Science in China Series E: Technological Sciences

www.scichina.com
tech.scichina.com
www.springerlink.com

Gradient-based Kriging approximate model and its application research to optimization design

XUAN Ying[†], XIANG JunHua, ZHANG WeiHua & ZHANG YuLin

College of Aerospace and Material Engineering, National University of Defense Technology, Changsha 410073, China

In the process of multidisciplinary design optimization, there exits the calculation complexity problem due to frequently calling high fidelity system analysis models. The high fidelity system analysis models can be surrogated by approximate models. The sensitivity analysis and numerical noise filtering can be done easily by coupling approximate models to optimization. Approximate models can reduce the number of executions of the problem's simulation code during optimization, so the solution efficiency of the multidisciplinary design optimization problem can be improved. Most optimization methods are based on gradient. The gradients of the objective and constrain functions are gained easily. The gradient-based Kriging (GBK) approximate model can be constructed by using system response value and its gradients. The gradients can greatly improve prediction precision of system response. The hybrid optimization method is constructed by coupling GBK approximate models to gradient-based optimization methods. An aircraft aerodynamics shape optimization design example indicates that the methods of this paper can achieve good feasibility and validity.

Kriging approximation function, gradient-based approximate model, hybrid optimization method, multidisciplinary design optimization

1 Introduction

In the process of multidisciplinary design optimization (MDO), there exits the calculation complexity problem due to frequently calling high fidelity system analysis models. Therefore, the high fidelity system analysis models should be surrogated by approximate models. The approximation method separates complex disciplinary analysis from the optimization process by constructing approximation function which is convenient for computing, then the function is coupled into the optimization algorithm for sequential optimization, and approximate optimal solution of the realistic condition is obtained after several iterations. The approximate model is widely used as the key technology of MDO since it can both filter out numerical noises and make the sensitivity analysis of optimal design point facile.

Sacks et al.[1] studied the design and analysis technology of computer simulation experiment based on Kriging approximation function, and evaluated the I/O relationship of confirmation computer simulation model through Kriging approximation function. Kriging function has such advantages as unbiased estimator at the training sample point, desirably strong non-linear approximating ability and flexible parameter selection of the model, and thus it is quite suitable for approximate models[2]. Kriging models show great promise for building accurate global approximations of a design space[3]. These models are extremely flexible because of the wide range of spatial correlation functions that can be chosen for building the approximation, provided that sufficient sample data are available to capture the trends in the system responses. As a result, Kriging models can approximate linear and nonlinear functions equally well. Furthermore, Kriging models can either “honor the data”, by providing an exact interpolation of the data, or

Received July 6, 2008; accepted January 27, 2009
doi: 10.1007/s11431-009-0096-2
[†]Corresponding author (email: xuanyingxy@126.com)
Supported by the National High Technology Research and Development Program of China (“863” Program)

"smooth the data", by providing an inexact interpolation[4,5].

Many researchers have employed Kriging modeling strategies specifically for numerical optimization[6−8]. Timothy et al.[9] investigated the use of Kriging models as alternatives to traditional second-order polynomial response surfaces for constructing global approximations for use in a real aerospace engineering application, namely the design of an aerospike nozzle. Han et al.[10] presented a Kriging model-based multidisciplinary design optimization framework for turbine blade. Jeong S et al.[11] applied the Kriging modeling method to two-dimensional (2D) transonic airfoil design. The Kriging was used with the empirical tool for wing optimization, and the method was much quicker to use than direct searches of the CFD[12]. And Wang[13] applied the Kriging model to the aerodynamic optimization design for airfoil.

The paper establishes gradient-based Kriging (GBK) approximate model with gradient value and function value, improves the predicted accuracy of Kriging function with gradient information, and constructs a hybrid optimization method by combining GBK model with gradient based optimization method. Finally, the validity of GBK model and the hybrid optimization method are confirmed respectively by aerodynamics shape optimization design example of a certain aircraft. The hybrid optimization method based on GBK model proposed in this paper is suitable for complex cases of optimization design using high fidelity system analysis model.

2 Kringing approximation function

The system analysis model can be represented as a function $y(x)$: The I/O relationship between the n dimensional input x and output y, while the approximation method is an approximation function $\hat{y}(x)$ sampled at m points: $(\boldsymbol{X},\boldsymbol{Y})^{\mathrm{T}}=\{(x^1,y^1),\cdots,(x^m,y^m)\}$ of $y(x)$. Kriging function combines linear regression model and random process model to predict the real function of I/O relationship[14]:

$$\hat{y}(x)=\boldsymbol{f}^{\mathrm{T}}(x)\boldsymbol{\beta}+z(x), \tag{1}$$

where $\boldsymbol{f}^{\mathrm{T}}(x)=\{f_1(x),\cdots,f_p(x)\}$. It is the polynomial regression function with $\boldsymbol{\beta}^{\mathrm{T}}=\{\beta_1,\cdots,\beta_p\}$ as the regression coefficient[2] and $z(x)$ as the stationary Gaussian random function with zero average. The random variable $\boldsymbol{Z}^{\mathrm{T}}=\{z(x^1),\cdots,z(x^m)\}$ of stationary Gaussian random function is of multivariate normal distribution with a covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma}=\sigma^2\boldsymbol{R},\quad \boldsymbol{R}=[\boldsymbol{R}(x^i-x^j)]\quad (i,j\in[1,m]), \tag{2}$$

where σ^2 is the stationary Gaussian random function variance, $\boldsymbol{R}(x^i-x^j)$ is the correlation function of random variables $z(x^i)$ and $z(x^j)$ with product form.

