

Surrogate-Based Optimization

Zhong-Hua Han and Ke-Shi Zhang

*School of Aeronautics, Northwestern Polytechnical University, Xi'an,
P.R. China*

1. Introduction

Surrogate-based optimization (Queipo et al. 2005, Simpson et al. 2008) represents a class of optimization methodologies that make use of surrogate modeling techniques to quickly find the local or global optima. It provides us a novel optimization framework in which the conventional optimization algorithms, e.g. gradient-based or evolutionary algorithms are used for sub-optimization(s). Surrogate modeling techniques are of particular interest for engineering design when high-fidelity, thus expensive analysis codes (e.g. Computation Fluid Dynamics (CFD) or Computational Structural Dynamics (CSD)) are used. They can be used to greatly improve the design efficiency and be very helpful in finding global optima, filtering numerical noise, realizing parallel design optimization and integrating simulation codes of different disciplines into a process chain. Here the term “surrogate model” has the same meaning as “response surface model”, “metamodel”, “approximation model”, “emulator” etc. This chapter aims to give an overview of existing surrogate modeling techniques and issues about how to use them for optimization.

2. Overview of surrogate modeling techniques

For optimization problems, surrogate models can be regarded as approximation models for the cost function (s) and state function (s), which are built from sampled data obtained by randomly probing the design space (called sampling via Design of Experiment (DoE)). Once the surrogate models are built, an optimization algorithm such as Genetic Algorithms (GA) can be used to search the new design (based on the surrogate models) that is most likely to be the optimum. Since the prediction with a surrogate model is generally much more efficient than that with a numerical analysis code, the computational cost associated with the search based on the surrogate models is generally negligible.

Surrogate modeling is referred to as a technique that makes use of the sampled data (observed by running the computer code) to build surrogate models, which are sufficient to predict the output of an expensive computer code at untried points in the design space. Thus, how to choose sample points, how to build surrogate models, and how to evaluate the accuracy of surrogate models are key issues for surrogate modeling.

2.1 Design of experiments

To build a surrogate model, DoE methods are usually used to determine the locations of sample points in the design space. DoE is a procedure with the general goal of maximizing

the amount of information gained from a limited number of sample points (Giunta et al., 2001). Currently, there are different DoE methods which can be classified into two categories: "classic" DoE methods and "modern" DoE methods. The classic DoE methods, such as full-factorial design, central composite design (CCD), Box-Behnken and D-Optimal Design (DOD), were developed for the arrangement of laboratory experiments, with the consideration of reducing the effect of random error. In contrast, the modern DoE methods such as Latin Hypercube Sampling (LHS), Orthogonal Array Design (OAD) and Uniform Design (UD) (Fang et al., 2000) were developed for deterministic computer experiments without the random error as arises in laboratory experiments. An overview of the classic and modern DoE methods was presented by Giunta et al. (2001). A more detailed description of existing DoE methods is beyond the scope of this chapter.

The schematics of 40 sample points selected by LHS and UD for a two-dimensional problem are sketched in Figure 1.

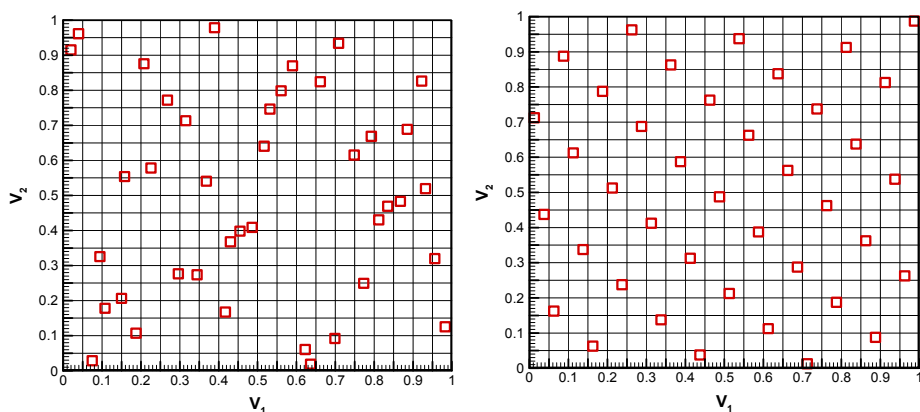


Fig. 1. Schematics of 40 sample points selected by Design of Experiments for a two-dimensional problem (left: Latin Hypercube Sampling; right: Uniform Design).

2.2 Surrogate models

There are a number of surrogate models available in the literatures. Here we limit our discussion to three popular techniques such as RSM (polynomial Response Surface Model), Kriging, RBFs (Radial Basis Functions).

For an m -dimensional problem, suppose we are concerned with the prediction of the output of a high-fidelity, thus expensive computer code, which is correspondent to an unknown function $y: \mathbb{R}^m \rightarrow \mathbb{R}$. By running the computer code, y is observed at n sites (determined by DoE)

$$\mathbf{S} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}]^T \in \mathbb{R}^{n \times m}, \mathbf{x} = \{x_1, \dots, x_m\} \in \mathbb{R}^m \quad (1)$$

with the corresponding responses

$$\mathbf{y}_S = [y^{(1)}, \dots, y^{(n)}]^T = [y(\mathbf{x}^{(1)}), \dots, y(\mathbf{x}^{(n)})]^T \in \mathbb{R}^n \quad (2)$$

The pair $(\mathbf{S}, \mathbf{y}_s)$ denotes the sampled data sets in the vector space.

With the above descriptions and assumptions, our objective here is to build a surrogate model for predicting the output of the computer code for any untried site \mathbf{x} (that is, to estimate $y(\mathbf{x})$) based on the sampled data sets $(\mathbf{S}, \mathbf{y}_s)$, in an attempt to achieve the desired accuracy with the least possible number of sample points.

2.2.1 Quadratic response surface method

Here we use “RSM” to denote a polynomial approximation model in which the sampled data is fitted by a least-square regression technique. In RSM-based optimization applications, the “quadratic” polynomial model usually provides the best compromise between the modeling accuracy and computational expense, when compared with the linear or higher order polynomial models. An advantage of RSM is that it can smooth out the various scales of numerical noise in the data while captures the global trend of the variation, which makes it very robust and thus well suited for optimization problems in engineering design.

The true quadratic RSM can be written in the following form:

$$y(\mathbf{x}) = \hat{y}(\mathbf{x}) + \varepsilon, \mathbf{x} \in \mathbb{R}^m, \quad (3)$$

where $\hat{y}(\mathbf{x})$ is the quadratic polynomial approximation and ε is the random error which is assumed to be normally distributed with mean zero and variance of σ^2 . The error ε_i at each observation is supposed to be independent and identically distributed. The quadratic RSM predictor $\hat{y}(\mathbf{x})$ can be defined as:

$$\hat{y}(\mathbf{x}) = \beta_0 + \sum_{i=1}^m \beta_i x_i + \sum_{i=1}^m \beta_{ii} x_i^2 + \sum_{i=1}^m \sum_{j \geq i}^m \beta_{ij} x_i x_j, \quad (4)$$

where β_0 , β_i , β_{ii} and β_{ij} are unknown coefficients to be determined. Since there are totally $p = (m+1)(m+2)/2$ unknown coefficients in Eq.(4), building a quadratic RSM with m variables requires at least p sample points. Let $\boldsymbol{\beta} \in \mathbb{R}^p$ be the column vector contains these p unknown coefficients. The least square estimator of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \mathbf{y}_s, \quad (5)$$

where

$$\mathbf{U} = \begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_m^{(1)} & x_1^{(1)} x_2^{(1)} & \cdots & x_{m-1}^{(1)} x_m^{(1)} & (x_1^{(1)})^2 & \cdots & (x_m^{(1)})^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(n)} & \cdots & x_m^{(n)} & x_1^{(n)} x_2^{(n)} & \cdots & x_{m-1}^{(n)} x_m^{(n)} & (x_1^{(n)})^2 & \cdots & (x_m^{(n)})^2 \end{bmatrix} \in \mathbb{R}^{n \times p}. \quad (6)$$

After the unknown coefficients in $\boldsymbol{\beta}$ are determined, the approximated response \hat{y} at any untried \mathbf{x} can be efficiently predicted by Eq. (4).

