ($M_{w,1}$ and $M_{w,suction}$)

(3) Sample distribution projected to 2D space ($M_{w,suction}$ and X_1)

(4) Sample distribution projected to 2D space ($M_{w,1}$ and X_1)

Figure 6: Sample distribution in objective space

DOI: 10.13009/EUCASS2019-403

5. Conclusion

In order to generate samples with good space-filling property in the objective space, a design of experiment method is proposed based on adaptive sampling strategy using the radial basis function response surface. The method sequentially generates new samples to improve RBFRS quality and objective diversity, while keeping the new samples within the bounds of pre-set constraints. The proposed method is tested and compared with Latin hypercube method and an adaptive sampling method on 3 test functions of 1, 2, and 10 dimensional input space. The proposed method can achieve a better objective space filling property in all these cases, especially in the high dimensional case. Then the method is used to generated airfoils to study the relation among shock wave location, wall Mach number in front of shock wave, and wall Mach number of suction peak. And the correlation between drag coefficient and Mach number in front of shock wave, as well as the correlation between Mach number of suction peak and shock wave position are presented.

References

[1] Bhosekar A, Ierapetritou M. Advances in surrogate-based modeling, feasibility analysis, and optimization: A review[J]. Computers & Chemical Engineering, 2018, 108: 250-267.
[2] Wang G G, Shan S. Review of metamodeling techniques in support of engineering design optimization[J]. Journal of Mechanical design, 2007, 129(4): 370-380.
[3] Yondo R, Andres E, Valero E. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses[J]. Progress in Aerospace Sciences, 2018, 96: 23-61.
[4] Forrester A I J, Keane A J. Recent advances in surrogate-based optimization[J]. Progress in aerospace sciences, 2009, 45(1-3): 50-79.
[5] Runze LI, Kaiwen D, Zhang Y, et al. Pressure distribution guided supercritical wing optimization[J]. Chinese Journal of Aeronautics, 2018, 31(9): 1842-1854.
[6] Mason W. Analytic models for technology integration in aircraft design[C]//Aircraft Design, Systems and Operations Conference. 1990: 3262.
[7] Inger G R. Application of Oswatitsch's theorem to supercritical airfoil drag calculation[J]. Journal of aircraft, 1993, 30(3): 415-416.
[8] McKay, M.D., Beckman, R.J., Conover, W.J., 1979. Comparison of three methods forselecting values of input variables in the analysis of output from a computercode. Technometrics 21 (2), 239–245
[9] Garud, 2017, Smart sampling algorithm for surrogate model development.
[10] Mackman, 2013, Comparison of adaptive sampling methods for generation of surrogate