$$\begin{cases}\boldsymbol{R}(x^i-x^j)=\boldsymbol{R}(d,\boldsymbol{\theta})=\prod_{k=1}^{n}\boldsymbol{R}_k(d_k,\theta_k),\\ d=x^i-x^j,d_k=x_k^i-x_k^j,\theta_k>0,\end{cases} \tag{3}$$

where $\boldsymbol{\theta}^{\mathrm{T}}=\{\theta_1,\cdots,\theta_n\}$ is the parameter of the correlation function. The random variables y and $\boldsymbol{Y}^{\mathrm{T}}=\{y^1,\cdots,y^m\}$ correspond to sample points at x and $\boldsymbol{X}^{\mathrm{T}}=\{x^1,\cdots,x^m\}$ of the Kriging approximation function in eq. (1) with multivariate normal distribution:

$$\begin{bmatrix}y\\ \boldsymbol{Y}\end{bmatrix}\sim N_{1+m}\left(\begin{bmatrix}\boldsymbol{f}^{\mathrm{T}}\\ \boldsymbol{F}\end{bmatrix}\boldsymbol{\beta},\sigma^2\begin{bmatrix}1 & \boldsymbol{r}^{\mathrm{T}}\\ \boldsymbol{r} & \boldsymbol{R}\end{bmatrix}\right). \tag{4}$$

In eq. (4), $\boldsymbol{f}^{\mathrm{T}}=\boldsymbol{f}^{\mathrm{T}}(x)$ is the p dimensional row vector of the regression function. $\boldsymbol{F}=\boldsymbol{f}_i^{\mathrm{T}}(\boldsymbol{x}^j)\,(j\in[1,m]$, $i\in[1,p])$ is the $m\times p$ dimensional regression function matrix; $\boldsymbol{r}^{\mathrm{T}}=\{\boldsymbol{R}(x-x^1),\cdots,\boldsymbol{R}(x-x^m)\}$ is the m dimensional column vector of correlation function, and $\boldsymbol{R}=[\boldsymbol{R}(x^i-x^j)](i,j\in[1,m])$ is m dimensional correlation function matrix. The distribution of $y\,|\,\boldsymbol{Y}$ is

$$(y|\boldsymbol{Y})\sim N_1\left(\boldsymbol{f}^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{r}^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{Y}-\boldsymbol{F}\boldsymbol{\beta}),\sigma^2\left(1-\boldsymbol{r}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{r}\right)\right). \tag{5}$$

Given sample data $\boldsymbol{Y}$, the minimal mean square error $\hat{y}$ of y is the condition mean of $y\,|\,\boldsymbol{Y}$:

$$\hat{y}(x)=E\{y|\boldsymbol{Y}\}=\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{\beta}+\boldsymbol{r}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{R}^{-1}(\boldsymbol{Y}-\boldsymbol{F}\boldsymbol{\beta}). \tag{6}$$

The undefined parameters of Kriging approximation function include regression coefficient $\boldsymbol{\beta}$, stationary Gaussian random function variance σ^2 and correlation function parameter $\boldsymbol{\theta}$, while the maximum likelihood estimator of β and σ^2 are defined as

$$\begin{cases}\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})=\left(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{Y}^n,\\ \hat{\sigma}^2=\hat{\sigma}^2(\boldsymbol{\theta})=\dfrac{1}{n}\left(\boldsymbol{Y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}\boldsymbol{R}^{-1}\left(\boldsymbol{Y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right).\end{cases} \tag{7}$$

And the maximum likelihood estimator $\hat{\theta}$ of $\boldsymbol{\theta}$ is at the extremum of the optimization problem:

$$\left\{\hat{\sigma}^2(\hat{\theta})\left|\boldsymbol{R}(\hat{\theta})\right|^{1/n}\right\}=\min_{\theta\in\Theta}\left\{\hat{\sigma}^2(\theta)\left|\boldsymbol{R}(\boldsymbol{\theta})\right|^{1/n}\right\}. \tag{8}$$

$\hat{\boldsymbol{\beta}}$ can be obtained by substituting $\hat{\theta}$ into eq. (7), and the Kriging approximation function estimation $\hat{y}(x)$ of $y(x)$ can be obtained by substituting $\hat{\theta}$ and $\hat{\boldsymbol{\beta}}$ into eq. (6).

