28th AIAA Applied Aerodynamics Conference
28 June - 1 July 2010, Chicago, Illinois

AIAA 2010-4363

Design Optimization Utilizing Gradient/Hessian Enhanced Surrogate Model

Wataru YAMAZAKI[1], Markus P. RUMPFKEIL[2] and Dimitri J. MAVRIPLIS[3]
Department of Mechanical Engineering, University of Wyoming, Laramie, Wyoming, 82072

In this paper, gradient/Hessian-enhanced surrogate models have been developed based on Kriging approaches. The gradient/Hessian-enhanced Kriging methods have been developed based on direct and indirect formulations. The efficiencies of these methods are compared by analytical function fitting, aerodynamic data modeling and 2D airfoil drag minimization problems. For the aerodynamic problems, efficient CFD gradient/Hessian calculation methods are utilized that make use of adjoint and automatic differentiation techniques. The gradient/Hessian-enhanced surrogate models are shown to be useful in the development of efficient design optimization, aerodynamic database construction, and uncertainty analysis.

I. Introduction

Numerical design optimization in the field of aerospace engineering is becoming more commonplace largely due to the maturity of Computational Fluid Dynamics (CFD) tools and the increase in computer hardware performance. Although gradient-based local optimization approaches are computationally efficient, more global optimization techniques are often required as the design space becomes more complex. The major bottleneck of aerodynamic design optimization with global search methods, such as genetic algorithms (GA), is the large number of required CFD function evaluations which can easily exceed several thousand, making these methods impractical for use with high-fidelity CFD tools.

Recently, global optimization approaches based on surrogate models have been gaining popularity as alternatives for reducing the cost of global methods. The idea of a surrogate model is to replace expensive functional evaluations (i.e. high-fidelity CFD simulations) with an analytical model which is constructed through selective sampling of the high-fidelity model. When a surrogate model is constructed with given training data, the most promising locations in the model are explored at very low computational cost. The accuracy of surrogate models can be increased efficiently by adding the exact function information in the most promising locations. This approach can save a lot of computational cost and enables the exploration of wider design space more efficiently. In Ref.[1], the performances of major surrogate models, such as least square polynomial, multi-layer perceptron, radial basis function (RBF) and Kriging, have been compared for a two dimensional turbomachinery problem. The Kriging and RBF models showed the best performance during this investigation. The Kriging model, which was originally

[1] Postdoctoral Research Associate, Dept. of Mechanical Engineering, University of Wyoming, Member AIAA
[2] Postdoctoral Research Associate, Dept. of Mechanical Engineering, University of Wyoming, Member AIAA
[3] Professor, Dept. of Mechanical Engineering, University of Wyoming, Associate Fellow AIAA

developed in the field of geological statistics, has often been found to perform well in other engineering fields and has thus gained popularity in aerospace engineering[1-12]. This surrogate model predicts the function value by using stochastic processes, and has the flexibility to represent multimodal/non-linear functions.

Because efficient gradient evaluation methods based on adjoint formulations have been developed and applied successfully in the field of aerodynamics[13-19], the introduction of gradient information within surrogate models as additional training data has also attracted attention. Two gradient-enhanced Kriging (called cokriging or GEK) approaches have been developed and beneficial results have been shown in the literature[6,9,10,12]. Additionally, efficient CFD Hessian evaluation methods have been developed recently[20], details of which are described in Subsection V.A. In this approach, the Hessian matrix of a CFD functional output with respect to design variables can be calculated efficiently by using the adjoint method and automatic differentiation (AD) tools such as TAPENADE[21]. Thus it is promising to utilize the Hessian information within surrogate models to enhance the accuracy of the surrogate model.

In addition to design optimization problems, surrogate models can be used to reduce the computational expense of creating aerodynamic database, and for uncertainty quantification[22,23]. For aerodynamic database construction, a surrogate model may be used to reduce the overall number of simulations required to fill out the entire flight envelope. With respect to uncertainty quantification, although most high-fidelity CFD computations typically assume perfect knowledge of all parameters, there is much uncertainty in reality due to manufacturing tolerances, in-service wear-and-tear, approximate modeling parameters and so on. The most straightforward and accurate method for uncertainty analysis is a full non-linear Monte-Carlo (MC) simulation. Although this method is easy to implement, it is still prohibitively expensive for high-fidelity CFD computations. Moment methods based on Taylor series expansions represent an alternative strategy to assess uncertainty, by estimating the mean and standard deviation of the output function. The use of a surrogate model is another approach for dramatically reducing the computational cost of uncertainty analysis, because the surrogate model is an analytic representation of the design space. Thus, the surrogate model can be utilized for an Inexpensive Monte-Carlo (IMC) simulation to obtain not only mean and standard deviation, but also an approximate probability density function (PDF) of the output function. The accurate response surfaces enhanced by gradient and Hessian information can increase the accuracy of uncertainty analysis based on an IMC simulation.

In this paper, gradient/Hessian-enhanced surrogate models are proposed based on Kriging approaches for efficient design optimization, aerodynamic database construction and uncertainty analysis. The efficiencies are investigated using analytical function fitting, aerodynamic data modeling and 2D airfoil drag minimization problems.

II. Kriging Model

In this section, gradient/Hessian-enhanced Kriging model approaches are proposed based on indirect/direct formulations. First, the basics of conventional Kriging, as well as direct and indirect GEK approaches are reviewed. Then, the two GEK approaches are extended to gradient/Hessian-enhanced Kriging models.

A. Conventional Kriging, direct/indirect GEK approaches

The Kriging method is a statistical prediction of a function from a set of exact function values obtained with arbitrary design variables or input parameters. Kriging prediction depends on the spatial correlations between given sample points. The correlation is given by a correlation function which is only dependent on the distance between

two points. These correlations are considered in a matrix (called correlation matrix) which appears in the Kriging formulation. The size of the matrix is $n \times n$, where n is the number of sample points. The formulation details are given in Subsection II.C for the case of a gradient/Hessian-enhanced direct Kriging model.

For a GEK model, the gradient information at several sample points is also used for the construction of the Kriging model. In the case of direct GEK, the gradient information is directly used in the modified Kriging formulation by including the correlations between function/gradient and gradient/gradient. These correlations can be modeled by differentiating the correlation function. If all sample points have function and gradient information, the size of the correlation matrix is $n(1+m)$, where m is the dimensionality of the design space (i.e. number of design variables), and n is the number of sample points, as previously.

The formulation of the indirect GEK is exactly the same as that of the original Kriging model although additional sample points are introduced using the gradient information. In this case, the additional sample points are created around a real sample point (i.e. a point where function and gradient information is evaluated) based on a first-order Taylor approximation as follows:

$$\begin{aligned} \mathbf{x}_i^{add} &= \mathbf{x}_i + \Delta\mathbf{x} \\ y_{(\mathbf{x}_i^{add})} &= y_{(\mathbf{x}_i)} + \Delta\mathbf{x}^T \left[\frac{\partial y_{(\mathbf{x}_i)}}{\partial \mathbf{x}} \right] \end{aligned} \tag{2.1}$$

where $\mathbf{x}_i$ and $y_{(\mathbf{x}_i)}$ are the location and exact function value of a real sample point, respectively. $\mathbf{x}_i^{add}$ and $y_{(\mathbf{x}_i^{add})}$ are the location and approximated function value of an additional sample point. $\Delta\mathbf{x}$ is a user-specified step size vector for the relative location of the additional sample point. Usually, one additional point is created in each direction of the design variable space as illustrated in Fig.1a (in total m additional points per real sample point). After augmentation of the sample points, the indirect GEK model is created by using both real and additional sample point information. If all sample points have function and gradient information, the total number of sample points is $n(1+m)$, which is the same as the size of the direct GEK correlation matrix. The advantage of the indirect GEK over the direct approach is the ease of implementation. On the other hand, the major disadvantage of indirect GEK is the issue associated with the step size used for creating additional sample points $\|\Delta\mathbf{x}\|$. Very small step sizes can lead to an ill-conditioned correlation matrix in the Kriging formulation due to the closely spaced sample points. The step size is usually determined by considering both the accuracy of the first-order Taylor approximation and the matrix conditioning. According to Ref.[9], direct and indirect GEK models are identical in the limit of small step size and same set of hyper parameters. However, ill-conditioning associated with small step sizes can be avoided by the direct GEK approach.

B. Gradient/Hessian-enhanced Indirect Kriging Approach

The extension of the indirect GEK approach to a gradient/Hessian-enhanced model can be achieved straightforwardly. This can be realized by modifying Eq.(2.1) to the second-order Taylor approximation as

$$\begin{aligned} \mathbf{x}_i^{add} &= \mathbf{x}_i + \Delta\mathbf{x} \\ y_{(\mathbf{x}_i^{add})} &= y_{(\mathbf{x}_i)} + \Delta\mathbf{x}^T \left[\frac{\partial y_{(\mathbf{x}_i)}}{\partial \mathbf{x}} \right] + \frac{1}{2} \Delta\mathbf{x}^T \left[\frac{\partial^2 y_{(\mathbf{x}_i)}}{\partial \mathbf{x} \partial \mathbf{x}} \right] \Delta\mathbf{x} \end{aligned} \tag{2.2}$$

There are many different ways to create additional sample points around a real sample point. Two approaches are typically used in this study. In the first approach, two additional points are created in each direction of the design variable space as illustrated in Fig.1b. In this approach, the diagonal components of the Hessian matrix are only used to create the additional sample points. If all real sample points have gradient/Hessian information, the augmented number of sample points becomes $n(1+2m)$. The second approach is more general, creating an arbitrary number of additional sample points, m_{add} by using a sampling method with a fixed distance from a real sample point as shown in Fig.1c.

