Flow Field Data Mining of Pareto-Optimal Airfoils Using Proper Orthogonal Decomposition

Akira Oyama

Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara 229-8510, Japan

Paul C. Verburg

University of Twente, 7500 AE Enschede, The Netherlands

Taku Nonomura

Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara 229-8510, Japan

Harry W. M. Hoeijmakers

University of Twente, 7500 AE Enschede, The Netherlands

and

Kozo Fujii

Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara 229-8510, Japan

The capability of a proper-orthogonal-decomposition-based data mining approach for the analysis of flow field data of Pareto-optimal solutions is demonstrated. This method enables a designer to extract design knowledge by examining baseline data and a limited number of eigenvectors and orthogonal base vectors. The flow data analyzed herein are the pressure field data of the Pareto-optimal solutions of an aerodynamic transonic airfoil shape optimization problem. The results of the present study indicate that the proper-orthogonal-decomposition-based data mining approach is useful for extracting design knowledge from the flow field data of the Pareto-optimal solutions.

Nomenclature

- $a_m(n)$ = eigenvector of mode $m$
- $c$ = chord length
- $C_d$ = drag coefficient
- $C_l$ = lift coefficient
- $j$ = index of grid points
- $j_{\text{max}}$ = number of grid points
- $m$ = index of mode
- $m_{\text{max}}$ = number of modes ($m_{\text{max}} = n_{\text{max}}$)
- $n$ = index of Pareto-optimal solutions
- $n_{\text{max}}$ = number of Pareto-optimal solutions
- $p(j,n)$ = pressure of Pareto-optimal solution $n$ at grid point $j$
- $q(j,n)$ = data of Pareto-optimal solution $n$ at grid point $j$ to be analyzed by POD
- $q_{\text{ld,ave}}(j)$ = data of maximum-lift-to-drag-ratio design at grid point $j$

1 Assistant Professor, Department of Space Transportation Engineering, 3-1-1 Yoshinodai, Member AIAA.
2 Graduate student, Department of Mechanical Engineering, Engineering Fluid Dynamics, PO Box 217.
3 Research Associate, Department of Space Transportation Engineering, 3-1-1 Yoshinodai, Member AIAA.
4 Professor, Department of Mechanical Engineering, Engineering Fluid Dynamics, PO Box 217, Senior Member AIAA.
5 Professor, Department of Space Transportation Engineering, 3-1-1 Yoshinodai, Fellow AIAA.

Copyright © 2010 by Akira Oyama, Paul C. Verburg, Taku Nonomura, Harry W. M. Hoeijmakers, and Kozo Fujii. Published by the American Institute of Aeronautics and Astronautics, Inc., with
\[ q'(j,n) = \text{fluctuation of data } q \text{ of Pareto-optimal solution } n \text{ at grid point } j \]
\[ q_{\text{base}}(j,m) = \text{orthogonal base vector of mode } m \]
\[ S_{m1,m2} = \text{covariance of } q_{\text{base}} \text{ of mode } m1 \text{ and mode } m2 \]
\[ x = \text{coordinate in the chordwise direction} \]
\[ y = \text{coordinate in the normal direction} \]

**I. Introduction**

MULTIOBJECTIVE design exploration\(^1\) (MODE) is a framework to extract essential knowledge of a multiobjective design optimization problem, such as tradeoff information between contradicting objectives and the effect of each design parameter on the objectives. In the framework of MODE, Pareto-optimal solutions are obtained by multiobjective optimization using, for example, a multiobjective evolutionary algorithm\(^2\), and important design knowledge is then extracted by analyzing objective function and design parameter values of the obtained Pareto-optimal solutions using data mining approaches, such as the self-organizing map\(^3\) (SOM) and analysis of variance\(^4\). Recently, MODE framework has been applied to a wide variety of design optimization problems, including multidisciplinary design of a regional-jet wing\(^5\)\(^6\), aerodynamic design of the fly-back booster of a reusable launch vehicle\(^7\), aerodynamic design of a flapping airfoil\(^8\), and aerodynamic design of a turbine blade for a rocket engine\(^9\).