2.2.2 Kriging model

Different from RSM, Kriging (Krige, 1951) is an interpolating method which features the observed data at all sample points. Kriging provides a statistic prediction of an unknown function by minimizing its Mean Squared Error (MSE). It can be equivalent to any order of polynomials and is thus well suited for a highly-nonlinear function with multi extremes. For the derivation of Kriging (Sacks et al., 1989), the output of a deterministic computer experiment is treated as a realization of a random function (or stochastic process), which is defined as the sum of a global trend function $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ and a Gaussian random function $Z(\mathbf{x})$ as following

$$y(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}), \mathbf{x} \in \mathbb{R}^m, \quad (7)$$

where $\mathbf{f}(\mathbf{x}) = [f_0(\mathbf{x}), \dots, f_{p-1}(\mathbf{x})]^T \in \mathbb{R}^p$ is defined with a set of the regression basis functions and $\boldsymbol{\beta} = [\beta_0, \dots, \beta_{p-1}]^T \in \mathbb{R}^p$ denotes the vector of the corresponding coefficients. In general, $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ is taken as either a constant or low-order polynomials. Practice suggests that the constant trend function is sufficient for most of the problems. Thus, $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ is taken as a constant β_0 in the text hereafter. In Eq.(7), $Z(\cdot)$ denotes a stationary random process with zero mean, variance σ^2 and nonzero covariance of

$$\text{Cov}[Z(\mathbf{x}), Z(\mathbf{x}')] = \sigma^2 R(\mathbf{x}, \mathbf{x}'). \quad (8)$$

Here $R(\mathbf{x}, \mathbf{x}')$ is the correlation function which is only dependent on the Euclidean distance between any two sites \mathbf{x} and \mathbf{x}' in the design space. In this study, a Gaussian exponential correlation function is adopted, and it is of the form

$$R(\mathbf{x}, \mathbf{x}') = \exp\left[-\sum_{k=1}^m \theta_k |x_k - x'_k|^{p_k}\right], 1 < p_k \leq 2, \quad (9)$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_m]^T$ and $\mathbf{p} = [p_1, p_2, \dots, p_m]^T$ denote the vectors of the unknown model parameters (hyper parameters) to be tuned. The schematics of a Gaussian exponential correlation function for one-dimensional problem is sketched in Figure 2.

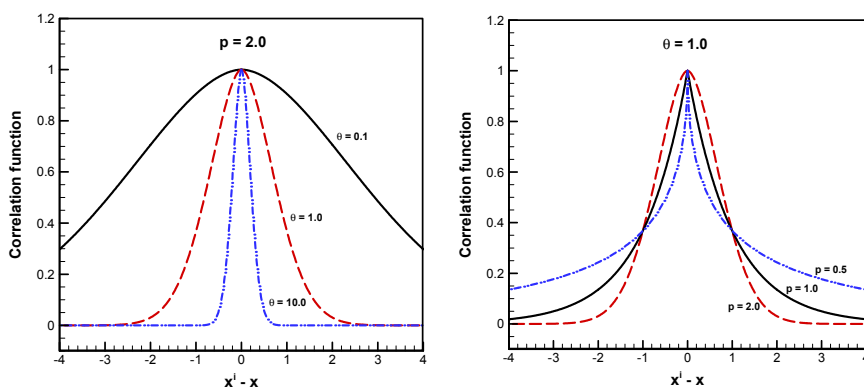


Fig. 2. Schematics of Gaussian exponential correlation function for different hyper parameters (left: varying θ with a fixed p ; right: varying p with a fixed θ)

From the derivation by Sacks et al. (1989) the Kriging predictor $\hat{y}(x)$ for any untried \mathbf{x} can be written as

$$\hat{y}(\mathbf{x}) = \beta_0 + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y}_S - \beta_0\mathbf{1}), \quad (10)$$

where the generalized least square estimation of β_0 is

$$\beta_0 = (\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}_S, \quad (11)$$

and $\mathbf{1} \in \mathbb{R}^n$ is a vector filled with ones, and \mathbf{R}, \mathbf{r} are the correlation matrix and the correlation vector, respectively. \mathbf{R} and \mathbf{r} are defined as

$$\mathbf{R} = \begin{bmatrix} R(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) & R(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) & \dots & R(\mathbf{x}^{(1)}, \mathbf{x}^{(n)}) \\ R(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) & R(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) & \dots & R(\mathbf{x}^{(2)}, \mathbf{x}^{(n)}) \\ \vdots & \vdots & \ddots & \vdots \\ R(\mathbf{x}^{(n)}, \mathbf{x}^{(1)}) & R(\mathbf{x}^{(n)}, \mathbf{x}^{(2)}) & \dots & R(\mathbf{x}^{(n)}, \mathbf{x}^{(n)}) \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad \mathbf{r} = \begin{bmatrix} R(\mathbf{x}^{(1)}, \mathbf{x}) \\ R(\mathbf{x}^{(2)}, \mathbf{x}) \\ \vdots \\ R(\mathbf{x}^{(n)}, \mathbf{x}) \end{bmatrix} \in \mathbb{R}^n, \quad (12)$$

where $R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ denotes the correlation between any two observed points $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$; $R(\mathbf{x}^{(i)}, \mathbf{x})$ denotes the correlation between the i -th observed point $\mathbf{x}^{(i)}$ and the untried point \mathbf{x} .