As understood from the above explanation, two major user-specified parameters appear in the gradient/Hessian-enhanced indirect Kriging formulation, which are the step size $\|\Delta\mathbf{x}\|$ and the number of additional sample points per real sample point m_{add}. To specify these parameters reasonably is more difficult than in the case of the indirect GEK. The matrix conditioning becomes worse when m_{add} is large. When m_{add} is small, conversely, it is difficult to make use of the Hessian information. In these cases, the step size should be larger than that used for the indirect GEK to make use of the Hessian information in the second-order Taylor approximation. However, identifying a reasonable step size is essentially problem/location dependent. Specification of smaller or larger step sizes yields worse matrix conditioning or inaccurate Taylor approximations. That is to say, there are severe tradeoffs for these user-specified parameters and these are difficult to determine. These issues led us to develop a gradient/Hessian-enhanced direct Kriging approach, which is described in the next subsection.

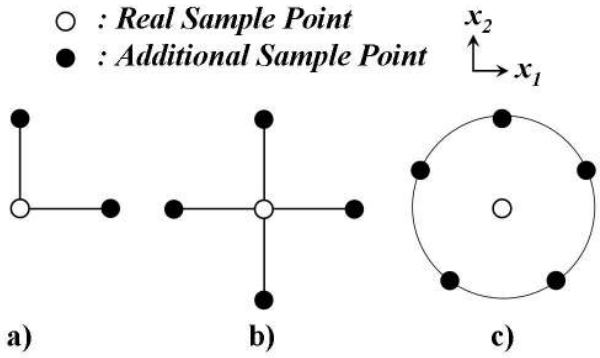


Fig.1 Arrangements of Additional Sample Points for 2D Design Space,
a) Gradient-enhanced, b) Gradient/Diagonal Hessian-enhanced, c) Gradient/Hessian-enhanced

C. Gradient/Hessian-enhanced Direct Kriging Approach

In this subsection, the gradient/Hessian-enhanced direct Kriging formulation is derived. The formulation is based on the conventional "ordinary Kriging" model for simplicity of the description. The mathematical form of a Kriging model has two terms as follows:

$$\hat{y}_{(\mathbf{x})} = \beta + Z_{(\mathbf{x})} \tag{2.3}$$

where $\hat{y}_{(\mathbf{x})}$ is the prediction of an unknown function at an arbitrary location of $\mathbf{x}$. The first term β is a constant model (or low-order polynomial regression model in the formulation of a "universal Kriging" model[7]) and

the second term $Z_{(\mathbf{x})}$ represents a random process model with zero mean, variance σ^2 and the covariance of two locations $\mathbf{x}_i$ and $\mathbf{x}_j$ is given as follows:

$$Cov\left[Z_{(\mathbf{x}_i)}, Z_{(\mathbf{x}_j)}\right] = \sigma^2 R\left(\mathbf{x}_i, \mathbf{x}_j\right) \tag{2.4}$$

where R is the spatial correlation function which is only dependent on the distance between the two locations. For the gradient/Hessian-enhanced direct Kriging approach, the following additional covariance is also required:

$$\begin{aligned}
&Cov\left[Z_{(\mathbf{x}_i)}, \frac{\partial Z_{(\mathbf{x}_j)}}{\partial x^m}\right] = \sigma^2 \frac{\partial R\left(\mathbf{x}_i, \mathbf{x}_j\right)}{\partial x_j^m} \\
&Cov\left[Z_{(\mathbf{x}_i)}, \frac{\partial^2 Z_{(\mathbf{x}_j)}}{\partial x^{m1} \partial x^{m2}}\right] = \sigma^2 \frac{\partial^2 R\left(\mathbf{x}_i, \mathbf{x}_j\right)}{\partial x_j^{m1} \partial x_j^{m2}} \\
&Cov\left[\frac{\partial Z_{(\mathbf{x}_i)}}{\partial x^k}, \frac{\partial Z_{(\mathbf{x}_j)}}{\partial x^m}\right] = \sigma^2 \frac{\partial^2 R\left(\mathbf{x}_i, \mathbf{x}_j\right)}{\partial x_i^k \partial x_j^m} \\
&Cov\left[\frac{\partial Z_{(\mathbf{x}_i)}}{\partial x^k}, \frac{\partial^2 Z_{(\mathbf{x}_j)}}{\partial x^{m1} \partial x^{m2}}\right] = \sigma^2 \frac{\partial^3 R\left(\mathbf{x}_i, \mathbf{x}_j\right)}{\partial x_i^k \partial x_j^{m1} \partial x_j^{m2}} \\
&\quad \cdots \\
&Cov\left[\frac{\partial^2 Z_{(\mathbf{x}_i)}}{\partial x^{k1} \partial x^{k2}}, \frac{\partial^2 Z_{(\mathbf{x}_j)}}{\partial x^{m1} \partial x^{m2}}\right] = \sigma^2 \frac{\partial^4 R\left(\mathbf{x}_i, \mathbf{x}_j\right)}{\partial x_i^{k1} \partial x_i^{k2} \partial x_j^{m1} \partial x_j^{m2}}
\end{aligned} \tag{2.5}$$

Suppose that there are $\left(n, n', n''\right)$ known function, gradient and Hessian components, and consider a linear combination of the exact information at given sample points as follows:

$$\hat{y}_{(\mathbf{x})} = \sum_{i=1}^{n} w_i y_i + \sum_{j=1}^{n'} \lambda_j y'_j + \sum_{k=1}^{n''} \phi_k y''_k \tag{2.6}$$

where $\left(y, y', y''\right)$ are the observed function, gradient and Hessian, and $\left(w, \lambda, \phi\right)$ are their weight coefficients. In this derivation, the numbers of function values, gradients and Hessian components can be arbitrary. Therefore, the sample points which have gradient and/or Hessian information are also arbitrary. Furthermore, users can limit arbitrarily the gradient/Hessian components which are enforced in the formulation (e.g., only the diagonal Hessian components can be enforced). The Kriging approach finds the best linear unbiased predictor which minimizes the mean square error (MSE) :

$$MSE\left[\hat{y}_{(\mathbf{x})}\right] = E\left[\sum_{i=1}^{n} w_i y_i + \sum_{j=1}^{n'} \lambda_j y'_j + \sum_{k=1}^{n''} \phi_k y''_k - y_{(\mathbf{x})}\right]^2 \tag{2.7}$$

subject to the unbiasedness constraint as follows:

$$E\left[\sum_{i=1}^{n} w_i y_i + \sum_{j=1}^{n'} \lambda_j y'_j + \sum_{k=1}^{n''} \phi_k y''_k\right] = E\left[y_{(\mathbf{x})}\right] \tag{2.8}$$

The weight coefficients can be found by solving this constrained minimization problem with the Lagrange

multiplier approach. They can be found by the following system of equations:

$$\begin{cases} \sum_{i=1}^{n} w_i Cov\left[Z_{(\mathbf{x}_i)}, Z_{(\mathbf{x}_l)}\right] + \sum_{j=1}^{n'} \lambda_j Cov\left[Z_{(\mathbf{x}_l)}, Z'_{(\mathbf{x}_j)}\right] + \sum_{k=1}^{n''} \phi_k Cov\left[Z_{(\mathbf{x}_l)}, Z''_{(\mathbf{x}_k)}\right] + \frac{1}{2}\mu \\ \qquad = Cov\left[Z_{(\mathbf{x}_l)}, Z_{(\mathbf{x})}\right] \ (l = 1, \cdots, n) \\ \sum_{i=1}^{n} w_i Cov\left[Z'_{(\mathbf{x}_i)}, Z_{(\mathbf{x}_l)}\right] + \sum_{j=1}^{n'} \lambda_j Cov\left[Z'_{(\mathbf{x}_l)}, Z'_{(\mathbf{x}_j)}\right] + \sum_{k=1}^{n''} \phi_k Cov\left[Z'_{(\mathbf{x}_l)}, Z''_{(\mathbf{x}_k)}\right] \\ \qquad = Cov\left[Z'_{(\mathbf{x}_l)}, Z_{(\mathbf{x})}\right] \ (l = 1, \cdots, n') \\ \sum_{i=1}^{n} w_i Cov\left[Z''_{(\mathbf{x}_i)}, Z_{(\mathbf{x}_l)}\right] + \sum_{j=1}^{n'} \lambda_j Cov\left[Z''_{(\mathbf{x}_l)}, Z'_{(\mathbf{x}_j)}\right] + \sum_{k=1}^{n''} \phi_k Cov\left[Z''_{(\mathbf{x}_l)}, Z''_{(\mathbf{x}_k)}\right] \\ \qquad = Cov\left[Z''_{(\mathbf{x}_l)}, Z_{(\mathbf{x})}\right] \ (l = 1, \cdots, n'') \\ \sum_{i=1}^{n} w_i = 1 \end{cases} \tag{2.9}$$

where μ is the Lagrange multiplier. Substituting Eqs.(2.4) and (2.5) in Eq.(2.9) finally yields the following matrix form:

$$\begin{bmatrix} \mathbf{R} & \mathbf{F} \\ \mathbf{F}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \widetilde{\mu} \end{bmatrix} = \begin{bmatrix} \mathbf{r} \\ 1 \end{bmatrix} \tag{2.10}$$