However, data mining of objective function and design parameter values is not sufficient. One reason is that the design knowledge of a shape design optimization problem that can be obtained depends on how the shape is parameterized. For example, if an airfoil shape is represented by B-Spline curves and the coordinates of the corresponding control points are considered to be design parameters, it is difficult to obtain design knowledge related to leading edge radius, thickness distribution, and so on. Another reason is that data mining of the objective function and design parameter values does not lead to an understanding of the physics behind the design problem. For example, if only the design parameters of a transonic airfoil were analyzed, it would not be possible to clarify the relation between shock wave generation and aerodynamic characteristics.

In Reference 10, Oyama et al. proposed a new approach based on the proper orthogonal decomposition (POD) for data mining of shape and flow data of Pareto-optimal solutions. In their method, shape data or flow data of the obtained Pareto-optimal solutions are decomposed into baseline data and eigenvectors and orthogonal base vectors of principal modes with POD. This enables a designer to extract design knowledge by examining baseline data and a limited number of eigenvectors and orthogonal base vectors. They demonstrated the capability of the POD-based method for data mining of shape data and one-dimensional flow data (surface pressure data) of the Pareto-optimal solutions of an aerodynamic transonic airfoil design optimization problem.

The objective of the present study is to demonstrate the capability of the POD-based method for data mining of two-dimensional flow field data of the Pareto-optimal solutions. In the present paper, the POD-based method is applied to the pressure field data of the Pareto-optimal solutions of an aerodynamic transonic airfoil design optimization problem. First, the results of data mining of shape data of the Pareto-optimal solutions are presented. Then, data mining of the flow field data of the Pareto-optimal solutions is discussed.

**II. Pareto-optimal Solutions**

The Pareto-optimal solutions of the following design optimization problem are analyzed.

**Objective functions:**
- lift coefficient (maximization)
- drag coefficient (minimization)

**Constraints:**
- lift coefficient must be greater than 0
- maximum thickness must be greater than 0.10 chord length

**Design parameters:**
- coordinates of six control points of the B-Spline curves representing an airfoil shape (Fig. 1)

**Flow conditions:**
- free stream Mach number of 0.8
- angle of attack of 2 degrees
The Pareto-optimal solutions are obtained by the multiobjective evolutionary algorithm (MOEA) used in Reference 8. The present MOEA adopts real number coding because the optimization problem considered herein is a real number optimization problem. The population size is maintained at 64, and the maximum number of generations is set to 60. The initial population is generated randomly so that the initial population covers the entire design space presented in Table 1. The fitness of each design candidate is computed according to Pareto-ranking, fitness sharing, and Pareto-based constraint handling based on its objective function and constraint function values. Here, Fonseca and Fleming’s Pareto-based ranking method and the fitness sharing method of Goldberg and Richardson are used for Pareto-ranking, where each individual is assigned a rank according to the number of individuals dominating the individual. In Pareto-based constraint handling, the rank of feasible designs is determined by the Pareto-ranking based on the objective function values, whereas the rank of infeasible designs is determined by the Pareto-ranking based on the constraint function values. The parents of the new generation are selected through roulette selection from the best 64 individuals among the present generation and the best 64 individuals in the previous generation. A new generation is reproduced through crossover and mutation operators. The term “crossover” refers to an operator that combines the genotype of the selected parents and produces new individuals with the intent of improving the fitness value of the next generation. Here, the blended crossover, the value of $\alpha$ of which is 0.5, is used for crossover between the selected solutions. Mutation is applied to the design parameters of the new generation to maintain diversity. Here, the probability of mutation occurring is 20%, which adds a random disturbance to the corresponding gene of up to 10% of the given range of each design parameter. The capability of the present MOEA to find quasi-optimal solutions has been well validated.

The lift and drag coefficients of each design candidate are evaluated using a two-dimensional Reynolds-averaged Navier-Stokes solver. This code employs total variation diminishing type upwind differencing, the lower-upper symmetric Gauss-Seidel scheme, the turbulence model of Baldwin and Lomax, and the multigrid method.

All of the design candidates and Pareto-optimal solutions are plotted in Fig. 2. The number of Pareto-optimal solutions obtained is 85. A strong tradeoff between lift maximization and drag minimization is observed. The static pressure distributions around the maximum-lift, maximum-lift-to-drag-ratio, and minimum-drag airfoils are also shown in the figure. Figure 3 compares the shapes and surface pressure distributions of the above three designs.