A unique feature of Kriging model is that it provides an uncertainty estimation (or MSE) for the prediction, which is very useful for sample-points refinement. It is of the form

$$\hat{\sigma}^2(\mathbf{x}) = \sigma^2 [1.0 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} + (\mathbf{r}^T \mathbf{R}^{-1} \mathbf{1} - 1)^2 / \mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}]. \quad (13)$$

Assuming that the sampled data are distributed according to a Gaussian process, the responses at sampling sites are considered to be correlated random functions with the corresponding likelihood function given by

$$L(\beta_0, \sigma^2, \boldsymbol{\theta}, \mathbf{p}) = \frac{1}{\sqrt{2\pi(\sigma^2)^n |\mathbf{R}|}} \exp \left(-\frac{1}{2} \frac{(\mathbf{y}_S - \beta_0 \mathbf{1})^T \mathbf{R}^{-1} (\mathbf{y}_S - \beta_0 \mathbf{1})}{\sigma^2} \right). \quad (14)$$

The optimal estimation of β_0 and the process variance

$$\begin{aligned} \beta_0(\boldsymbol{\theta}, \mathbf{p}) &= (\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}_S \\ \sigma^2(\beta_0, \boldsymbol{\theta}, \mathbf{p}) &= \frac{1}{n} (\mathbf{y}_S - \beta_0 \mathbf{1})^T \mathbf{R}^{-1} (\mathbf{y}_S - \beta_0 \mathbf{1}) \end{aligned} \quad (15)$$

are obtained analytically, yet depends on the remaining hyper-parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_m]^T$ and $\mathbf{p} = [p_1, p_2, \dots, p_m]^T$. Substituting it into the associated Eq. (14) and taking the logarithm, we are left with maximizing

$$. \text{MLE}(\boldsymbol{\theta}, \mathbf{p}) = -n \ln \sigma^2(\boldsymbol{\theta}) - \ln |\mathbf{R}(\boldsymbol{\theta})|, \quad (16)$$

which can be solved by a numerical optimization algorithm such as GA. The hyperparameters tuning strategies are discussed by Toal et al. (2008). Note that the above Kriging formulation can be extended by including gradient information obtained by Adjoint method (Han et al. 2009, Laurenceau et al. 2008) or lower-fidelity data by lower-fidelity analysis code (Han et al. 2010, Forrester et al. 2007).

2.2.3 Radial basis functions

In addition to Kriging, RBFs model (Hardy, 1971) is known as an alternative interpolation method for surrogate modeling. For the RBFs approach by Powell (1987), the approximation of the unknown function $y(\mathbf{x})$ at an untried \mathbf{x} is formally defined as the linear combination of the radial basis functions and a global trend function as

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^n \omega_i \varphi(\mathbf{x}) + P(\mathbf{x}), \quad (17)$$

where ω_i are the i -th unknown weight coefficient, $\varphi(\mathbf{x}) = \varphi(\|\mathbf{x}^{(i)} - \mathbf{x}\|)$ are the basis functions that depend on the Euclidean distance between the observed point $\mathbf{x}^{(i)}$ and the untried point \mathbf{x} (similar to the correlation function of kriging model); $P(\mathbf{x})$ is the global trend function which is taken as a constant β_0 here. To ensure the function values at observed points are reproduced by the RBFs predictor, the following constraints should be satisfied:

$$\hat{y}(\mathbf{x}^{(i)}) = y^{(i)}, i = 1, \dots, n. \quad (18)$$

Then the additional constraints for $P(\mathbf{x})$ should be imposed as

$$\sum_{i=0}^n \omega_i = 0. \quad (19)$$

Solving the linear equations formed by Eq. (18) and Eq. (19) for ω_i and β_0 , and substituting into Eq.(17) yields the RBFs predictor as

$$\hat{y}(\mathbf{x}) = \beta_0 + \boldsymbol{\varphi}^T(\mathbf{x}) \boldsymbol{\Psi}^{-1} (\mathbf{y}_S - \beta_0 \mathbf{1}). \quad (20)$$

Where $\beta_0(\boldsymbol{\theta}) = (\mathbf{1}^T \boldsymbol{\Psi}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \boldsymbol{\Psi}^{-1} \mathbf{y}_S$ and $\boldsymbol{\Psi}$, $\boldsymbol{\varphi}(\mathbf{x})$ are defined as

$$\boldsymbol{\Psi} := [\varphi(\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|)]_{ij} \in \mathbb{R}^{n \times n}, \quad \boldsymbol{\varphi}(\mathbf{x}) := [\varphi(\|\mathbf{x}^{(i)} - \mathbf{x}\|)]_i \in \mathbb{R}^n. \quad (21)$$

When the above RBFs predictor is compared with the Kriging predictor (see Eq. (10)), one can observe that they are essentially similar, only with the basis-function matrix $\boldsymbol{\Psi}$ (also called Gram matrix) and the basis function vector $\boldsymbol{\varphi}(\mathbf{x})$ being different from the correlation matrix \mathbf{R} and the correlation vector $\mathbf{r}(\mathbf{x})$ of the Kriging predictor, respectively. In addition, RBFs differs from Kriging at the following two aspects: 1) RBFs doesn't provide the uncertainty estimation of the prediction; 2) The model parameters can't be tuned by MLE like Kriging. Generally, Kriging can be regarded as a particular form of RBFs.

To build a RBFs model, one needs to prescribe the type of basis functions that only depends on the Euclidean distance $r = \|\mathbf{x} - \mathbf{x}'\|$ between any two sites \mathbf{x} and \mathbf{x}' . Compared to the correlation function used for a Kriging model, more choices are available for a RBFs model, which are partially listed in Table 1.

Basis functions	Formulations
Gaussian (GAUSS)	$\phi(r) = e^{-r^2/2\sigma^2}$ (e.g. $\sigma^2 = 1$)
Power function (POW)	$\phi(r) = r^\beta, 1 \leq \beta \leq 3$ (e.g. $\beta = 1.8$)
Thin Plate Spline (TPS)	$\phi(r) = r^2 \ln(r)$
Hardy's Multiquadric (HMQ)	$\phi(r) = \sqrt{1+r^2}$
Hardy's Inverse Multiquadric (HIMQ)	$\phi(r) = 1/\sqrt{1+r^2}$

Table 1. Basis functions for RBFs surrogate model

All the basis functions listed in Table 1 can be classified into two categories: decaying functions (such as GAUSS and HIMQ) and growing functions (POW, TPS and HMQ). The decaying functions can yield positive definite matrix Ψ , which allows for the use of Cholesky decomposition for its inversion; the growing functions generally result in a non-positive definite matrix Ψ and thus LU decomposition is usually used alternatively. The schematics of the basis functions for one-dimensional problem is sketched in Figure 3.

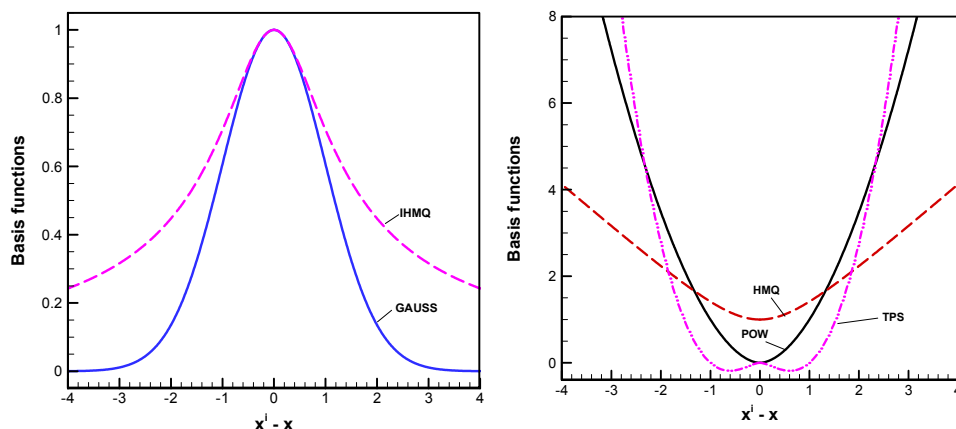


Fig. 3. Schematics of basis functions for Radial Basis Functions (left: decaying functions; right: growing functions).