where $\mathbf{w} = \left(w_1, \cdots, w_n, \lambda_1, \cdots, \lambda_{n'}, \phi_1, \cdots, \phi_{n''}\right)^T \in \Re^{n+n'+n''}$, $\mathbf{F} = \left(1, \cdots 1, 0, \cdots, 0\right)^T \in \Re^{n+n'+n''}$ (n units and $n' + n''$ zeros), and $\widetilde{\mu} = \mu / \left(2\sigma^2\right)$. $\mathbf{R} \in \Re^{(n+n'+n'')\times(n+n'+n'')}$ is the correlation matrix which expresses the correlation between the observed data. $\mathbf{r} \in \Re^{n+n'+n''}$ is the correlation vector which expresses the correlation between the observed data and $\mathbf{x}$. The expressions of $\mathbf{R}$ and $\mathbf{r}$ are too complicated to present in this paper. $\mathbf{R}$ and $\mathbf{r}$ respectively include the correlation terms up to the fourth- and second-order derivatives (see Ref.[12] for the expressions of gradient-enhanced case, the expressions of our gradient/Hessian case are their straightforward extensions). The unknown weight coefficients are determined by inverting the matrix of Eq.(2.10). The final form of the gradient/Hessian-enhanced direct Kriging approach is:

$$\hat{y}_{(\mathbf{x})} = \beta + \mathbf{r}_{(\mathbf{x})}^T \mathbf{R}^{-1} \left(\mathbf{Y} - \beta \mathbf{F}\right) \tag{2.11}$$

where $\beta = \left(\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F}\right)^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{Y}$ and $\mathbf{Y} = \left(y_1, \cdots, y_n, y'_1, \cdots, y'_{n'}, y''_1, \cdots, y''_{n''}\right)^T$. The MSE of Eq.(2.7) can be expressed as:

$$MSE\left[\hat{y}_{(\mathbf{x})}\right] = s^2_{(\mathbf{x})} = \sigma^2 \left(1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} + \left(\mathbf{F}^T \mathbf{R}^{-1} \mathbf{r} - 1\right)^T \left(\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F}\right)^{-1} \left(\mathbf{F}^T \mathbf{R}^{-1} \mathbf{r} - 1\right)\right) \tag{2.12}$$

There is no sensitive parameter in the formulation of the gradient/Hessian-enhanced direct approach, although the expression is more complicated than the indirect approach. Since additional sample points are not created around real sample points, the matrix conditioning is much better than that of the indirect approach.

D. Correlation Function

The correlation matrix and vector are specified by a spatial correlation function and its derivatives. General correlation functions depend only on the distance between two locations as follows:

$$R(\mathbf{x}_i, \mathbf{x}_j) = \prod_{k=1}^{m} scf\left(\theta_k, \left|x_i^k - x_j^k\right|\right) \tag{2.13}$$

where θ_k is the hyper parameter for each direction of design variables. Gaussian or cubic spline functions are the most common forms for the spatial correlation function. The Gaussian function is defined as follows:

$$scf(\theta_k, h_k) = \exp\left(-\theta_k h_k^2\right) \qquad (\theta_k > 0) \tag{2.14}$$

One of the cubic spline functions is defined as follows:

$$scf(\theta_k, h_k) = \begin{cases} 1 - 6(h_k\theta_k)^2 + 6(h_k\theta_k)^3 & for \quad h_k \le 1/(2\theta_k) \\ 2(1 - h_k\theta_k)^3 & for \quad 1/(2\theta_k) < h_k < 1/\theta_k \\ 0 & for \quad h_k \ge 1/\theta_k \end{cases} \qquad (\theta_k > 0) \tag{2.15}$$

The hyper parameter expresses the distance weight for both spatial correlation functions. Generally, cubic spline functions yield a better conditioned correlation matrix than Gaussian functions. However, the cubic spline functions cannot be used in our practical cases since the fourth-derivatives of the correlation function are required as shown in Eq.(2.5). Therefore, the following RBF is used for the spatial correlation function in this research:

$$scf(\theta_k, h_k) = \begin{cases} \frac{1}{3}(1 - \theta_k h_k)^6 \left(35\theta_k^2 h_k^2 + 18\theta_k h_k + 3\right) & for \quad h_k \le 1/\theta_k \\ 0 & else \end{cases} \qquad (\theta_k > 0) \tag{2.16}$$

This function also yields a better conditioned correlation matrix than the Gaussian function. The derivatives up to fourth-order are calculated by employing the TAPENADE automatic differentiation tool[21].

E. Model Fitting by Maximum Likelihood Estimation

The Kriging model still includes the undetermined parameters θ and σ^2. These parameters can be estimated by maximizing the likelihood (joint probability) function of the given samples. This empirical approach finds the parameters which are most consistent with the sample data. Optimal mean and variance are analytically determined as:

$$\beta = \left(\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F}\right)^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{Y} \tag{2.17}$$

$$\sigma^2 = \frac{1}{n + n' + n''} (\mathbf{Y} - \beta\mathbf{F})^T \mathbf{R}^{-1} (\mathbf{Y} - \beta\mathbf{F}) \tag{2.18}$$

One can notice that the mean of Eq.(2.17) is equal to that of Eq.(2.11). Since there is no analytical form solution for θ, these are estimated by a numerical optimization which corresponds to the maximization of a likelihood function. For given θ, β and σ^2, the following log-likelihood function is calculated:

$$\ln(L(\theta)) = -\frac{1}{2}(n + n' + n'')\ln(\sigma^2) - \frac{1}{2}\ln(|\mathbf{R}|) \tag{2.19}$$

The most consistent θ is determined through the maximization of the log-likelihood function. One of the issues associated with the maximum likelihood estimation is the multimodality of the log-likelihood function. Therefore, some global optimization method, such as GA, is preferable to find the global optimal parameters. However, the

computational cost for this optimization problem can become relatively expensive above a certain number of sample points. This is because the evaluation of a likelihood function requires the inversion of $\mathbf{R}$ and the calculation of the determinant, and a large number of function evaluations with different sets of θ is required for a global optimization. In our approach, a master-slave type MPI-parallelized GA[24,25] is used for the optimization. The correlation matrix inversion and the determinant evaluation are solved with a Cholesky decomposition method.

F. Addition of New Sample Points by Expected Improvement

Once the Kriging model is constructed with given function/gradient/Hessian information, design optimization or uncertainty analysis can be realized on the surrogate model at low computational cost. For design optimization, it is a straightforward approach to find the approximate optimal (best $\hat{y}_{(\mathbf{x})}$) location on the surrogate model by using any type of optimizer. The surrogate model can be iteratively updated by adding new sample point information where exact function evaluation is required during the optimization process. However, this approach usually converges to a local optimal point depending on the locations of the initial sample points, because it does not take into account the uncertainty of the surrogate model.

Expected Improvement (EI)[4] expresses a potential for improvement which considers both estimated function values and uncertainties in the surrogate model. At a point $\mathbf{x}$, the EI for minimization problems is given as follows:

$$EI_{(\mathbf{x})} = \begin{cases} \left(y_{\min} - \hat{y}_{(\mathbf{x})}\right)\Phi\left(\dfrac{y_{\min} - \hat{y}_{(\mathbf{x})}}{s_{(\mathbf{x})}}\right) + s_{(\mathbf{x})}\phi\left(\dfrac{y_{\min} - \hat{y}_{(\mathbf{x})}}{s_{(\mathbf{x})}}\right) & if \;\; s_{(\mathbf{x})} > 0 \\ 0 & if \;\; s_{(\mathbf{x})} = 0 \end{cases} \tag{2.20}$$

where Φ and ϕ are the normal cumulative distribution function and probability density function, respectively. $s_{(\mathbf{x})}$ is the standard deviation expressed by Eq.(2.12). y_{min} is the minimum (optimal) function value among n given sample points. The new sample point is added in the location where the EI is maximal. This EI-based iterative approach can increase the accuracy of Kriging models by maintaining a balance between global and local search criteria. In our approach, a master-slave type MPI-parallelized GA is used for the maximization of EI.

III. Validation in Analytical Function Problems

In this section, the developed gradient/Hessian-enhanced surrogate models are used to fit analytical functions with various sets of sample points. The sets of sample points are generated by a Latin Hypercube Sampling (LHS) method[3].

A. 2D Rastrigin Function Fitting

Firstly, one of the most impressive results which make use of gradient/Hessian information within the surrogate model is shown. Surrogate models are used to fit the 2D Rastrigin function using 80 sample points generated by LHS. The gradient and Hessian components are analytically evaluated on these sample points. The Rastrigin function is defined as:

$$Rastrigin_{(\mathbf{x})} = 10m + \sum_{i=1}^{m}\left[x_i^2 - 10\cos\left(2\pi x_i\right)\right] \qquad \left(-5.12 \le x_i \le 5.12\right) \tag{3.1}$$

where m denotes the dimension (i.e. $m = 2$ in this case). The shape of the 2D Rastrigin function is illustrated

in Fig.2, which is a highly multimodal function. In Fig.3, three different surrogate models constructed by function, function/gradient and function/gradient/Hessian information are compared. The gradient and gradient/Hessian-enhanced Kriging models are constructed by the direct Kriging approach. Although the function values show agreement at the 80 sample locations, the function-based conventional Kriging model cannot capture the global trend of the Rastrigin function at all. Although the local trends of the Rastrigin function around the given sample points can be captured by including the gradient information, the gradient-enhanced Kriging model still does not capture the global trend. By including the Hessian information, on the other hand, the gradient/Hessian-enhanced Kriging model reproduces the global trend very accurately. Thus, the gradient/Hessian-enhanced model is promising for constructing an accurate surrogate model of complex multimodal functions.