3 Gradient-based Kriging approximate model

The random process $y(x)$ in eq. (1) has an average of $\boldsymbol{f}^{\mathrm{T}}(x)\boldsymbol{\beta}$, and a variance of σ^2. If random process $z(x)$ adopts product Gaussian correlation function as

$$\boldsymbol{R}\left(x^i-x^j\right)=\prod_{k=1}^{n}e^{-\theta_k(x_k^i-x_k^j)^2}, \tag{9}$$

then the first order partial derivative with respect to variable x_k of random process $y(x)$ is

$$y^{(k)}(x)=\frac{\partial y}{\partial x_k}(x), k\in[1,n]. \tag{10}$$

The above mentioned is also a stationary Gaussian random function[15] with covariance function and cross-covariance function as

$$\begin{cases}C_{yy}\left(x^i,x^j\right)=\sigma^2\boldsymbol{R}\left(x^i-x^j\right),\\ C_{yy^{(k)}}\left(x^i,x^j\right)=2\theta_k(x_k^i-x_k^j)C_{yy}\left(x^i,x^j\right),\\ C_{y^{(l)}y^{(k)}}\left(x^i,x^j\right)=-4\theta_l\theta_k(x_l^i-x_l^j)(x_k^i-x_k^j)C_{yy}\left(x^i,x^j\right),\\ C_{y^{(k)}y^{(k)}}\left(x^i,x^j\right)=[2\theta_k-4\theta_k^2(x_k^i-x_k^j)^2]C_{yy}\left(x^i,x^j\right).\end{cases} \tag{11}$$

If the stationary Gaussian random function $y(x)$ and $y^{(k)}(x)$ are applied to estimate the system output and its first order derivative about the input variable at x, then y estimated at x, $\boldsymbol{Y}^{(0)}=\{y(x^1),\ldots,y(x^m)\}^{\mathrm{T}}$ and $\{\boldsymbol{Y}^{(1)},\cdots,\boldsymbol{Y}^{(n)}\}^{\mathrm{T}}$ estimated at $\boldsymbol{X}=\{x^1,\cdots,x^m\}^{\mathrm{T}}$ are submitted to multivariate normal distribution as

$$\begin{pmatrix}y\\ \boldsymbol{Y}\end{pmatrix}\sim N_{1+m+mn}\left(\begin{pmatrix}\boldsymbol{f}^{\mathrm{T}}\\ \boldsymbol{F}\end{pmatrix}\boldsymbol{\beta},\begin{pmatrix}1 & \boldsymbol{c}^{\mathrm{T}}\\ \boldsymbol{c} & \boldsymbol{C}\end{pmatrix}\right), \tag{12}$$

where $\boldsymbol{\beta}^{\mathrm{T}}=\{\beta_1,\cdots,\beta_p\}$, $\boldsymbol{f}^{\mathrm{T}}=\{f_1(x),\cdots,f_p(x)\}$, $\boldsymbol{Y}=\{\boldsymbol{Y}^{(0)},\boldsymbol{Y}^{(1)},\cdots,\boldsymbol{Y}^{(n)}\}^{\mathrm{T}}$,

$$\boldsymbol{Y}^{(k)}=\left\{\frac{\partial y\left(x^1\right)}{\partial x_k},\ldots,\frac{\partial y\left(x^m\right)}{\partial x_k}\right\}^{\mathrm{T}},\ \boldsymbol{F}=\{\boldsymbol{F}^{(0)},\boldsymbol{F}^{(1)},\cdots,\boldsymbol{F}^{(n)}\}^{\mathrm{T}},$$

$$\boldsymbol{F}^{(0)}=\begin{Bmatrix}f_1(x^1),\cdots,f_p(x^1)\\ \vdots\\ f_1(x^m),\cdots,f_p(x^m)\end{Bmatrix},$$

$$\boldsymbol{F}^{(k)}=\begin{Bmatrix}\dfrac{\partial f_1(x^1)}{\partial x_k},\cdots,\dfrac{\partial f_p(x^1)}{\partial x_k}\\ \vdots\\ \dfrac{\partial f_1(x^m)}{\partial x_k},\cdots,\dfrac{\partial f_p(x^m)}{\partial x_k}\end{Bmatrix},$$

$$\boldsymbol{c}=\left\{\left(\boldsymbol{c}_{YY}\right),\left(\boldsymbol{c}_{YY^{(1)}}\right),\cdots,\left(\boldsymbol{c}_{YY^{(n)}}\right)\right\}^{\mathrm{T}},$$

$$\boldsymbol{c}_{YY}=\left\{C_{yy}\left(x,x^1\right),\cdots,C_{yy}\left(x,x^m\right)\right\}^{\mathrm{T}},$$

$$\boldsymbol{c}_{YY^{(k)}}=\left\{C_{yy^{(k)}}\left(x,x^1\right),\cdots,C_{yy^{(k)}}\left(x,x^m\right)\right\}^{\mathrm{T}},$$