2.2.4 The big (ger) picture

In addition to what we mentioned above, there are also a few surrogate models available in the literatures, such as Artificial Neural Network (ANN) (Elanayar et al. 1994, Park et al. 1991), Multiple Adaptive Regression Splines (MARS) and Support Vector Regression (SVR)

(Smola & Schoelkopf 2004). Although these methods are coming from different research communities, the idea is similar when using them for function prediction in surrogate modeling. They are not described in detail here due to the limited space. The readers are referred to read the paper by Wang & Shan (2007) and the books written by Keane et al. (2005) and by Forrester et al. (2008) for more description about surrogate modeling techniques.

2.3 Evaluation of approximation models

An important issue for the surrogate modeling is how to estimate the error of the approximation models. Only when the surrogate model with sufficient accuracy is built can the reliable optimum be obtained. Here we use two variables (\bar{e} and σ_e) to evaluate the error of the approximation models at test points, which are also chosen by DoE method. The average relative error is

$$\bar{e} = \frac{1}{n_t} \sum_{i=1}^{n_t} e^{(i)}, \quad e^{(i)} = \left\| \frac{\hat{y}_t^{(i)} - y_t^{(i)}}{y_t^{(i)}} \right\|, \quad (22)$$

where n_t is number of the test points; $y_t^{(i)}$ and $\hat{y}_t^{(i)}$ are the true value and predicted value corresponding to the i -th test point, respectively. The root mean squared error is defined by

$$\sigma_e = \sqrt{\sum_{i=1}^{n_t} (e^{(i)})^2 / n_t}. \quad (23)$$

2.4 Framework of building surrogate models

A Generic framework of building a surrogate model is sketched in Figure 4. Note that the initial surrogate model can be evaluated by Eqs. (22) and (23) and a branch for resampling is denoted by black dashed line in Figure 4.

3. Use of surrogate models for optimization

Suppose we are concerned with solving a general optimization problem as

$$\begin{aligned} \text{Objective} \quad & \text{minimize } y(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, i = 1, \dots, n_c, \\ & \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \end{aligned} \quad (24)$$

where n_c is the number of state functions which is in line with the number of inequality constraints (assuming that all the equality constraints have been transformed into inequality constraints.); \mathbf{x}_l and \mathbf{x}_u are the lower and upper bound of design variables, respectively; the object function $y(\mathbf{x})$ and state functions $g_i(\mathbf{x})$ are evaluated by an expensive analysis code. Traditionally, the optimization problem is solved by either a gradient-based algorithm or a gradient-free algorithm such as GA. It may become prohibitive due to the large computational cost associated running the expensive analysis code. Alternatively, here we are concerned with using surrogate modeling techniques to solve the optimization problem, in an attempt to dramatically improve the efficiency.

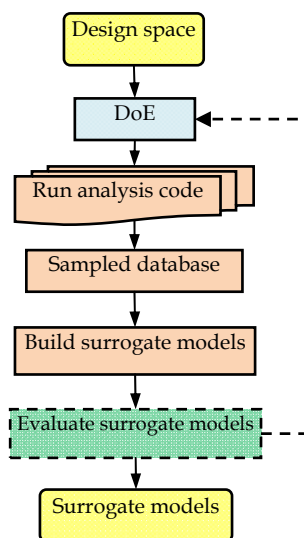


Fig. 4. Frameworks of building surrogate models

3.1 Framework of surrogate-based optimization

3.1.1 A simple framework

The basic idea of using surrogate models in optimization can be quite simple. First, the surrogate models for the object function(s) and state function(s) with sufficient accuracy are built (see Figure 2); second, the optimum is found by an optimizer, with the object function(s) or state function(s) evaluated by surrogate models, rather than by the expensive analysis code. Since prediction with the surrogate models is much more efficient than that by the expensive analysis code, the optimization efficiency can be greatly improved. The comparison of the conventional optimization and surrogate-based optimization is sketched in Figure 5. In addition to improve optimization efficiency, surrogate models also serve as an interface between the analysis code and the optimizer, which makes the establishment of an optimization procedure much easier. One of the examples for such a surrogate-based optimization framework can be found in paper by Han et al. (2010).

3.1.2 A bi-level framework

Although the framework of the surrogate-based optimization sketched in Figure 5. (b) is very intuitive and simple, questions may arise: are the surrogate models accurate enough? has the true optimum been reached? In fact, the optimum gained by the surrogate models is only an approximation to the true optimum (see Figure 5. (a)). One has to refine the surrogate models by adding new sample points, which is to be observed by running the analysis code. The procedure of augmenting new sample point(s) to the current sampled data sets is the so-called “sample-point refinement”. The rules of determining the new sample sites towards the true optimum are called “infill criteria”, which will be discussed in section 3.2. The flowchart of a surrogate-based optimization with additional process of

sample-point refinement is sketched in Figure 6. It can be regarded as a bi-level optimization framework, with the process of building and refining the surrogate models (which needs to run the expensive analysis code) acting as the main optimization and the process of using surrogate models to determine the new sample sites acting as the sub-optimization(s).

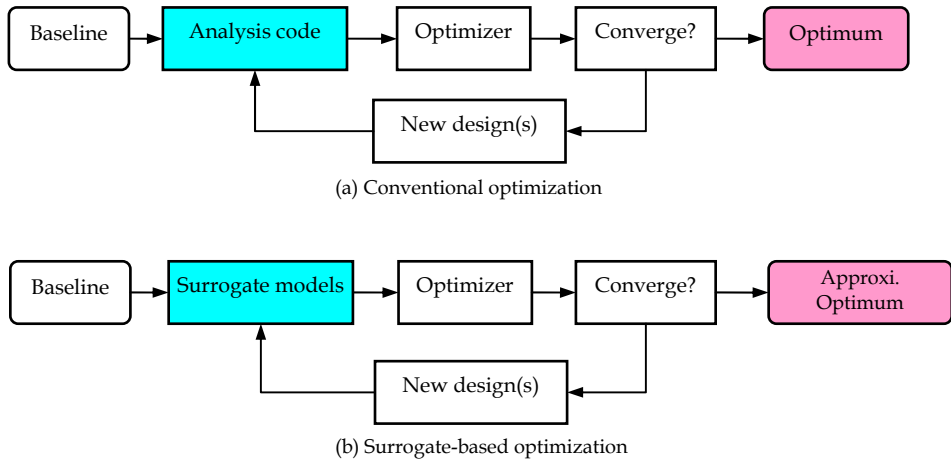


Fig. 5. Comparison of frameworks for conventional optimization and surrogate-based optimization with a simple framework

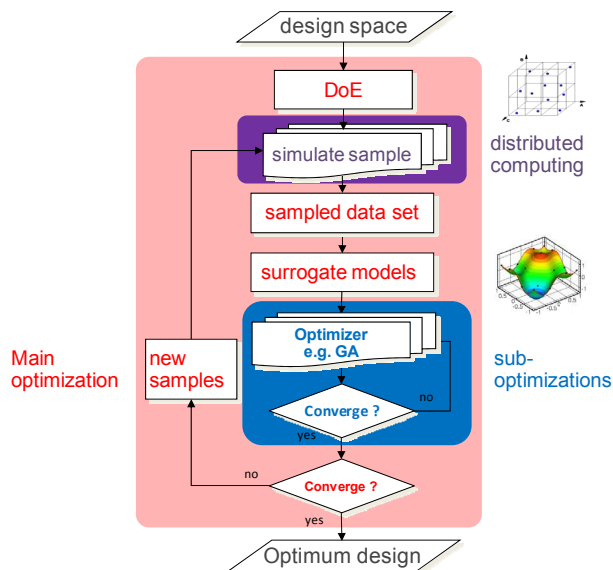


Fig. 6. Flowchart of the surrogate-based optimization with a bi-level framework (main optimization: building and refining the surrogate models which needs to run the expensive analysis code; sub-optimization(s): using surrogate models to determine new sample sites)

3.2 Infill criteria

The infill criterion is related to the determination of the new sample sites by solving sub-optimization problem(s). Three infill criteria are discussed here: Searching Surrogate Model(s) (SSM), Expected Improvement (EI) and Lower-Bounding Confidence (LCB).