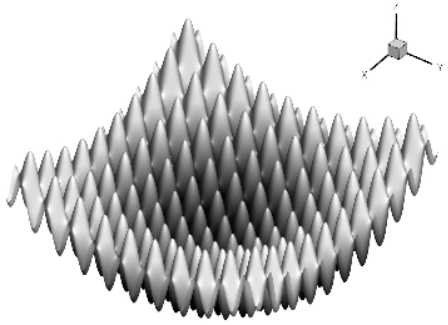


Fig.2 Visualization of 2D Rastrigin Function

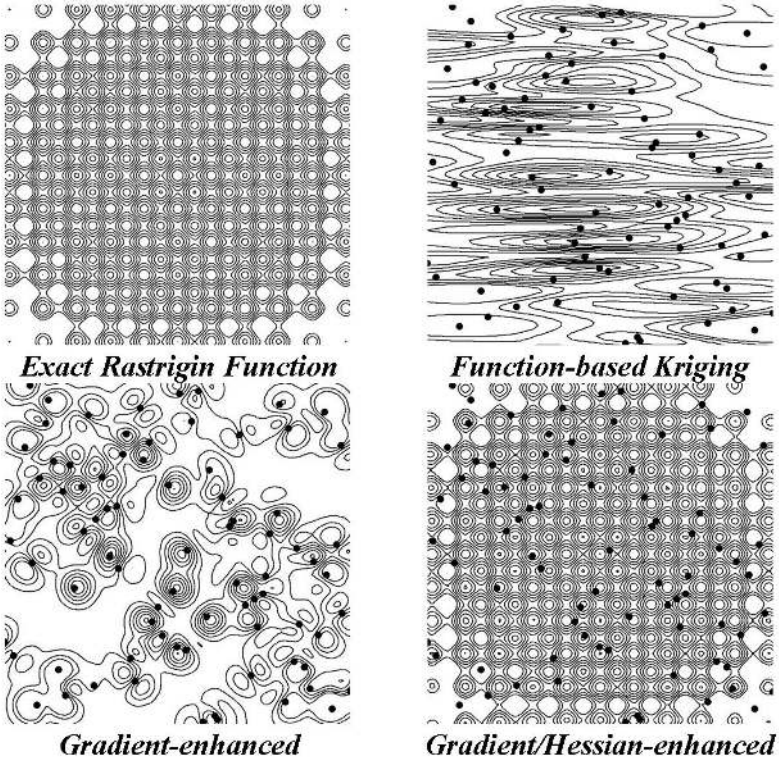


Fig.3 Kriging Models Used to Fit Rastrigin Function with 80 Sample Points

B. 5D Rosenbrock Function Fitting

For investigating the performance of the developed approaches, the developed surrogate models are used to fit the 5D Rosenbrock function by using various sets of sample points (from 10 to 600) generated by LHS. The Rosenbrock function is defined as:

$$Rosenbrock_{(\mathbf{x})} = \sum_{i=1}^{m-1}\left[(1-x_i)^2 + 100\left(x_{i+1} - x_i^2\right)^2\right] \qquad \left(-2 \le x_i \le 3\right) \tag{3.2}$$

where m denotes the dimension of the function (i.e. $m = 5$ in this case). Once the surrogate model is constructed, the accuracy of the model is evaluated by using the root mean squared error (RMSE) between the exact function and approximate function values given by the surrogate model. The RMSE is given as follows:

$$RMSE = \frac{1}{M}\sqrt{\sum_{j=1}^{M}\left(\hat{y}_{(\mathbf{x}_j)} - y_{(\mathbf{x}_j)}^{exact}\right)^2} \tag{3.3}$$

where the coordinates $\mathbf{x}_j \ \left(j = 1, \cdots, M\right)$ define an equally spaced Cartesian mesh which covers the entire design space.

In Fig.4 the RMSE values are compared between the direct/indirect Kriging approaches. For the gradient/Hessian-enhanced indirect Kriging approach, three runs are executed with different sets of $\left\|\Delta\mathbf{x}\right\|$ and m_{add}. For the gradient/Hessian-enhanced direct Kriging approach, two runs are executed for gradient/diagonal Hessian-enhanced and gradient/full Hessian-enhanced cases. Although the gradient/Hessian-enhanced indirect Kriging approaches mostly show better performance than the gradient-enhanced indirect Kriging approach, the performance is sensitive to the user-specified parameters. The direct approaches do not have any parameters to be specified, and the performance is much better than the indirect approaches. This superiority is due to the exact enforcement of derivative information and the better conditioning of the correlation matrix.

In Fig.5, the x-axis is modified to represent the number of pieces of information and the approximated computational time factor for a detailed comparison. The number of pieces of information is the sum of the number of function values, gradient as well as Hessian net components among all sample points, which is generally equal to the size of the correlation matrix. The approximated computational time factor is defined as follows, considering the gradient/Hessian computational times in our practical CFD case:

$$TF = \sum_{i=1}^{n} T_i, \quad T_i = \begin{cases} 1 & if\ i\ has \quad function\ value \\ 2 & \qquad\quad function\,/\,gradient \\ 3 & \qquad\quad function\,/\,gradient\,/\,Hessian \end{cases} \tag{3.4}$$

This definition assumes that the computational times of gradient/Hessian evaluations are respectively comparable to one function evaluation. Although this is the case when using an adjoint approach for obtaining the gradient information, the Hessian construction scales with the number of design variables or input parameters. On the other hand, our implementation parallelizes this aspect of the Hessian (see Subsection V.A) resulting in an overall wall clock timing close to that assigned above. The original function-based Kriging shows the best performance when compared by the number of pieces of information. This result implies that the scattering of only function information for the whole design space is better than the concentration of function/gradient/Hessian information at a limited number of sample points to obtain a more accurate fitting of the whole design space. When compared by

the approximated computational time factor, the gradient/Hessian-enhanced direct Kriging approaches show better performance than the original Kriging approach. This result indicates that the developed approaches can provide better performance by using our efficient gradient/Hessian evaluation methods. Since the computational times for CFD gradient evaluations are independent of the number of design variables, the gradient (and to a lesser degree Hessian) enhanced surrogate model approaches can be expected to be more effective as the dimensionality of the problem increases.

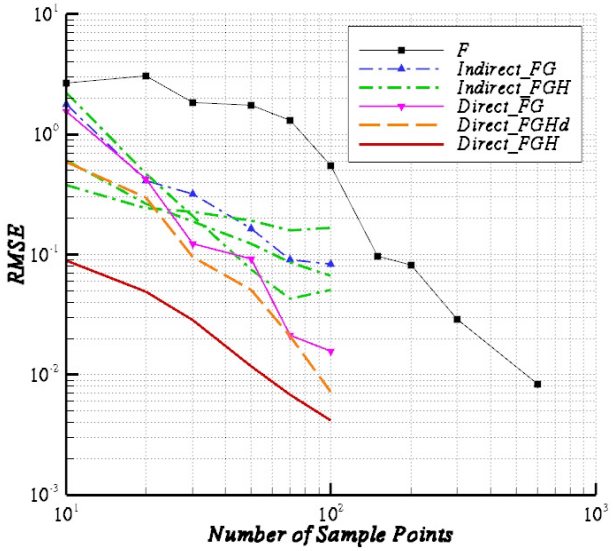


Fig.4 Comparison of RMSEs between Kriging Approaches,
(F: Function-based Original Kriging, FG: Gradient-enhanced,
FGH: Gradient/Hessian-enhanced, FGHd: Gradient/Diagonal Hessian-enhanced)

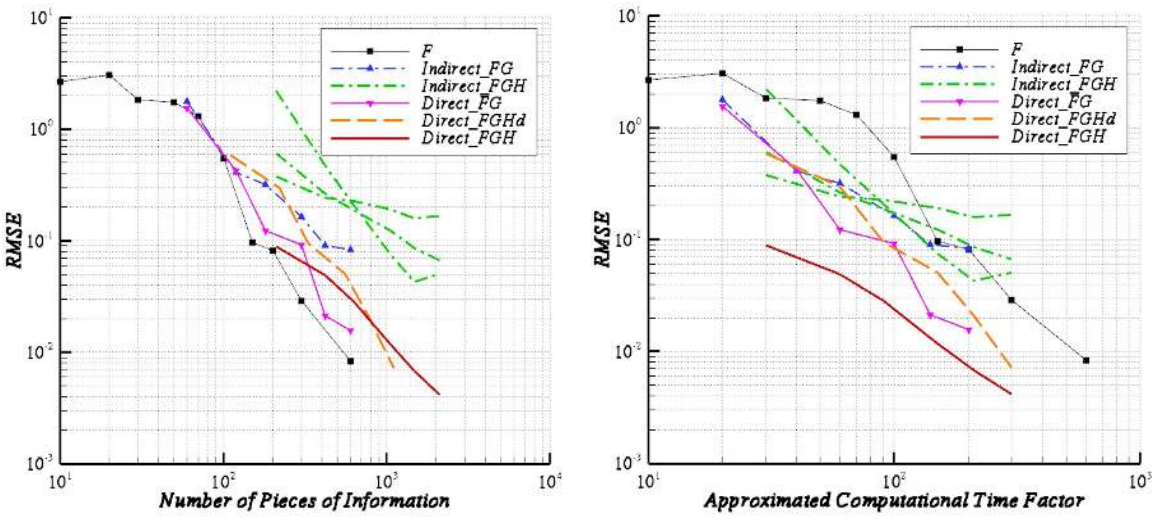
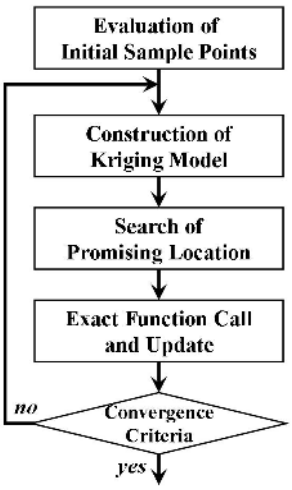


Fig.5 Comparison of RMSEs between Kriging Approaches,
Left: By Number of Pieces of Information, Right: By Approximated Computational Time Factor

C. Minimization of 20D Rosenbrock Function

In this subsection, the 20-dimensional Rosenbrock function is minimized by using a surrogate model-based optimization framework as shown in Fig.6. The most promising location is searched by maximizing the EI value. The optimization starts from 30 initial sample points that are generated by LHS. To reduce the computational cost to create a surrogate model, the number of sample points that have gradient/Hessian information is limited to the current optimal region. The optimization histories are shown in Fig.7. The gradient/Hessian-enhanced direct Kriging approach shows much faster convergence towards the global optimal than the other Kriging approaches.