$$\boldsymbol{C}=\begin{Bmatrix}\boldsymbol{C}_{YY} & \boldsymbol{C}_{YY^{(1)}}^{\mathrm{T}} & \cdots & \boldsymbol{C}_{YY^{(k)}}^{\mathrm{T}} & \cdots & \boldsymbol{C}_{YY^{(n)}}^{\mathrm{T}}\\ \boldsymbol{C}_{YY^{(1)}} & \boldsymbol{C}_{Y^{(1)}Y^{(1)}} & \cdots & \vdots & \cdots & \vdots\\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots\\ \boldsymbol{C}_{YY^{(k)}} & \cdots & \cdots & \boldsymbol{C}_{Y^{(k)}Y^{(k)}} & \cdots & \vdots\\ \vdots & \vdots & \vdots & \vdots & \ddots & \boldsymbol{C}_{Y^{(n-l)}Y^{(n)}}^{\mathrm{T}}\\ \boldsymbol{C}_{YY^{(n)}} & \cdots & \cdots & \cdots & \boldsymbol{C}_{Y^{(n-l)}Y^{(n)}} & \boldsymbol{C}_{Y^{(n)}Y^{(n)}}\end{Bmatrix},$$

$$\boldsymbol{C}_{YY}=\begin{Bmatrix}C_{yy}\left(x^1,x^1\right) & \cdots & C_{yy}\left(x^1,x^m\right)\\ \vdots & \ddots & \vdots\\ C_{yy}\left(x^m,x^1\right) & \cdots & C_{yy}\left(x^m,x^m\right)\end{Bmatrix},$$

$$\boldsymbol{C}_{YY^{(k)}}=\begin{Bmatrix}C_{yy^{(k)}}\left(x^1,x^1\right) & \cdots & C_{yy^{(k)}}\left(x^1,x^m\right)\\ \vdots & \ddots & \vdots\\ C_{yy^{(k)}}\left(x^m,x^1\right) & \cdots & C_{yy^{(k)}}\left(x^m,x^m\right)\end{Bmatrix},$$

$$\boldsymbol{C}_{Y^{(i)}Y^{(j)}}=$$

$$\begin{Bmatrix}C_{y^{(i)}y^{(j)}}\left(x^1,x^1\right) & \cdots & C_{y^{(i)}y^{(j)}}\left(x^1,x^m\right)\\ \vdots & \ddots & \vdots\\ C_{y^{(i)}y^{(j)}}\left(x^m,x^1\right) & \cdots & C_{y^{(i)}y^{(j)}}\left(x^m,x^m\right)\end{Bmatrix}\ (i,j,k\in[1,n]).$$

In eq. (12), $\boldsymbol{Y}$ represents the $(m+mn)\times 1$ random variable vector which is composed of the values of the random function and its partial derivatives at the sample points, $\boldsymbol{f}$ represents the value of $p\times 1$ regression function vector at x, $\boldsymbol{\beta}$ is the unknown regression coefficient of $p\times 1$, $\boldsymbol{F}$ is the $(m+mn)\times p$ regression function matrix, $\boldsymbol{c}$ is the covariance function vector of $x_{(m+mn)\times 1}$ and the sample points, and $\boldsymbol{C}$ is the covariance function matrix of the $(m+mn)\times(m+mn)$ sample points. Because c and $\boldsymbol{C}$ both include σ^2, we let

$$\boldsymbol{c}=\sigma^2\boldsymbol{r},\ \boldsymbol{C}=\sigma^2\boldsymbol{R}.$$

Then

$$\begin{pmatrix} y \\ \boldsymbol{Y} \end{pmatrix} \sim N_{1+m+mn}\left(\begin{pmatrix} \boldsymbol{f}^{\mathrm{T}} \\ \boldsymbol{F} \end{pmatrix}\boldsymbol{\beta},\sigma^2\begin{pmatrix} 1 & \boldsymbol{r}^{\mathrm{T}} \\ \boldsymbol{r} & \boldsymbol{R} \end{pmatrix}\right). \quad (13)$$

Eqs. (4) and (13) are represented in the same form while the dimensions of $\boldsymbol{Y}$, $\boldsymbol{r}$ column vectors increase from m to $(m+mn)$; the dimension of matrix $\boldsymbol{F}$ increases from $m\times p$ to $(m+mn)\times p$, and the dimension of matrix $\boldsymbol{R}$ increases from $m\times m$ to $(m+mn)\times(m+mn)$. Similar to the deduction of Kriging approximation function, GBK approximation function is obtained as

$$\begin{cases} \hat{y}(x)=\boldsymbol{f}^{\mathrm{T}}(x)\boldsymbol{\beta}+\boldsymbol{r}^{\mathrm{T}}(x)\boldsymbol{R}^{-1}\left(\boldsymbol{Y}-\boldsymbol{F}\boldsymbol{\beta}\right), \\ \boldsymbol{\beta}=\left(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{R}^{-1}(\hat{\theta})\boldsymbol{F}\right)^{1}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{R}^{-1}(\hat{\theta})\boldsymbol{Y}^{n}, \\ \sigma^2=\dfrac{1}{n}\left(\boldsymbol{Y}-\boldsymbol{F}\boldsymbol{\beta}\right)^{\mathrm{T}}\boldsymbol{R}^{-1}(\hat{\theta})\left(\boldsymbol{Y}-\boldsymbol{F}\boldsymbol{\beta}\right), \end{cases} \quad (14)$$