### Table 1. Search range of each design parameter

<table>
<thead>
<tr>
<th>Design parameter</th>
<th>lower bound</th>
<th>upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>0.66</td>
<td>0.99</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0.33</td>
<td>0.66</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0.01</td>
<td>0.33</td>
</tr>
<tr>
<td>$x_4$</td>
<td>0.01</td>
<td>0.33</td>
</tr>
<tr>
<td>$x_5$</td>
<td>0.33</td>
<td>0.66</td>
</tr>
<tr>
<td>$x_6$</td>
<td>0.66</td>
<td>0.99</td>
</tr>
<tr>
<td>$y_1$</td>
<td>-0.1</td>
<td>0.10</td>
</tr>
<tr>
<td>$y_2$</td>
<td>-0.1</td>
<td>0.10</td>
</tr>
<tr>
<td>$y_3$</td>
<td>-0.1</td>
<td>0.10</td>
</tr>
<tr>
<td>$y_4$</td>
<td>0.0</td>
<td>0.20</td>
</tr>
<tr>
<td>$y_5$</td>
<td>0.0</td>
<td>0.20</td>
</tr>
<tr>
<td>$y_6$</td>
<td>0.0</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Figure 1. Parameterization of the airfoil shape. The coordinates of six control points of the B-Spline curves representing an airfoil shape are considered as design parameters.
These figures indicate that the minimum-drag design avoids the generation of strong shock waves, whereas the maximum-lift design generates a strong and large negative pressure region. These figures also show that the maximum-lift-to-drag-ratio design has a shape that is similar to supercritical airfoils. These facts indicate that the obtained Pareto-optimal solutions are good approximations of the true Pareto-optimal solutions.

Figure 2. Distribution of the Pareto-optimal solutions and other design candidates with the pressure distribution around the minimum-drag, maximum-lift-to-drag-ratio, and maximum-lift airfoils.

Figure 3. Shape and surface pressure distributions of the minimum-drag, maximum-lift-to-drag-ratio, and maximum-lift airfoils.
III. Data Mining of Pareto-Optimal Solutions Using Proper Orthogonal Decomposition

In the present study, the shape and pressure field data of Pareto-optimal airfoils are analyzed using the snapshot POD proposed by Sirovich\textsuperscript{22}. The Pareto-optimal solutions are numbered from the minimum-drag design to the maximum-lift design, as shown in Fig. 4. For data mining of the shape data, the $y$ coordinates defined on all grid points around the airfoil are analyzed, and the pressure defined at all grid points is analyzed for the pressure field analysis. The number of grid points around the airfoil is 137, and total number of grid points is 9,849 (201 (chordwise) x 49 (normal)).

![Figure 4. Index of the Pareto-optimal solutions. For the minimum-drag design, $n = 1$; for the maximum-lift design, $n = n_{\text{max}} = 85$.](image)

In the original snapshot POD, the data to be analyzed are decomposed into the mean vector and the fluctuation vector from the mean vector to minimize the error of the reconstructed flow field with a certain number of modes. However, for the analysis of Pareto-optimal solutions, the fluctuation from the mean shape or flow field is not intuitive. Thus, it is reasonable to analyze the fluctuation from a single representative design, for example, the median design. Here, the fluctuation from the maximum-lift-to-drag-ratio design is analyzed. Analysis of the fluctuation from the maximum-lift-to-drag-ratio design had minimal impact on the accuracy of the reconstruction for the present data. The data of the Pareto-optimal solutions are decomposed into the data of the maximum-lift-to-drag-ratio design and fluctuation data as follows:

$$
\begin{bmatrix}
q(1, n) \\
q(2, n) \\
\vdots \\
q(j \text{ max} - 1, n) \\
q(j \text{ max}, n)
\end{bmatrix}
\begin{bmatrix}
q_{l/d_{\text{max}}}(1) \\
q_{l/d_{\text{max}}}(2) \\
\vdots \\
q_{l/d_{\text{max}}}(j \text{ max} - 1) \\
q_{l/d_{\text{max}}}(j \text{ max})
\end{bmatrix}
+ 
\begin{bmatrix}
q'(1, n) \\
q'(2, n) \\
\vdots \\
q'(j \text{ max} - 1, n) \\
q'(j \text{ max}, n)
\end{bmatrix}.
$$

(1)