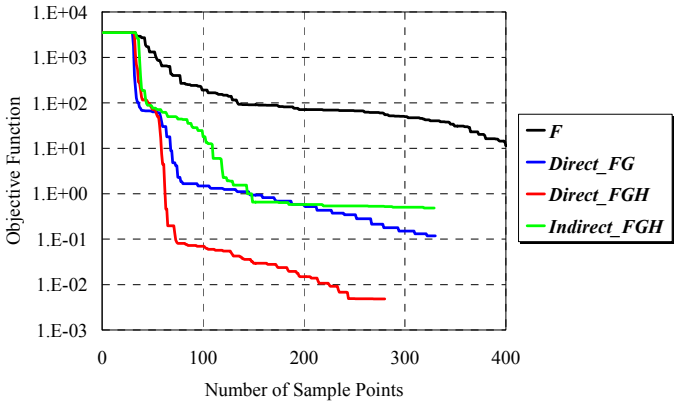


Fig.6 Surrogate Model-based Optimization Framework **Fig.7 Optimization History**

D. Uncertainty Analysis on 2D Rosenbrock Function

In this subsection, uncertainty analysis is conducted on an exact/approximated 2D Rosenbrock function. The center location for the uncertainty analysis is (1.0, 1.0), where the exact function has minimal value. The center location and the exact function distribution are shown in Fig.8. Five sample points which are generated by LHS are used to create direct Kriging models, and then inexpensive Monte-Carlo (IMC) simulations are executed on the surrogate models (Note that there is no sample point on the center location in this case). The performance of the IMC simulations is compared with a full-MC simulation and the second-order moment method. In the full-MC simulation, a large number of exact function calls is required around the center location. The moment method is based on a Taylor series expansion and makes use of derivative information at the center location. The mean and variance of a function are given by the moment method as follows[22]:

$$\mu_{MM2} = f_{(\mathbf{x}_c)} + \frac{1}{2}\sum_{i=1}^{m}\left[\left.\frac{d^2 f}{dD_i^2}\right|_{\mathbf{x}_c} \sigma_{D_i}^2\right] \tag{3.5}$$

$$\sigma_{MM2}^2 = \sum_{i=1}^{m}\left[\left.\frac{df}{dD_i}\right|_{\mathbf{x}_c} \sigma_{D_i}\right]^2 + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left[\left.\frac{d^2 f}{dD_i dD_j}\right|_{\mathbf{x}_c} \sigma_{D_i}\sigma_{D_j}\right]^2 \tag{3.6}$$

where $\mathbf{x}_c$ is the center location and σ_{D_i} is the standard deviation of design variable D_i.

In Fig.9, the variations of mean and variance with respect to the variation of σ_D are shown (a same standard deviation value was used for the all design variables). The full and inexpensive MC simulations use a total of 100,000 function calls. The second-order moment method (MM2) shows good agreement with the full-MC simulation for smaller values of σ_D since the method of MM2 uses the derivative information at the center location. For larger σ_D values, however, the discrepancy increases because the trend of the exact function deviates from the quadratic approximation. On the other hand, the result of IMC with the gradient/Hessian-enhanced Kriging model shows good agreement at both smaller and larger σ_D values. The PDF (and then cumulative density function, CDF) also can be obtained by the IMC approach. In Fig.10, the CDFs of the full-MC and IMC simulations are compared for the case of $\sigma_D = 0.15$. The CDF of IMC with the gradient/Hessian-enhanced model shows good agreement with that of the full-MC simulation.

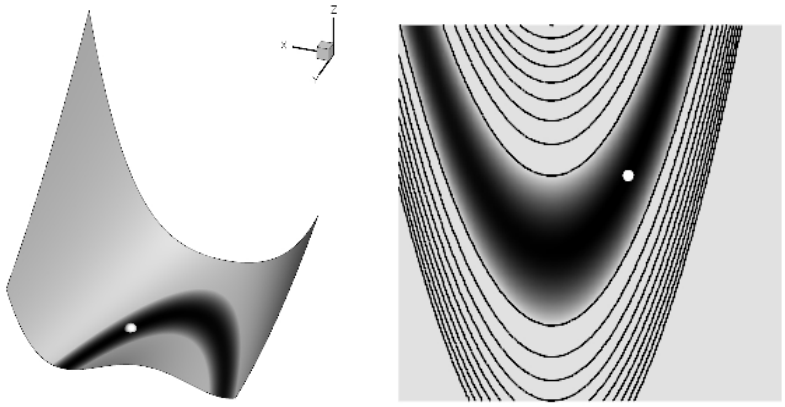


Fig.8 Visualizations of 2D Rosenbrock Function and the Center Location for Uncertainty Analysis

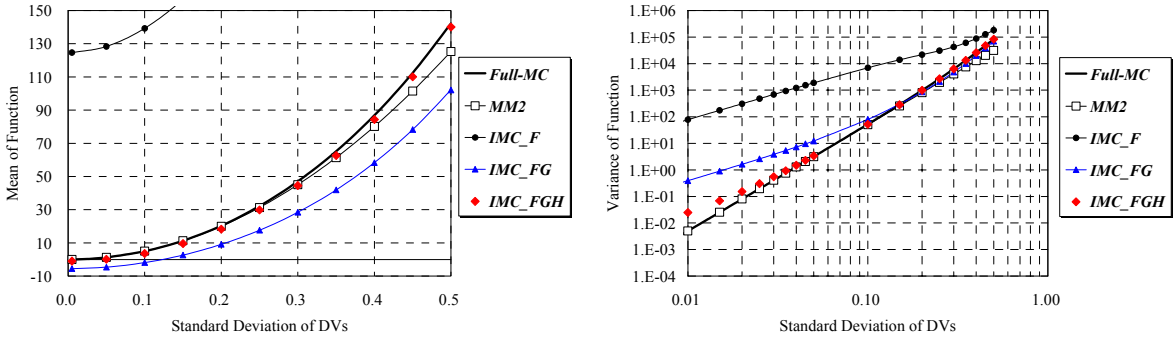


Fig.9 Variations of Mean/Variance of Function by Uncertainty Analysis

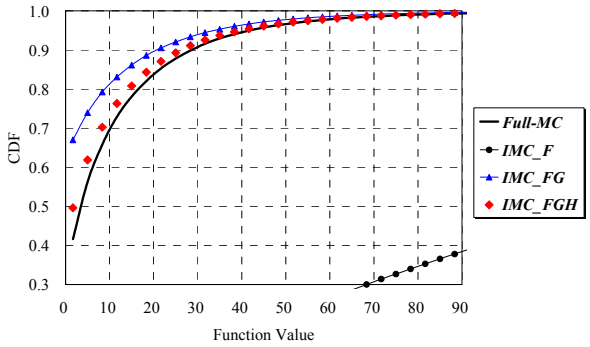


Fig.10 Comparison of CDFs obtained by Full-MC and IMC Simulations

IV. Fitting on Aerodynamic Data of 2D Airfoil

In this section the developed Kriging approaches are used to fit an airfoil two-dimensional aerodynamic database. We consider the steady inviscid flow around a NACA0012 airfoil as a flow example which is described in more detail in Mani and Mavriplis[26,27]. The governing Euler equations of the flow problem are discretized by a finite-volume approach and are solved with second-order spatial accuracy. The details of our adjoint derivative calculation method are described in Subsection V.A. The computational mesh contains approximately 20,000 triangular elements and is shown in Fig.11. In this study, two parameters are considered; Mach number and angle of attack. Their ranges are specified as $M_\infty \in [0.5,1.5]$ and $\alpha_{[\deg]} \in [0,5]$. The exact surface or validation data is obtained through a series of 21x21=441 flow computations covering the entire parameter space at equally spaced increments.

The exact hypersurfaces of lift and drag coefficients are shown in Figs.12-13. It can be observed that their behaviors are complicated at the transonic Mach numbers. Since the CFD functional outputs on this design space are relatively noisy, the CFD Hessian information is not used to construct Kriging models. The Kriging models are constructed by utilizing the function and adjoint gradient values on ten sample points, and their hypersurfaces are also illustrated in Figs.12-13. It can be understood that the adjoint derivative information is helpful to construct accurate surrogate models for these practical non-linear aerodynamic functional outputs. The discrepancies between the exact and approximate hypersufaces are sufficiently reduced by utilizing the gradient information to construct the surrogate models.