where the maximum likelihood estimator $\hat{\theta}$ of correlation parameter $\boldsymbol{\theta}$ is the extremum of the following optimization problem

$$\left\{\hat{\sigma}^2(\hat{\theta})\left|\boldsymbol{R}(\hat{\theta})\right|^{1/n}\right\}=\min_{\boldsymbol{\theta}\in\Theta}\left\{\hat{\sigma}^2(\boldsymbol{\theta})\left|\boldsymbol{R}(\boldsymbol{\theta})\right|^{1/n}\right\}$$

$$=\min_{\boldsymbol{\theta}\in\Theta}\left\{\frac{1}{n}\left(\boldsymbol{Y}-\boldsymbol{F}\boldsymbol{\beta}\right)^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{\theta})\left(\boldsymbol{Y}-\boldsymbol{F}\boldsymbol{\beta}\right)\left|\boldsymbol{R}(\boldsymbol{\theta})\right|^{1/n}\right\}. \quad (15)$$

4 Realization of GBK approximate model

Directly calculating eq. (14) needs the inverse matrix $\boldsymbol{R}^{-1}$ of correlation matrix $\boldsymbol{R}$ of $(m+mn)\times(m+mn)$ dimensions, which is a complex calculation process that might lead to great computing errors. In addition, inappropriate selection of the sample points may result in the ill-condition of correlation matrix. Provided that correlation function matrix is symmetrical and sparse, with correlation function matrix $\boldsymbol{R}$ being positive definite, the Cholesky factorization is as follows

$$\begin{cases} \boldsymbol{R}=V\Lambda V^{\mathrm{T}}=\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}, \\ VV^{\mathrm{T}}=\boldsymbol{I}, \\ \boldsymbol{C}=V\Lambda^{1/2}, \\ \Lambda=\mathrm{diag}\left(\lambda_1,\cdots,\lambda_{m+mn}\right), \end{cases} \quad (16)$$

where λ_i is the eigenvalue of $\boldsymbol{R}$, $\boldsymbol{C}$ is a lower-triangular matrix, and $\boldsymbol{I}$ is an unit matrix.

The orthogonal transformation matrix of $\boldsymbol{F}$ is

$$\boldsymbol{F}_t=\boldsymbol{C}^{-1}\boldsymbol{F}, \quad (17)$$

and Householder transform for upper-triangular decomposing to $\boldsymbol{F}_t$ is

$$\boldsymbol{F}_t=\boldsymbol{Q}\boldsymbol{G}, \quad (18)$$

where $\boldsymbol{Q}$ is $(m+mn)\times p$ orthogonal matrix and $\boldsymbol{G}$ is a $p\times p$ upper-triangular matrix. Substituting eqs. (16)－(18) into eq. (14), the regression coefficient, stationary Gaussian random function variance and GBK approximate model can be computed as

$$\begin{cases} \hat{\boldsymbol{\beta}}=\boldsymbol{G}^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{Y}, \\ \hat{\sigma}_z^2=\dfrac{1}{m}\left(\boldsymbol{C}^{-1}\boldsymbol{Y}-\boldsymbol{F}_t\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}\left(\boldsymbol{C}^{-1}\boldsymbol{Y}-\boldsymbol{F}_t\hat{\boldsymbol{\beta}}\right), \\ \hat{y}(x)=\boldsymbol{f}^{\mathrm{T}}(x)\hat{\boldsymbol{\beta}}+\boldsymbol{r}^{\mathrm{T}}(x)\boldsymbol{C}^{-\mathrm{T}}\left(\boldsymbol{C}^{-1}\boldsymbol{Y}-\boldsymbol{F}_t\hat{\boldsymbol{\beta}}\right). \end{cases} \quad (19)$$

To avoid errors from different quantification levels of the training samples, the training sample data $(\boldsymbol{X},\boldsymbol{Y}^{(0)}, \boldsymbol{Y}^{(1)},\cdots,\boldsymbol{Y}^{(\mathrm{n})})$ should be unified into $(\boldsymbol{U},\boldsymbol{V}^{(0)},\boldsymbol{V}^{(1)},\cdots, \boldsymbol{V}^{(\mathrm{n})})$:

$$\begin{cases} u_i^j=\dfrac{x_i^j-\mu_{x_i}}{\sigma_{x_i}},v^j=\dfrac{y^j-\mu_y}{\sigma_y}, \quad \left(j\in[1,m]\right), \\ v^{(k)j}=\dfrac{\dfrac{\partial y}{\partial x_k}\left(x^j\right)-\mu_{y^{(k)}}}{\sigma_{y^{(k)}}}, \quad \left(i,k\in[1,n]\right), \end{cases} \quad (20)$$

where μ_{x_i}, σ_{x_i}, μ_y, σ_y and $\mu_{y^{(k)}}$, $\sigma_{y^{(k)}}$ are the average and standard variances of x_i, y and $\partial y/\partial x_k$, respectively.