3.2.1 Searching surrogate model(s)

Provided that the initial surrogate models have been built, an optimizer such as GA can be used to find the optimum, which in turn can be employed to refine the surrogate models.

The mathematical model of the sub-optimization for determining the new sample site is of the form

$$\begin{aligned} \text{Objective} \quad & \text{minimize } \hat{y}(\mathbf{x}) \\ \text{s.t.} \quad & \hat{g}_i(\mathbf{x}) \leq 0, i=1, n_c, \\ & \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \end{aligned} \quad (25)$$

where $\hat{y}(\mathbf{x})$ and $\hat{g}_i(\mathbf{x})$ are surrogate models of $y(\mathbf{x})$ and $g_i(\mathbf{x})$, respectively. With the optimal design variables $\hat{\mathbf{x}}_{\text{opt}}$ gained by the surrogate models in hand, one needs to run the expensive analysis code to compute the corresponding true function value and compare it with what predicted by the surrogate models. If the error between them is blow a threshold, the optimization process can be terminated; if not, the new sample point is augmented to the sampled data sets and the surrogate models are rebuilt; the process is repeated until the optimum solution is approached.

This criterion applies for all the surrogate models and is very efficient for local exploitation of the promising region in the design space.

3.2.2 Expected improvement

Surrogate model such as Kriging provides not only a function prediction but also an estimation of the mean squared error (MSE). In fact, the prediction by a Kriging model, $\hat{y}(\mathbf{x})$, at any point can be regarded as a Gaussian random variable with the mean given by the Kriging predictor, and the variance given by the mean squared error, $s^2(\mathbf{x})$ (see section 2.2.2). Viewed in this way, a probability can be computed that the function value at any untried \mathbf{x} would fall below the minimum among the sample points observed so far. Then Expected Improvement (EI) function (Jones et al 1998, Jeong et al. 2005) can be calculated to account for the improvement of the object function we expect to achieve at any untried \mathbf{x} . The definition of EI is of the form

$$E[I(\mathbf{x})] = \begin{cases} (y_{\min} - \hat{y}(\mathbf{x}))\Phi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) + \hat{s}(\mathbf{x})\phi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) & \text{if } \hat{s} > 0 \\ 0 & \text{if } \hat{s} = 0 \end{cases} \quad (26)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function and probability density function of a standard normal distribution, respectively. $y_{\min} = \text{Min}(y^{(1)}, y^{(2)}, \dots, y^{(n)})$

denotes the minimum of the observed data so far. The greater the EI, the more improvement we expect to achieve. The point with maximum EI is located by a global optimizer such as GA then observed by running the analysis code. For this infill criterion, the constraints can be accounted by introducing the probability that the constraints are satisfied. The corresponding sub-optimization problem can be modeled as

$$\begin{aligned} \text{Objective} \quad & \text{maximize } E[I(\mathbf{x})] \cdot \prod_{i=1}^{n_c} P[G_i(\mathbf{x}) \leq 0], \\ \text{s.t.} \quad & \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \end{aligned} \quad (27)$$

where $P[G_i(\mathbf{x}) \leq 0]$ denotes the probability that i -th constraint may be satisfied and $G_i(\mathbf{x})$ is a random function corresponding to i -th state function $g_i(\mathbf{x})$. $P[G_i(\mathbf{x}) \leq 0] \rightarrow 1$ when the constraint is satisfied and $P[G_i(\mathbf{x}) \leq 0] \rightarrow 0$ when the constraint is violated. $P[G_i(\mathbf{x}) \leq 0]$ can be calculated by

$$P[G_i(\mathbf{x}) \leq 0] = \frac{1}{\hat{s}_i(\mathbf{x})\sqrt{2\pi}} \int_{-\infty}^0 e^{-[G_i(\mathbf{x}) - \hat{g}_i(\mathbf{x})]^2 / 2\hat{s}_i^2(\mathbf{x})} dG_i(\mathbf{x}) = \Phi\left(\frac{-\hat{g}_i(\mathbf{x})}{\hat{s}_i(\mathbf{x})}\right) \quad (28)$$

where $\hat{s}_i(\mathbf{x})$ denotes the estimated standard error corresponding to the surrogate model $\hat{g}_i(\mathbf{x})$.

The optimum site $\hat{\mathbf{x}}_{\text{opt}}$ obtained by solving Eq. (27) is observed by running analysis code and the new sample point is added to the sampled data sets; the surrogate models are rebuilt and the whole process is repeated until the global optimum is approached.

3.2.3 Lower-bounding confidence (LCB)

The LCB function is defined as the weighted sum of predicted function value $\hat{y}(\mathbf{x})$ and the standard error of the prediction $\hat{s}(\mathbf{x})$. For an optimization problem of finding the minimum of the unknown function $y(\mathbf{x})$, a simple expression for LCB function is of the form

$$\text{LCB} = \hat{y}(\mathbf{x}) - A\hat{s}(\mathbf{x}), \quad (29)$$

where A is a constant which balances the influence of the predicted function and the corresponding uncertainty. Best practice suggests $A=1$ works well for a number of realistic problems. The corresponding sub-optimization problem can be modeled as

$$\begin{aligned} \text{Objective} \quad & \text{minimize } \hat{y}(\mathbf{x}) - A\hat{s}(\mathbf{x}) \\ \text{s.t.} \quad & \hat{g}_i(\mathbf{x}) \leq 0, i = 1, n_c, \\ & \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \end{aligned} \quad (30)$$

The above optimization problem can be solved via a global optimizer such as GA. Since the point with smallest value of LCB indicates the possible minimum of the unknown function, the optimum site $\hat{\mathbf{x}}_{\text{opt}}$ is then observed and added to sampled data sets to refine the surrogate models. This procedure is performed iteratively until the global optimum is reached.