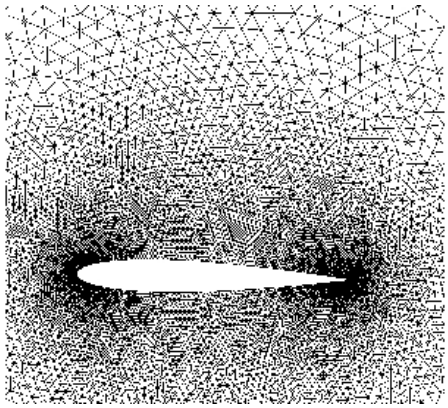


Fig. 11 Unstructured Mesh Visualization around NACA0012 Airfoil

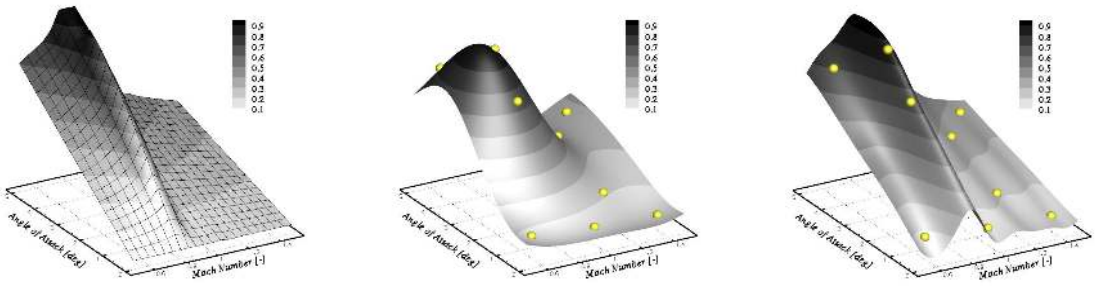


Fig.12 Hypersurfaces of Lift Coefficient,
From Left to Right: Exact, Function-based, Gradient-enhanced Kriging

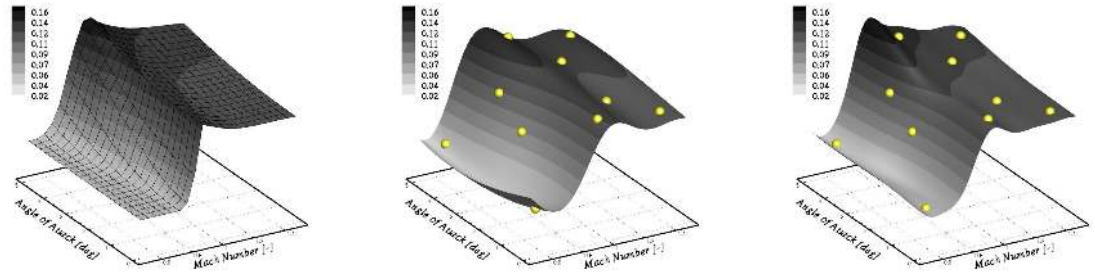


Fig.13 Hypersurfaces of Drag Coefficient,
From Left to Right: Exact, Function-based, Gradient-enhanced Kriging

V. Application to Drag Minimization Problem of 2D Airfoil

In this section, the developed gradient/Hessian-enhanced direct Kriging approach is used for a 2D airfoil shape optimization.

A. Problem Definition and CFD Methods

We consider again the steady inviscid flow around a NACA0012 airfoil. The governing Euler equations are discretized and solved in the same manner as described previously and the same computational mesh of Fig.11 is used in this study. The required deformation and movement of the mesh are performed via a linear tension spring analogy[26-28]. The free-stream Mach number is M_∞=0.755 with an angle of attack of 1.25 degrees. This flow condition gives a lift coefficient of C_l = 0.268 and a drag coefficient of C_d = 0.00521 for the NACA0012 airfoil.

The optimization example consists of an inverse design given by the following objective function:

$$F = \frac{1}{2}\left(C_l - C_l^*\right)^2 + \frac{100}{2}\left(C_d - C_d^*\right)^2 \tag{5.1}$$

where a star denotes a target lift or drag coefficient and the factor of one hundred is introduced since the drag coefficient is about an order of magnitude smaller than the lift coefficient in this particular flow example. We use 16 design variables, half of which are placed at upper and the other half at lower surface points which control the

magnitude of Hicks-Henne sine bump functions[29]. This steady inverse design problem is initialized with the NACA0012 airfoil profile and the target lift and drag coefficients are set to 0.675 and 0.0, respectively, that is, we attempt a lift constrained drag minimization problem. We must use bound constraints on the design variables to prevent the generation of invalid geometries from the mesh movement algorithm. In addition, a quadratic penalty term is added to the objective function if the sectional area of the new design is less than 90% of the initial one.

The details of our CFD gradient/Hessian evaluation methods are described in Ref.[20]. There are two options to evaluate the Hessian referred to as exact Hessian and approximate Hessian evaluation methods. For steady flow cases, the procedures to evaluate the exact Hessian are as follows:

i. The grid deformation residual equations $s(D, x(D)) = 0$ are solved

ii. The flow residual equations $R(D, x(D), w(D)) = 0$ are solved

iii. The flow adjoint problem $(\partial R/\partial w)^T \Psi + (\partial F/\partial w) = 0$ is solved

iv. The grid adjoint problem $(\partial s/\partial x)^T \lambda + (\partial F/\partial x + \Psi^T \partial R/\partial x)^T = 0$ is solved

v. m linear problems are solved each for dx/dD_i

vi. m linear problems are solved each for dw/dD_i

vii. Each Hessian element can be evaluated as $(d^2F)/(dD_i dD_j) = \Im_{ij}(F) + \Psi^T \Im_{ij}(R) + \lambda^T \partial_{ij}(s)$

where D, x and w are the design variables, grid coordinates and flow variables, respectively. Ψ and λ are respectively the flow and grid adjoint variables. $\Im_{ij}$ and ∂_{ij} indicate all possible second-order derivatives for F, R and s, and they are evaluated by employing the TAPENADE automatic differentiation tool. For a CFD gradient evaluation, only the first four steps are required. Since the computational costs of the grid deformation related steps (i, iv, v) and of the final combination (vii) are much cheaper than the others, the major computational costs are given by one flow solution, one flow adjoint solution and m linear solutions for dw/dD_i. Taking the computational cost of a flow solution as one unit, the total computational cost for a CFD exact Hessian is roughly $m+2$ units. Since there is no dependency between the third to sixth steps, they can be parallelized with $m+1$ processors as shown in Fig.14 a). Then the wall clock time for a CFD exact Hessian can be reduced to two units.

The approximate Hessian can be evaluated at reduced computational cost for inverse design objective functions of the type shown in Eq.(5.1). This type of function can be written in the form:

$$F(D) = \frac{1}{2}\sum_{k=1}^{K} W_k \left(F_k(D, x(D), w(D)) - F_k^*\right)^2 \tag{5.2}$$

where F_k, F_k^* and W_k are respectively k-th flow function value, its target value and weighting factor. The first derivatives of F with respect to D are given by

$$\frac{dF}{dD_i} = \sum_{k=1}^{K} W_k \left(\frac{dF_k}{dD_i}\right)^T \left(F_k - F_k^*\right) \tag{5.3}$$

and can be calculated inexpensively using the adjoint approach. Differentiating Eq.(5.3) again yields

$$\frac{d^2F}{dD_i dD_j} = \sum_{k=1}^{K} W_k \left(\frac{dF_k}{dD_i} \right)^T \frac{dF_k}{dD_j} + \sum_{k=1}^{K} W_k \left(F_k - F_k^* \right) \frac{d^2F_k}{dD_i dD_j}$$
$$\approx \sum_{k=1}^{K} W_k \left(\frac{dF_k}{dD_i} \right)^T \frac{dF_k}{dD_j} \tag{5.4}$$

The last approximation is true if one is close to the optimum where $F_k \approx F_k^*$ for $k = 1, \cdots, K$. Eq.(5.4) implies that the Hessian can be approximated by determining only the first derivatives dF_k / dD_i for $k = 1, \cdots, K$ where one can use the adjoint to determine each derivative. The total computational cost for an approximate CFD Hessian is $K+1$ units, which is independent of the number of design variables. Again, all adjoint calculations can be parallelized with K processors as shown in Fig.14 b). The accuracies of the exact Hessian at any location and the approximate Hessian close to the optimum location have been confirmed by comparing the results of a finite differenced Hessian[20].