Similar to the standard Kriging approximation function, to compute GBK model needs to obtain the maximum likelihood estimator $\hat{\theta}$ of correlation function parameter $\boldsymbol{\theta}$. A common method for obtaining the optimal value of correlation function parameter is to use optimization techniques. Since we couple the GBK approximate model into the optimization process and construct a hybrid optimization method, the optimization model includes internal and external hierarchies: the internal hierarchy optimization determines the optimal value of correlation function parameter, and the external hierarchy optimization determines the optimal design variable value. In order to overcome computing complexity resulted by two hierarchy optimization, the Latin Hypercube Sample in ref. [16] with fixed sample number is adopted instead of optimization methods using convergence criteria to obtain the minimal value of the target of maximum likelihood estimator, and the optimal correlation function parameter is determined.

5 Hybrid optimization method with GBK approximate model

The control flow of the hybrid optimization method with GBK approximate model is shown in Figure 1.

The main steps include the following.

Step 1: construct initial GBK approximate model. If there are no sample data, then it is necessary to design the sample points and call the system analysis model to obtain function value and first order derivatives at the sample points by finite difference.

Step 2: design optimization based on GBK approximate model. Sequence quadratic programming (SQP) is adopted in the optimization algorithm.

Step 3: compute the optimal design result of the approximated optimization.

Step 4: judge whether the whole optimization process is convergent. End the optimization if the process is convergent, otherwise add the optimal design point of approximation optimization into sample space, calculate the distances among all the sample points, and select the midpoint of the two points corresponding to the largest distance as the new sample point.

Step 5: calculate the value and gradient of new added sample points.

Step 6: replace the original GBK approximate model with the new GBK model based on increased sample data, and repeat step 2.

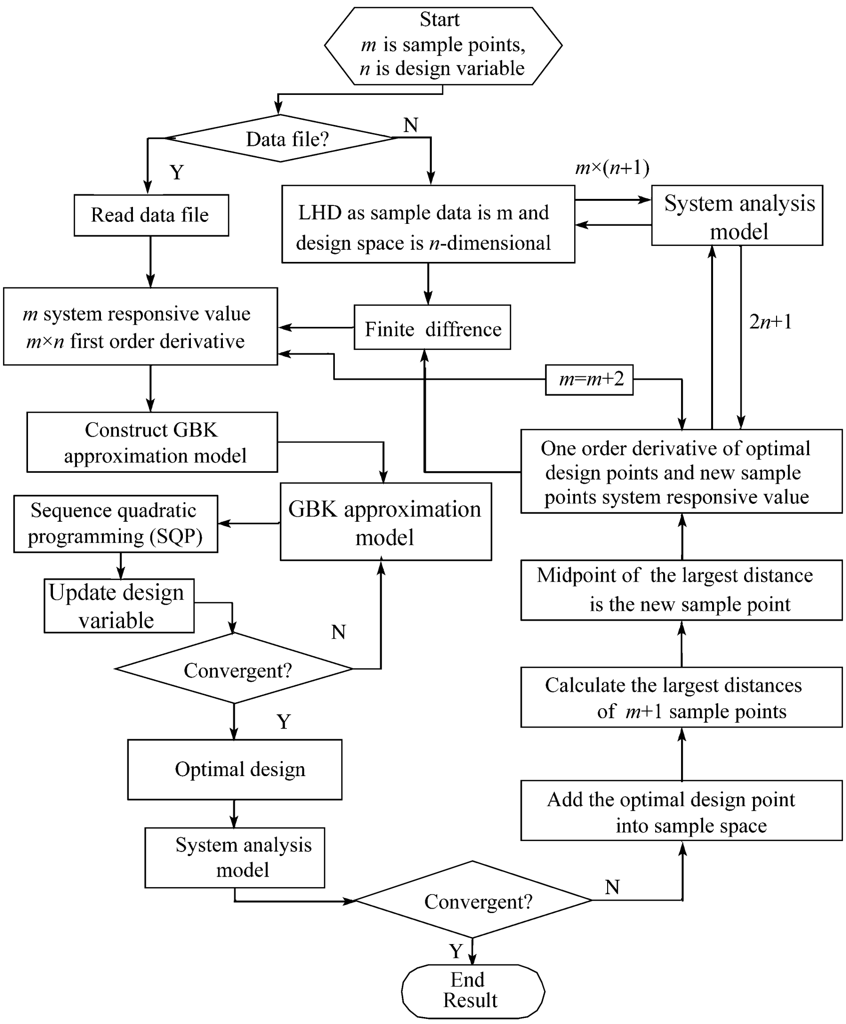

Figure 1 Control flow of the hybrid optimization method with GBK approximate model.

6 Example of aircraft aerodynamic shape optimization design

Take the aerodynamic shape optimization design of reentry vehicle as an example to confirm the precision and efficiency of GBK model. The reentry vehicles adopt the aerodynamic shape of the lifting body to improve the maneuverability, for example, Common Aero Vehicle (CAV) X-41 and maneuver warhead of Pershing II. Suppose that the aerodynamic configuration of the researched lifting body reentry aircraft adopts an elliptic cross section biconical body with cruciform flaps. The aerodynamic shape and size of the reentry vehicle are as shown in Figure 2.