4. Examples for applications to aircraft design

4.1 Airfoil design

Using an in-house Navier-Stokes flow solver, the objective of the problem is to minimize the drag of an RAE2822 airfoil at the flow condition of $Ma = 0.73$, $\alpha = 2.7^\circ$, $Re = 6.5 \times 10^6$, subject to 3 constraints:

$$\begin{aligned} \text{Objective} &: \text{Minimize } C_d \\ \text{st.} &: (1) \text{Area} \geq 0.99 \times \text{Area}_0 \\ &: (2) C_l \geq 0.99 \times C_{l0} \\ &: (3) |C_m| \leq |C_{m0}| \end{aligned} \quad (31)$$

where Area_0 , C_{l0} , C_{m0} are the area, lift coefficient, and moment coefficient of the baseline airfoil, respectively. The first constraint is in consideration of the structural design of the wing to guarantee the volume of the wing; the second one is to enforce a constant lift of the wing in order to balance the weight of the aircraft at cruise condition; the third one is to control the pitching moment of the airfoil to avoid large drag penalty of the horizontal tail paid for balancing the aircraft.

The initial number of samples for Kriging is set to 20, selected by the Latin Hypercube Sampling (LHS). The airfoil is parameterized by 10 Hicks-Henne bump functions (Hicks & Henne, 1978); and the maximum amplitude of each bump is $A^{\max}/c = 0.544\%$. Both of the SSM and EI infill strategies are adopted in the surrogate refinement. Table 2 presents the optimization results of the two optimization method. The optimized and initial airfoils and the corresponding pressure coefficient distributions are compared in Figure 7. Note that the aerodynamic coefficients of the initial airfoil RAE2822 are set to 100. Obviously, the Kriging-based optimization method gives better result, and with higher efficiency, and is more likely to find the global optimum.

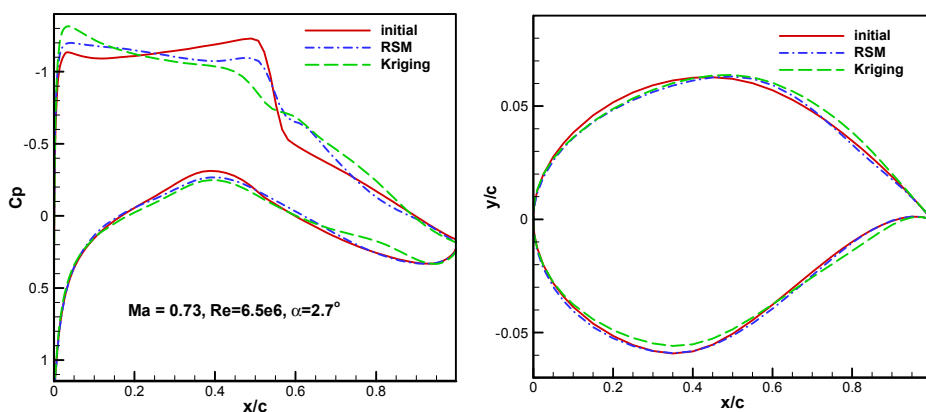


Fig. 7. Aerodynamic shape optimization of a transonic airfoil (RAE 2822) via Kriging and quadratic Response Surface Model (left: pressure distribution; right: airfoil shape); by using Kriging model with Expected Improvement infill criteria, the drag is reduced by 33.6% with only 56 calling of Navier-Stokes flow solver.

	C_d	C_l	$ C_m $	Area	NS flow solver calls
baseline	100	100	100	100	-
RSM	73.0 (-27.0%)	101.1 (+1.1%)	99.3 (-0.7%)	99.9 (-0.1%)	102
Kriging	70.7 (-29.3%)	101.1 (+1.1%)	96.5 (-3.5%)	100.4 (+0.4%)	57

Table 2. Drag reduction of an RAE 2822 airfoil via Kriging and RSM-based optimizations

4.2 Wing design

Here we are concerned with the preliminary design for a high-subsonic transport-aircraft wing of a wing/body combination, considering aerodynamic, structure and static aeroelastic effect. The calculation of the external flow is carried out by numerical solutions of the full potential equation in conservative form (Kovalev & Karas, 1991). The FEM-based commercial software ANSYS is used for analyzing the structural performance of the wing with double-beams sandwich structure. Weak coupling method is adopted for static aeroelastic analysis.

The optimization objectives are to maximize the aircraft lift-to-drag ratio and minimize the weight of wing for a fixed maximum take-off weight of 54 tons and cruise Mach number of 0.76 at 10,000 meters high. The wing is composed of inner and outer wing. The reference area of wing is 105 square meter. The mathematical model for optimization is of the form

$$\begin{aligned}
 &\max L/D \\
 &\min W_{\text{wing}} \\
 &\text{s.t. } L \geq 54 \quad / 10^3 \text{ kg} \\
 &\quad 100 \leq S_{\text{wing}} \leq 110 \quad / \text{m}^2 \\
 &\quad \sigma_{\max} \leq \sigma_b \quad / 10^9 \text{ pa} \\
 &\quad \delta_{\max} \leq 1 \quad / \text{m}
 \end{aligned} \tag{32}$$

Eight supercritical airfoils are configured along the span. The optimization is subject to 4 constraints. The first constraint is to enforce a constant lift of the wing in order to balance the weight of the aircraft at cruise condition; the second one is to guarantee a near constant wing loading; the third and fourth constraints are to make sure that the strength and rigidity requirements are satisfied. The definition for the limits of design variables is listed in Table 3. The first four design variables define the aerodynamic configuration of the wing. The four remain are for structure design. The detail can be found in paper by Zhang et al. (2008).

The uniform design table, U100(10^8), is used to create one hundred candidate wings for building surrogate models. The other forty-five candidate wings are created by the uniform design table, U45(10^8), for evaluating the approximation model. For each wing, the static aeroelastic analysis is performed to obtain the responses of lift (L), lift-to-drag ratio (L/D), wing area (S_{wing}), maximum stress (σ_{\max}), maximum deformation (δ_{\max}) and wing weight

(W_{wing}). Then the average relative errors and the root mean squared errors are calculated to evaluate the approximation models, as listed in Table 4. In this case Kriging and RSM have comparative high accuracy.

Design variable	Unit	Lower limit	Upper limit
Span	m	26	34
Taper ratio	/	0.2	0.4
Linear twist angle		-3	-1
Sweep angle on leading edge	degree	25	35
Thickness of front-spar web	mm	2	6
Thickness of back-spar web	mm	2	6
Thickness of lower skin	mm	3	7
Thickness of Upper skin	mm	3	7

Table 3. Definition of design variables for preliminary design of a high-subsonic transport-aircraft wing

Parameter	Surrogate model	\bar{e}	σ_e	Parameter	Surrogate model	\bar{e}	σ_e
L	RSM	0.0360	0.0213	σ_{max}	RSM	0.0563	0.0535
	Kriging	0.0362	0.0213		Kriging	0.0515	0.0522
L/D	RSM	0.0122	0.0099	δ_{max}	RSM	0.0227	0.0241
	Kriging	0.0123	0.0097		Kriging	0.0227	0.0241
S_{wing}	RSM	0.0071	0.0051	W_{wing}	RSM	0.0140	0.0104
	Kriging	0.0071	0.0051		Kriging	0.0142	0.0107

Table 4. Evaluation of modeling accuracy

Then the multi-objective optimization for the supercritical wing is performed based on RSM due to its higher computational efficiency. Weighted sum method is used to transform the multi-objective optimization into a single-objective optimization. Sequential quadratic programming method is employed to solve the optimization. One of the candidate wings with better performance, are selected as the initial point for optimization. The optimal design is observed by running the analysis codes and the results are listed in Table 5. Where \mathbf{X}^0 and \mathbf{Y}^0 is the initial wing scheme and its response, respectively; \mathbf{X}^* and \mathbf{Y}^* is the optimal wing scheme and its actual response, respectively; $\hat{\mathbf{Y}}$ is the response at \mathbf{X}^* calculated by the approximation models. For the optimal wing scheme, the largest relative error of approximation models is no more than 3 percent. It again proves the high accuracy of the approximation models.