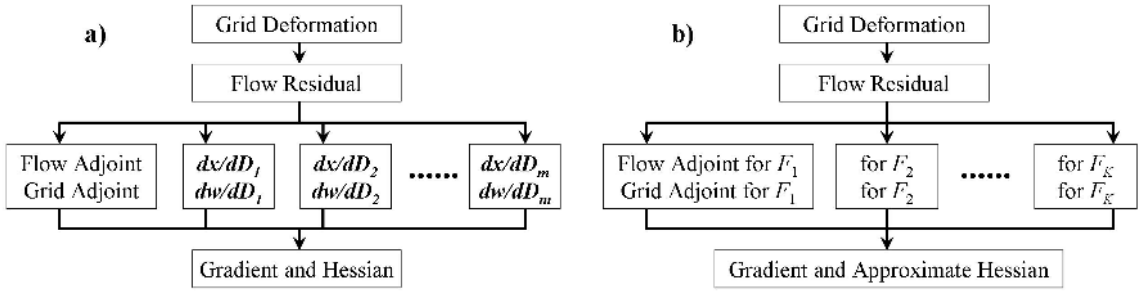


Fig.14 Flowcharts of Hessian Evaluation Approaches,
a) Exact CFD Hessian, b) Approximate CFD Hessian

B. Optimization Results

The direct Kriging approaches are used for this optimization. The optimization histories are shown in Fig.15. For these optimizations, a set of 16 initial sample points has been chosen by the LHS method. The gradient and Hessian have not been evaluated for the initial samples. Therefore, the same surrogate model is firstly constructed for all available cases. When constructing the surrogate model, new sample points are added which are chosen by maximizing the EI value. For the cases utilizing gradient and/or Hessian-enhanced surrogate models, the gradient/Hessian evaluations are invoked only when a new design has better performance than all previous designs. In addition, the number of sample points that have gradient/Hessian information is limited to the current optimal region to reduce the computational cost to construct a surrogate model.

It was confirmed that the gradient and/or Hessian-enhanced model approaches delivered better optimization histories than the conventional function-based Kriging approach. However, it was difficult to observe the superiority of gradient/Hessian-enhanced approaches over the gradient-enhanced approach. This was because the deformable range of the Hicks-Henne bump functions, that is the design space, was fairly limited to prevent negative thickness airfoils, and to ensure convergence and accuracy of the Hessian evaluation. Actually, several particular design variables of optimal designs were located on the bound values. Another reason is the presumably

simple objective functional shape due to the inviscid flow model, resulting in similar performance for both gradient- and gradient/Hessian-enhanced approaches. However, the absolute best design was obtained still by the gradient/exact Hessian-enhanced model approach. In Fig.16, the flowfields of the initial and optimal design given by the gradient/exact Hessian-enhanced approach are compared. It can be confirmed that the optimal design results in a supercritical airfoil shape which increases lift while reducing the strength of the shock wave on the upper surface of the airfoil.

Faster convergence towards the global optimal direction, i.e. the improvement of the search efficiency, is considered to be one of the major benefits of derivative-enhanced surrogate model approaches. These approaches should yield more beneficial results in higher-dimensions and/or for more complex physical models.

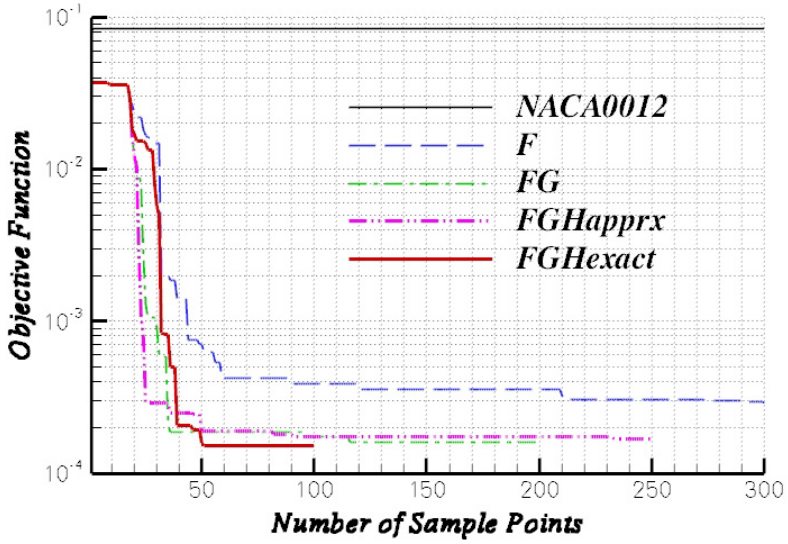


Fig.15 Optimization History

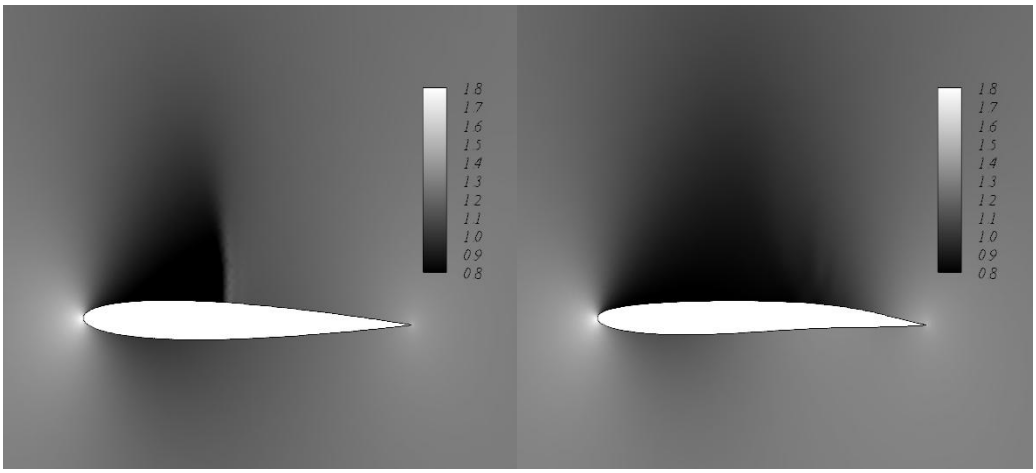


Fig.16 Pressure Distributions around 2D Airfoil Shapes, Left: Initial, Right: Optimal

C. Uncertainty Analysis at the Optimal Location

Subsequent to the optimization procedure, an uncertainty analysis is executed at the optimal location obtained by the gradient/exact Hessian-enhanced approach. The results are compared between a full Non-Linear Monte-Carlo simulation (NLMC), a second-order moment method (MM2), and IMC simulations (IMC1, IMC2). For NLMC, a total of 3000 non-linear flow evaluations per specified standard deviation of all design variables (σ_D) is executed around the center (optimal) location. The gradient and exact Hessian information at the center location is used for the MM2 approach. For the IMC1 method, the final gradient/exact Hessian-enhanced Kriging model given in the optimization loop is used for the IMC simulation. For the IMC2 method, function/gradient/Hessian information is evaluated on 10 different sample points which are created by using a sampling method for $\sigma_D = 0.01$. The gradient/Hessian-enhanced surrogate model for IMC2 is then created by using the information at 11 points (center and the 10 sample points). The number of (approximate) function calls for IMC is also 3000, and the computational cost of IMC is much cheaper than that of NLMC. The computational time of a NLMC simulation is over 6 days using one node (four cores) of our cluster (of course the computational time can be reduced by parallelizing the 3000 non-linear flow evaluations). On the other hand, the computational time for IMC simulation is less than one minute.

In Fig.17, the variations in the mean of the objective function with respect to the variation of σ_D are compared. For an intuitive understanding, the mean airfoil (optimal airfoil) and the airfoils resulting from perturbations of ± 0.1 for all design variables are shown in Fig.18. The IMC results showed good agreements with the NLMC results for smaller values of σ_D while the discrepancies increased for larger σ_D values. The IMC2 method showed the worst performance for larger σ_D values since the 10 different sample points were generated for $\sigma_D = 0.01$. For IMC1, there are two major reasons for the discrepancy at larger σ_D values. Firstly, the sample point locations were decided by the EI maximization criteria in the optimization procedure, that did not take into account the total model accuracy. Therefore, the distribution of sample points was biased towards the optimization purpose. Secondly, several optimized design variables were located on the bound values. Therefore, various of the function calls for the IMC simulation were at the outside of design space where there was no sample point. Although the Kriging model can predict function values even the outside of the design space, this essentially corresponds to extrapolation and is less accurate than interpolation.

In Fig.19, the CDFs of the objective function at $\sigma_D = 0.01$ are compared. The CDF of a quadratic surrogate model constructed by using the gradient/Hessian information at the center location is also included in Fig.19. Although the CDFs of the IMC approaches showed slightly better agreement with the NLMC results than the quadratic model, the accuracy difference is not justified by the additional complexity and cost. In this case, additional sample criteria will be essential to achieve an efficient design optimization and accurate uncertainty analysis simultaneously.

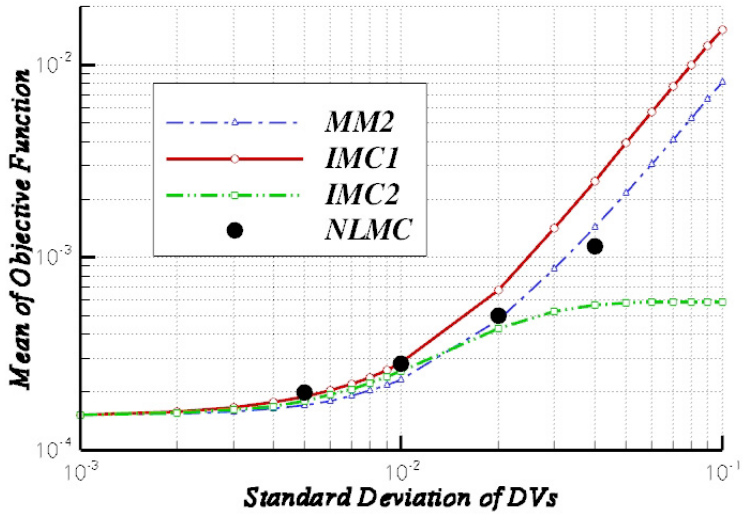


Fig.17 Variations of Mean of Objective Function by Uncertainty Analysis

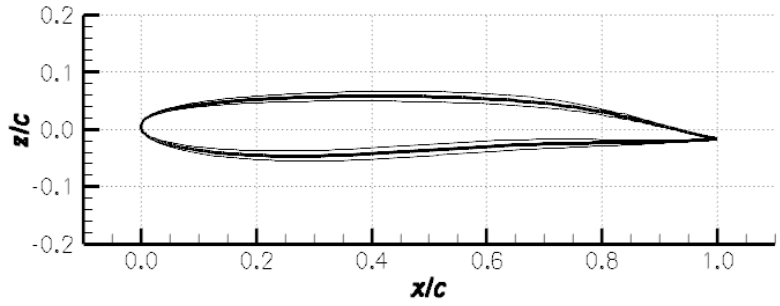


Fig.18 Perturbed Airfoils (thin) from the Mean Airfoil (thick)