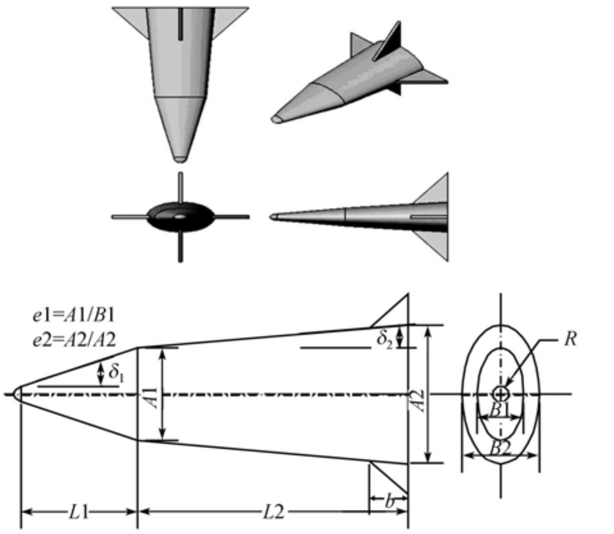

Figure 2 Aerodynamic shape and size of the reentry vehicle.

Given the length and maximum coved circle diameter, the optimization design variables of the reentry vehicle aerodynamic shape were selected as follows: Nosed radius R, frustum length $L1$, frustum half-angle $\delta 1$ & $\delta 2$, ellipse section eccentricities $e1$ & $e2$ and rudder radix chord b.

In the optimization problem, C_L/C_D was selected as the objective function:

$$object = \max \frac{1}{9}\sum_{j=1}^{3}\sum_{i=1}^{3}\frac{C_L}{C_D}\left(M_{\infty j}, \alpha_i\right),$$

$$\text{s.t.} \quad \begin{aligned} & M_{\infty 1}=10,\ M_{\infty 2}=12,\ M_{\infty 3}=14, \\ & \alpha_1=8^\circ,\ \alpha_2=9^\circ,\ \alpha_3=10^\circ. \end{aligned} \tag{21}$$

The constraints included:

1) The average value of the lift coefficients is greater than 3.0,

$$G_1 = \frac{1}{3}\sum_{j=1}^{3} C_y(M_{\infty j}, \alpha = 20^\circ) \geqslant 3.0; \tag{22}$$

2) the centre of pressure is greater than 0.55 and meets the stationary stability requirements,

$$G_2 = \frac{1}{9}\sum_{j=1}^{3}\sum_{i=1}^{3} x_{cp}(M_{\infty j}, \alpha_i) \geqslant 0.55; \tag{23}$$

3) the sizes of the vehicle meet certain loading demands,

$$\begin{aligned} G_3 &= L_2 \geqslant 3.2, \\ G_4 &= B_1 \geqslant 0.3, \\ G_5 &= B_2 \geqslant 0.5; \end{aligned} \tag{24}$$

4) the outer diameter of the missile body is less than 1.0,

$$G_6 = A_2 \leqslant 1.0. \tag{25}$$

The numerical method based on N-S equations and unstructured mesh was applied to calculate the vehicle aerodynamic coefficients. MUSCL finite-volume schemes were applied, and van-Leer flux vector splitting scheme was used to solve the Riemann problem on the boundary surface.

The hybrid optimization method with GBK approximate model is shown in Figure 1. Before constructing the initial GBK approximate model, the initial design space needed to be constructed first, of which sample quantity was m. When m=21, the initial design space was obtained through using LHD method as shown in Table 1.

To construct aerodynamic force coefficient GBK approximate model of design variables R, $L1$, $\delta 1$, $\delta 2$, $e1$, $e2$ and b, first order derivatives of design variables needed to be calculated. Since the numerical method for aerodynamic force was complicated, the first-order derivative with respect to design variable of aerodynamic force coefficient y was obtained by using forward difference in the paper. And the material method was as below:

$$\begin{cases} y'(R) = \dfrac{y(R+\Delta R) - y(R)}{\Delta R}, \\ y'(L1) = \dfrac{y(L1+\Delta L1) - y(L1)}{\Delta L1}, \\ y'(\delta 1) = \dfrac{y(\delta 1+\Delta \delta 1) - y(\delta 1)}{\Delta \delta 1}, \\ y'(\delta 2) = \dfrac{y(\delta 2+\Delta \delta 2) - y(\delta 2)}{\Delta \delta 2}, \\ y'(e1) = \dfrac{y(e1+\Delta e1) - y(e1)}{\Delta e1}, \\ y'(e2) = \dfrac{y(e2+\Delta e2) - y(e2)}{\Delta e2}, \\ y'(b) = \dfrac{y(b+\Delta b) - y(b)}{\Delta b}, \end{cases} \tag{26}$$