Figure 8 shows the contour of the equivalent stress of the optimal wing. It shows that the stress is larger in the location of intersection of inner wing and outer wing due to the inflexion. Figure 9 shows pressure distribution of the optimal wing. It shows that the wing

basically meets the design requirements of a supercritical wing. A little bit non-smoothness of the pressure distribution may be caused by non-uniform deformation of the skin. Figure 10 shows the convergence history of aeroelastic deformation, which shows that fast convergence of the aeroelastic deformation of the optimal wing.

The optimization, together with the aeroelastic analysis of all candidate wings, only takes about two days on a personal computer of Pentium(R) 4 CPU 2.8GHz. If more computers are used to concurrently calculate the performance of different candidate wings, the cost can be further greatly reduced.

	B/m	λ	$\theta/(^\circ)$	$\Lambda/(^\circ)$	T_{FS}/mm	T_{BS}/mm	T_{LS}/mm	T_{US}/mm
X^0	34.00	0.244	-1.667	29.444	3.333	3.778	6.556	3.889
X^*	31.69	0.200	-1.563	28.233	2.232	2.000	4.396	4.057
	$L/10^3kg$	L/D	S_{wing}/m^2	$\sigma_{max}/10^9pa$	δ_{max}/m	$W_{wing}/10^3kg$		
Y^0	51.50	27.81	111.94	0.311	1.191	3.850		
Y^*	53.14	27.40	107.25	0.275	0.991	3.063		
\hat{Y}	54.00	27.45	106.27	0.267	1.000	3.003		
Modeling error	1.61%	0.16%	0.91%	2.97%	0.87%	1.85%		

Table 5. Optimization results when considering the aeroelastic effect

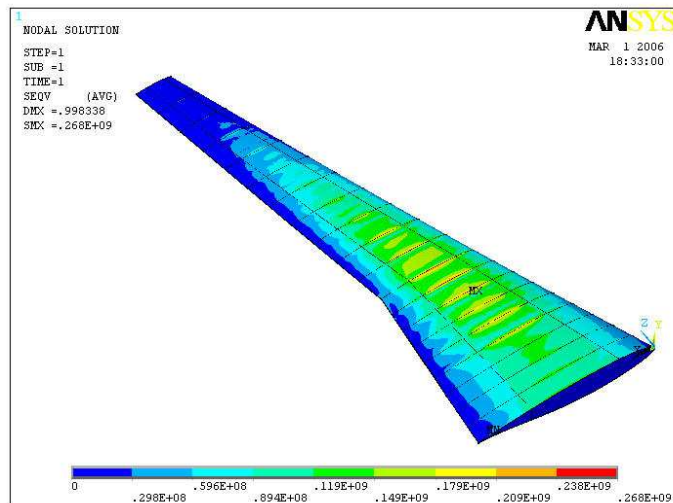


Fig. 8. Contour of equivalent stress of the optimal wing

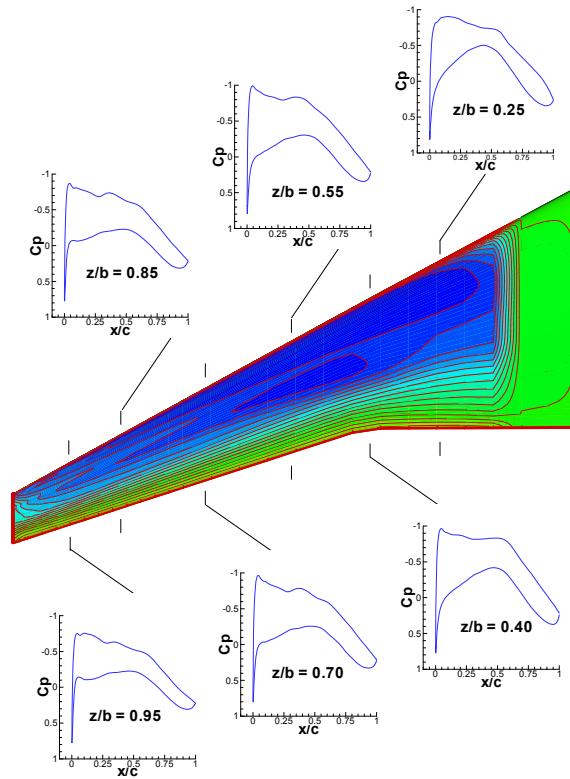


Fig. 9. Pressure distribution of the optimal wing ($Ma=0.76$, $Re=0.257E+08$, $\alpha=0^\circ$)

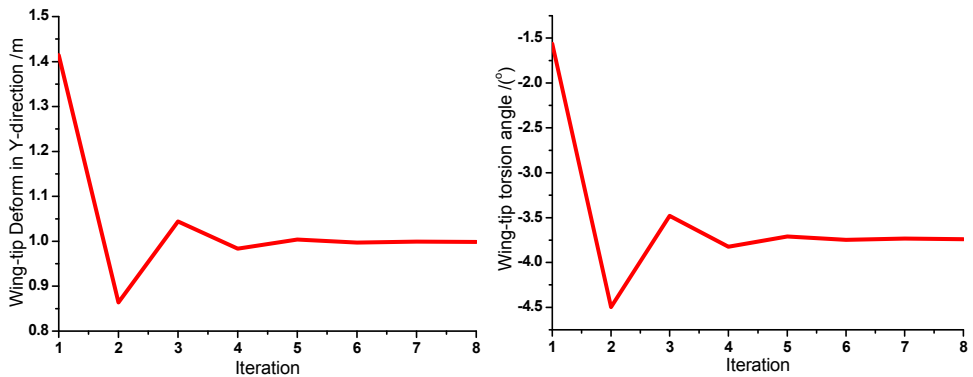


Fig. 10. Convergence history of Y-direction deform and torsion deform on wing tip

5. Conclusion

An overview of the existing surrogate models and the techniques about how to use them for optimization is presented in this chapter. Among the surrogate models, the regression model such as the quadratic response surface model (RSM) is well suited for a local optimization problem with relatively simpler design space; interpolation models such as Kriging or RBFs can be used for highly-nonlinear, multi-modal functions, and thus well suited for a global problem with relatively more complicated design space. From an application point of view, the simple framework of surrogate-based optimization is a good choice for an engineer design, due to the fact that surrogate model can act as an interface between the expensive analysis code and the optimizer and one doesn't need to change the analysis code itself. The drawback of this framework is that the accuracy of optimum only depends on the approximation accuracy of surrogate model and we generally get an approximation to the true optimum. In contrast, the bi-level framework with different infill criteria provides an efficient way to quickly find true optimum without the need of building globally accurate surrogate models. Multiple infill criteria seem to be a better way to overcome the drawback of the single infill criterion.

Examples for airfoil and wing designs show that surrogate-based optimization is very promising for aerodynamic problem with number of design variables being less than about 10. For higher-dimensional problem, the computational cost increases very quickly, which can be prohibitive. Thus, use of surrogate model for high(er)-dimensional optimization problems would become an important issue of future work.