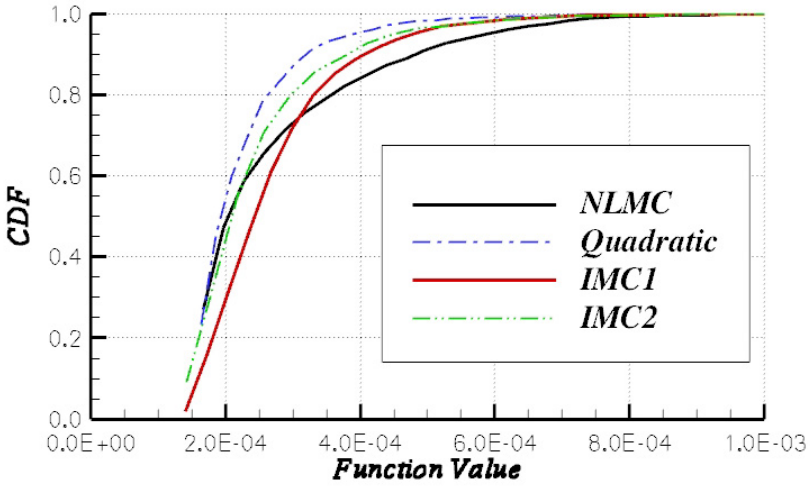


Fig.19 Comparison of CDFs at $\sigma_D = 0.01$

VI. Conclusion

In this paper, gradient/Hessian-enhanced surrogate models have been developed based on direct/indirect Kriging approaches. The direct approaches showed better performance than the indirect approaches, thanks to the exact enforcement of derivative information and better matrix conditioning. Firstly, analytical function fitting problems were investigated. The accuracy of surrogate models was sufficiently increased by adding gradient/Hessian information in the Kriging formulation. It was demonstrated that the gradient/Hessian-enhanced surrogate models were the most accurate and computationally least expensive even when taking into account the computational costs of gradient/Hessian evaluations. Design optimization and uncertainty analysis on analytical functions also showed the superiority of the developed gradient/Hessian-enhanced direct Kriging approach as well as IMC simulation utilizing the surrogate model.

Then, the developed Kriging model approaches were used to fit aerodynamic functions of a 2D airfoil in which Mach number and angle of attack were utilized as the parameters. It was confirmed that adjoint derivative information was helpful to construct accurate surrogate models for the practical non-linear aerodynamic functions.

Finally, a drag minimization problem was solved by combining the developed surrogate models with the efficient CFD gradient/Hessian evaluation methods. One could observe a faster reduction of the objective function value for gradient/Hessian-enhanced model approaches which will be more beneficial in higher dimensions and/or for more complex physical models. The uncertainty analysis at the obtained optimal location was also conducted and showed reasonable results compared to full non-linear Monte-Carlo simulations at much lower cost.

Thus, the developed gradient/Hessian-enhanced direct Kriging approach is promising for efficient design optimization, aerodynamic database construction and uncertainty analysis problems. The developed methods will be further applied and validated in more practical and higher dimensional problems.

Acknowledgements

The funding of the first author by the Japan Society for the Promotion of Science is gratefully acknowledged. We are also very grateful to Karthik Mani for making his flow and adjoint solver available to us.

References

[1]Peter, J., and Marcelet, M., "Comparison of Surrogate Models for Turbomachinery Design," *WSEAS Transactions on Fluid Mechanics*, Issue.1, Vol.3, 2008.

[2]Cressie, N., "The Origins of Kriging," *Mathematical Geology*, Vol.22, No.3, pp.239-252, 1990.

[3]Koehler, J. R., and Owen, A. B., "Computer Experiments," In Ghosh, S., Rao, C.R., (Eds.), *Handbook of Statistics*, pp.261-308, 1996.

[4]Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol.13, pp.455–492, 1998.

[5]Simpson, T. W., Korte, J. J., Mauery, T. M., and Mistree, F., "Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization," AIAA-98-4755, 1998.

[6]Chung, H. S., and Alonso, J. J., "Using Gradients to Construct Cokriging Approximation Models for High-Dimensional Design Optimization Problems," AIAA-2002–0317, 2002.

[7]Martin, J. D., and Simpson, T. W., "Use of Kriging Models to Approximate Deterministic Computer Models," *AIAA Journal*, Vol.43, No.4, pp.853-863, 2005.

[8]Jeong, S., Murayama, M., and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol.42, No.2, pp.413-420, 2005.

[9]Laurenceau, J., and Sagaut, P., "Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging," *AIAA Journal*, Vol.46, No.2, pp.498-507, 2008.

[10]Laurenceau, J., and Meaux, M., "Comparison of Gradient and Response Surface Based Optimization Frameworks Using Adjoint Method," AIAA-2008-1889, 2008.

[11]Yamazaki, W., Mouton, S., and Carrier, G., "Efficient Design Optimization by Physics-Based Direct Manipulation Free-Form Deformation," AIAA-2008-5953, 2008.

[12] Han, Z. H., Görtz, S., and Zimmermann, R., "On Improving Efficiency and Accuracy of Variable-Fidelity Surrogate Modeling in Aero-data for Loads Context," CEAS 2009 European Air and Space Conference, 2009.

[13]Jameson, A., "Efficient Aerodynamic Shape Optimization," AIAA Paper 2004-4369, 2004.

[14]Mavriplis, D. J., "A Discrete Adjoint-Based Approach for Optimization Problems on Three-Dimensional Unstructured Meshes," AIAA Paper 2006-0050, 2006.

[15]Kim, H. J., Sasaki, D., Obayashi, S., and Nakahashi, K., "Aerodynamic Optimization of Supersonic Transport Wing Using Unstructured Adjoint Method," *AIAA Journal*, Vol.39, No.6, pp.1011-1020, 2001.

[16]Nielsen, E. J. and Anderson, W. K., "Recent Improvements in Aerodynamic Optimization on Unstructured Meshes," *AIAA Journal*, Vol.40, No.6, pp.1155–1163, 2002.

[17]Giles, M. B., Duta, M. C., Muller, J. D., and Pierce, N. A., "Algorithm Developments for Discrete Adjoint Methods," *AIAA Journal*, Vol.41, No.2, pp.198-205, 2003.

[18]Nielsen, E. J., and Park, M. A., "Using an Adjoint Approach to Eliminate Mesh Sensitivities in Computational Design," *AIAA Journal*, Vol.44, No.5, pp.948-953, 2006.

[19]Peter, J., and Mayeur, M., "Improving Accuracy and Robustness of a Discrete Direct Differentiation Method and Discrete Adjoint Method for Aerodynamic Shape Optimization," Proceedings of the ECCOMAS CFD Conference, 2006.

[20]Rumpfkeil, M. P., and Mavriplis, D. J., "Efficient Hessian Calculations using Automatic Differentiation and the Adjoint Method," AIAA-2010-1268, 2010.

[21]Hascoët, L., "TAPENADE: A Tool for Automatic Differentiation of Programs," Proceedings of the ECCOMAS Conference, Jyvaskyla, Finland, 2004.

[22]Ghate, D., and Giles, M. B., "Inexpensive Monte Carlo Uncertainty Analysis," *Recent Trends in Aerospace Design and Optimization*, *Tata McGraw-Hill, New Delhi*, pp.203-210, 2006.

[23]Chalot, F., Dinh, Q. V., Herbin, E., Martin, L., Ravachol, M., and Roge, G., "Estimation of the Impact of Geometrical Uncertainties on Aerodynamic Coefficients Using CFD," AIAA-2008-2068, 2008.

[24]Fonseca, C. M., and Fleming, P. J., "Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization," Proceedings of the 5th International Conference on Genetic Algorithms, Morgan Kaufmann Publishers, Inc., San Mateo, pp.416-423, 1993.

[25]Kim, H. J., and Liou, M.-S., "New Multi-Objective Genetic Algorithms for Diversity and Convergence Enhancement," AIAA Paper 2009-1168, 2009.

[26]Mani, K., and Mavriplis, D. J., "An Unsteady Discrete Adjoint Formulation for Two-Dimensional Flow Problems with Deforming Meshes," AIAA-2007-60, 2007.

[27]Mani, K., and Mavriplis, D. J., "Unsteady Discrete Adjoint Formulation for Two-Dimensional Flow Problems

with Deforming Meshes," *AIAA Journal*, Vol.46, No.6, pp.1351-1364, 2008.

[28]Batina, J. T., "Unsteady Euler Airfoil Solutions Using Unstructured Dynamic Meshes," *AIAA Journal*, Vol.28, No.8, pp.1381-1388, 1990.

[29]Hicks, R. M., and Henne, P. A., "Wing Design by Numerical Optimization," *Journal of Aircraft*, Vol.15, No.7, pp.407-412, 1978.