The fluctuation vector is then expressed by the linear sum of normalized eigenvectors and orthogonal base vectors as follows:
where each eigenvector is determined so that the energy defined by Eq. (3) is maximized as follows:

\[
\sum_{j=1}^{j_{\text{max}}} q_{\text{base}}^2(j, m), \quad m = 1, 2, \ldots, m_{\text{max}}.
\] (3)

The eigenvectors that maximize the energy defined by Eq. (3) can be obtained by solving the eigenvalue problem of the following covariance matrix:

\[
\begin{pmatrix}
S_{1,1} & \cdots & S_{m1,1} & \cdots & S_{m\text{max},1} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
S_{1,m2} & \cdots & S_{m1,m2} & \cdots & S_{m\text{max},m2} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
S_{1,m_{\text{max}}} & \cdots & S_{m1,m_{\text{max}}} & \cdots & S_{m\text{max},m_{\text{max}}}
\end{pmatrix},
\]

where

\[
S_{m1,m2} = \sum_{j=1}^{j_{\text{max}}} q'(j,m1)q'(j,m2).
\] (5)

**IV. Data Mining of Airfoil Shape Using POD**

First, the airfoil shape data of the Pareto-optimal solutions are analyzed. The shape data analyzed here are the \(y\) coordinates defined on all grid points on the airfoil shape. The energy ratios of 10 principal orthogonal base vectors (principal POD modes) to the total energy are shown in Fig. 5. The first mode is dominant (more than 83\%), and the first two modes represent more than 94\% of the total energy.

Figure 6 shows the components of the eigenvectors of the first four modes with respect to the index of the non-dominated solutions \(n\) (left) and the lift coefficient \(C_l(n)\) (right), respectively. This figure indicates that the obtained non-dominated airfoil shapes are categorized into three groups: low-drag designs (1 \(\leq n \leq 39\)), high-lift-to-drag-ratio designs (40 \(\leq n \leq 52\)), and high-lift designs (53 \(\leq n \leq 85\)). As for the low-drag designs, the second mode is dominant as the eigenvector of the first mode is approximately zero. Among the high-lift-to-drag-ratio designs, the airfoil shape does not change much. Among the high-lift designs, the first mode is dominant because the eigenvector of the second mode is approximately zero.
Figure 7 presents the maximum-lift-to-drag airfoil shape and orthogonal base vectors of the first four modes. This figure indicates that the first mode contributes primarily to the most part of the lower surface change. The base vector of the first mode also indicates that thickness near the leading edge should be increased as the lower surface moves upward. This is due to the constraint on the maximum thickness imposed on the design optimization problem. The base vector of the second mode indicates that the second mode contributes primarily to the camber near the trailing edge. Recalling the shapes of the Pareto-optimal solutions are represented by equations (1) and (2), Figures 6 and 7 indicate that the Pareto-optimal low-drag designs increase lift by changing the camber near the trailing edge, while the other parts of the airfoil shape are approximately fixed. For the high-lift designs, lift is increased by moving the lower surface upward without significant changing the trailing edge angle. This movement of the lower surface corresponds to an increase in camber. The thickness near the leading edge is increased as the lower surface moves upward in order to satisfy the constraint applied to the airfoil maximum thickness near the leading edge.
V. Data Mining of Pressure Field Distribution Using POD

To demonstrate the capability of the POD-based data mining method for the analysis of the flow field data, the pressure field data defined on all grid points of the Pareto-optimal airfoil shapes are analyzed. The energy ratios of the 10 principal orthogonal base vectors are presented in Fig. 8. The first mode is dominant (more than 79%) and the first two modes represent more than 92% of the total energy. These results are qualitatively the same as the airfoil shape data mining results.

Figure 8. Energy ratio of the top 10 principal modes of the pressure field distribution.

Figure 9 plots the components of the eigenvector of the first four modes with respect to the index of the Pareto-optimal solutions (left) and the lift coefficient (right). This figure indicates that the pressure field of the Pareto-optimal solutions can be categorized into three groups as the result of the shape data mining, namely, low-drag designs (1 ≤ n ≤ 39), high-lift-to-drag-ratio designs (40 ≤ n ≤ 52), and high-lift designs (53 ≤ n ≤ 85). Among the low-drag designs, the components of the first and second modes increase monotonically to zero as n or C_l(n) increases. Among the high-lift-to-drag-ratio designs, the first mode increases monotonically as n or C_l(n) increases, whereas the second mode is approximately zero. Among the high-lift designs, the first mode increases monotonically as n or C_l(n) increases, whereas the second mode decreases monotonically as n or C_l(n) increases. In this figure, a large jump in the components of the eigenvectors is also observed between n = 52 and n = 53. This jump indicates a significant change in the flow field between the high-lift-to-drag-ratio designs and high-lift designs.