6. Acknowledgements

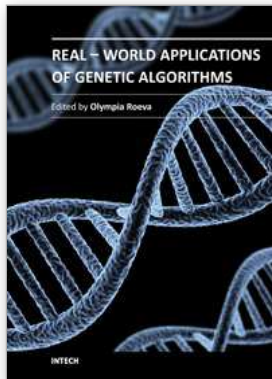
This research was sponsored by the National Natural Science Foundation of China (NSFC) under grant No. 10902088 and the Aeronautical Science Foundation of China under grant No. 2011ZA53008

7. References

- Elanayar, S. V. T., and Shin, Y. C., "Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems," *IEEE Transactions on Neural Networks*, Vol. 5, No. 4, 1994, pp. 594-603.
- Fang, K. T., Lin, D., Winker, P., Zhang, Y., "Uniform design: Theory and application," *Technometrics*, Vol. 42, No. 3, 2000, pp. 237-248.
- Forrester, A. I. J., Sóbester, A., and Keane, A. J., "Multi-Fidelity Optimization via Surrogate Modeling," *Proceedings of the Royal Society A*, Vol. 463, No. 2088, 2007, pp. 3251-3269.
- Forrester, A. I. J., Sóbester, A., and Keane, A., "Engineering Design via Surrogate Modeling: A Practical Guide," *Progress in Astronautics and Aeronautics Series*, 226, published by John Wiley & Sons, 2008.
- Giunta, A. A., Wojtkiewicz Jr, S. F., and Eldred, M. S. , "Overview of Modern Design of Experiments Methods for Computational Simulations," *AIAA paper 2003-649*, 2001.

- Han, Z. -H., Zhang, K. -S., Song, W. -P., and Qiao, Z. -D., "Optimization of Active Flow Control over an Airfoil Using a Surrogate-Management Framework," *Journal of Aircraft*, 2010, Vol. 47, No. 2, pp. 603-612.
- Han, Z.-H., Görtz, S., Zimmermann, R. "On Improving Efficiency and Accuracy of Variable-Fidelity Surrogate Modeling in Aero-data for Loads Context," Proceeding of CEAS 2009 European Air and Space Conference, Manchester, UK, Oct. 26-29, 2009.
- Han, Z.-H., Zimmermann, R., and Görtz, S., "A New Cokriging Method for Variable-Fidelity Surrogate Modeling of Aerodynamic Data," AIAA Paper 2010-1225, 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, Orlando, Florida, Jan. 4-7, 2010.
- Hardy, R. L., "Multiquadric Equations of Topography and Other Irregular Surface," *Journal of Geophysical Research*, Vol. 76, March, 1971, pp. 1905-1915.
- Hicks, R. M., and Henne, P. A., "Wing Design by Numerical Optimization," *Journal of Aircraft*, Vol. 15, No. 7, 1978, pp. 407-412.
- Jeong, S., Murayama, M., and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol. 42, No. 2, 2005, pp. 413- 420.
- Jones, D., Schonlau, M., Welch W., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, 1998, Vol. 13, pp. 455-492.
- Keane, A. J., Nair, P. B., "Computational Approaches for Aerospace Design: The Pursuit of Excellence", John Wiley & Sons, Ltd, Chichester, 2005.
- Kovalev, V.E., Karas, O. V. "Computation of a Transonic Airfoil Flow Considering Viscous Effects and Thin Separated Regions." *La Recherche Aerospatiale (English Edition)* (ISSN 0379-380X), No. 1, 1991, pp. 1-15.
- Krige, D. G., "A Statistical Approach to Some Basic Mine Valuations Problems on the Witwatersrand," *Journal of the Chemical, Metallurgical and Mining Engineering Society of South Africa*, Vol. 52, No. 6, 1951, pp. 119-139.
- Laurenceau, J., and Sagaut, P. "Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging," *AIAA Journal*, Vol. 46, No. 2, 2008, pp. 498-507.
- Park. J., and Sandberg, I. W., "Universal Approximation Using Radial-Basis-Function Networks," *Neural Computation*, Vol. 3, No. 2, 1991, pp. 246-257.
- Powell, M. J. D. "Radial Basis Functions for Multivariable Interpolation: A Review, " *Algorithms for Approximation*, edited by J. C. Mason and M. G. Cox, Oxford Univ. Press, New York, 1987, Chap. 3, pp. 141-167.
- Queipo, N. V., Haftka, R. T., Shyy W., Goel, T., Vaidyanathan, R., and Tucher, P. K., "Surrogate-based Analysis and Optimization," *Progress in Aerospace Sciences*, Vol. 41, 2005, pp. 1-28.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, 1989, pp. 409-423.
- Simpson, T. W., Mauery, T. M., Korte, J. J., et al., "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization", *AIAA Journal*, Vol. 39, No. 12, 2001, pp. 2233-2241.
- Simpson, T. W., Toropov, V., Balabanov, V., and Viana, F. A. C., "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: a Review of How Far We Have Come - or Not," AIAA Paper 2008-5802, 2008.

- Smola, A. J. and Schoelkopf, B., "A tutorial on support vector regression," *Statistics and Computing*, Vol. 14, 2004, pp. 199-222.
- Toal, D. J. J., Bressloff, N. W., Kean, J., "Kriging Hyperparameter Tuning Strategies," *AIAA Journal*, Vol. 46, No. 5, 2008, pp. 1240-1252.
- Wang, G. G., Shan S., "Review of Metamodeling Techniques in Support of Engineering Design Optimization," *Journal of Mechanical Design*, Vol. 129, No. 4. 2007, pp. 370-380.
- Zhang, K. -S., Han, Z. -H., Li, W. -J., and Song, W. -P., "Coupled Aerodynamic/Structural Optimization of a Subsonic Transport Wing Using a Surrogate Model," *Journal of Aircraft*, 2008, Vol. 45, No. 6, pp. 2167-2171.



Real-World Applications of Genetic Algorithms

Edited by Dr. Olympia Roeva

ISBN 978-953-51-0146-8

Hard cover, 376 pages

Publisher InTech

Published online 07, March, 2012

Published in print edition March, 2012

The book addresses some of the most recent issues, with the theoretical and methodological aspects, of evolutionary multi-objective optimization problems and the various design challenges using different hybrid intelligent approaches. Multi-objective optimization has been available for about two decades, and its application in real-world problems is continuously increasing. Furthermore, many applications function more effectively using a hybrid systems approach. The book presents hybrid techniques based on Artificial Neural Network, Fuzzy Sets, Automata Theory, other metaheuristic or classical algorithms, etc. The book examines various examples of algorithms in different real-world application domains as graph growing problem, speech synthesis, traveling salesman problem, scheduling problems, antenna design, genes design, modeling of chemical and biochemical processes etc.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Zhong-Hua Han and Ke-Shi Zhang (2012). Surrogate-Based Optimization, Real-World Applications of Genetic Algorithms, Dr. Olympia Roeva (Ed.), ISBN: 978-953-51-0146-8, InTech, Available from:
<http://www.intechopen.com/books/real-world-applications-of-genetic-algorithms/surrogate-based-optimization>

INTeCH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821