Figure 9. Eigenvectors of the first four modes of the pressure field distribution.
The orthogonal base vectors of the first and second modes are shown in Fig. 10. These vectors indicate that the major changes among the pressure fields of the Pareto-optimal solutions are 1) on the lower surface side near the trailing edge (region 1), 2) on the lower surface side near the leading edge (region 2), and 3) on the upper surface (region 3). These vectors also indicate that the pressure on the lower surface side near the leading and trailing edges decreases as the pressure on the upper surface side decreases.

Recalling that the pressure fields of the Pareto-optimal solutions are represented by Eqs. (1) and (2) and that the first and second modes are dominant (more than 92%), the eigenvectors (Fig. 9) and base vectors (Fig. 10) of the first and second modes and the pressure field of the maximum-lift-to-drag-ratio design (Fig. 11) provide the following observations:

1) In region 1, the second mode is dominant because the base vector of the first mode is approximately zero. Since the base vector of the second mode in region 1 is positive, the eigenvector of the second mode indicates that the high-lift-to-drag-ratio designs have the highest pressure near the trailing edge on the lower surface and that the pressure in region 1 increases monotonically as \( n \) (or lift) increases among the low-drag designs.

2) In region 2, the base vector of the first mode is negative, whereas that of the second mode is positive. In addition, the absolute value of the second mode is approximately half that of the first mode. Among the low-drag designs, the eigenvectors of the first and second modes increases monotonically as \( n \) (or lift) increases, and the absolute value of the second mode is approximately double that of the first mode, which indicates that the pressure field in region 2 does not change much among the low-drag designs because the first and second modes cancel each other. Among the high-lift-to-drag-ratio designs, the eigenvector of the first mode increases monotonically, whereas that of the second mode is approximately zero, which indicates that pressure in this region decreases as \( n \) (or lift) increases. Among the high-lift designs, the pressure in this region decreases drastically as \( n \) (or lift) increases.

3) In region 3, as in region 2, the pressure field does not change much among the low-drag designs because the first and second modes (the first and second terms of the right-hand side of Eq. (2)) approximately cancel each other. Among the high-lift-to-drag-ratio designs and high-lift designs, the pressure in region 3 increases as \( n \) (or lift) increases. The jump in the components of the eigenvectors of the first and second modes is due to strong shock wave generation on the upper surface.
VI. Conclusions

The capability of the proper-orthogonal-decomposition-based data mining approach for the analysis of the flow field data of the Pareto-optimal solutions was demonstrated. This method enables a designer to extract design knowledge by examining baseline data and a limited number of eigenvectors and orthogonal base vectors. The flow data analyzed here were the pressure field data of the Pareto-optimal solutions of an aerodynamic transonic airfoil shape optimization problem.

The results of the airfoil shape data mining and the pressure field data mining revealed that the Pareto-optimal solutions of the aerodynamic transonic airfoil shape optimization problem can be categorized into three groups: low-drag designs, high-lift-to-drag-ratio designs, and high-lift designs. For the low-drag designs, an increase in trailing edge camber contributes primarily to the pressure increase on the lower surface near the trailing edge, which leads to an increase in lift. Among the high-lift-to-drag-ratio designs, the changes in the airfoil shape and the pressure field are moderate. In the high-lift designs, the camber increases in order to decrease the pressure on the upper surface and increase lift. Interestingly, in these designs, the camber is increased by moving the lower surface upward, while the upper surface does not change much. In the high-lift designs, the airfoil thickness near the leading edge also increases as the lift increases to satisfy the given constraint on the maximum thickness near the leading edge, which results in a pressure decrease on the lower surface.

The present study reveals that the proper-orthogonal-decomposition-based data mining approach is useful for extracting design knowledge from shape data and flow field data of Pareto-optimal solutions.

Acknowledgments

The present research was supported in part by KAKENHI (20760552).

References