Table 1 Initial design space

m	R (mm)	$L1$ (mm)	$\delta 1$ (°)	$\delta 2$ (°)	$e1$	$e2$	b (mm)
1	20	700	14.55	5.6	2.4	1.1	650
2	24	1050	11.85	3.8	3.6	1	740
3	28	950	8.7	2	2.2	1.65	620
4	32	1450	6.45	3	2.9	1.35	800
5	36	1100	6.9	4.4	3.3	1.25	710
6	40	1350	14.1	2.2	3.1	1.7	560
7	44	1400	9.6	3.6	3.4	1.85	665
8	48	1250	6	6	3	1.4	680
9	52	1500	13.65	5	3.7	1.3	605
10	56	1000	13.2	2.8	3.2	1.05	515
11	60	1200	7.35	5.4	4	1.55	770
12	64	800	12.3	3.2	2.6	1.6	500
13	68	600	11.4	2.4	2.8	1.95	575
14	72	1300	10.95	4.8	3.5	1.45	635
15	76	750	8.25	2.6	2.1	1.5	725
16	80	500	10.5	4	2.3	1.9	755
17	84	1150	12.75	4.6	3.8	1.75	530
18	88	900	10.05	5.2	2.7	1.15	545
19	92	550	15	5.8	2.5	1.8	695
20	96	650	9.15	4.2	3.9	1.2	785
21	100	850	7.8	3.4	2	2	590

where ΔR, $\Delta L1$, $\Delta\delta 1$, $\Delta\delta 2$, $\Delta e1$, $\Delta e3$, Δb are the respective small disturbance quantities of the design variables, and their values are 10% of the present design variables. The design space for the first order derivative calculation at the first sample point of initial design space is shown in Table 2.

Table 2 Design space for first order derivative calculation

R (mm)	$L1$ (mm)	$\delta 1$ (°)	$\delta 2$ (°)	$e1$	$e2$	b (mm)
20	700	14.55	5.6	2.4	1.1	650
22	700	14.55	5.6	2.4	1.1	650
20	**770**	14.55	5.6	2.4	1.1	650
20	700	**16.01**	5.6	2.4	1.1	650
20	700	14.55	**6.16**	2.4	1.1	650
20	700	14.55	5.6	**2.64**	1.1	650
20	700	14.55	5.6	2.4	**1.21**	650
20	700	14.55	5.6	2.4	1.1	**715**

The system analysis model was called 168 (i.e. 21×(7+1)) times to obtain the initial GBK approximate model of aerodynamic coefficients of the reentry vehicle. Since complicated numerical value calculation method was adopted in the system analysis model, three distributed computers were used to compute aerodynamic coefficient at the 168 sample points to improve efficiency of calculating.

The optimization process converged after 5 iterations and the objective function value i.e. the optimal design average lift to drag ratio was 2.934. The optimization design results of GBK approximate model are shown in Table 3 and the optimal aerodynamic shape of the reentry vehicle is shown in Figure 3. The convergence process is shown in Figure 4.

Table 3 Optimization design results of GBK approximate model

Para-meter	R (mm)	$L1$ (mm)	$\delta 1$ (°)	$\delta 2$ (°)	$e1$	$e2$	b (mm)	Average lift to drag ratio	Maximum average lift coefficient	Average pressure center coefficient
1	60.2	1245.6	9.56	4.53	2.61	1.69	721.5	2.819	3.154	0.573
2	56.9	1188.9	9.12	4.98	2.66	1.65	727.5	2.836	3.265	0.558
3	55.4	1233.2	9.64	5.01	2.55	1.46	742.1	2.895	3.137	0.561
4	52.4	1234.5	9.87	4.32	2.45	1.76	756.7	2.931	3.015	0.554
5	50.7	1173.9	9.98	4.52	2.25	1.62	768.5	2.934	3.112	0.553

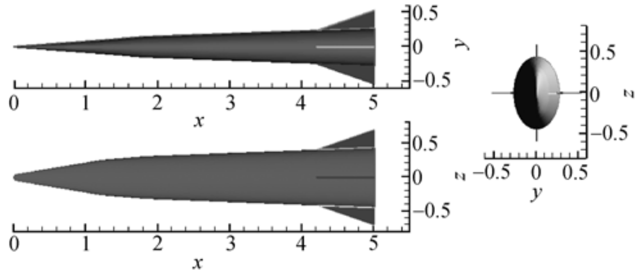

Figure 3 Optimal aerodynamic shape of the reentry vehicle.

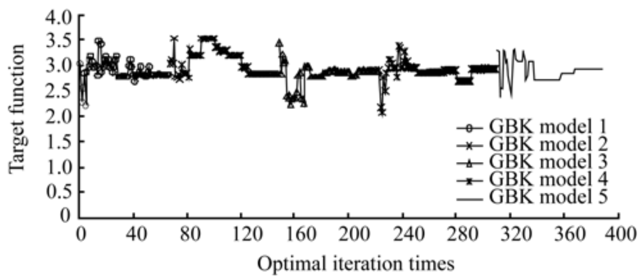

Figure 4 Convergence process of average lift to drag ratio.

7 Conclusions

1) The GBK approximation function establishes the Kriging approximation function based on m values and $m\times n$ first order partial derivatives of the sample points. The estimation precision can be improved by using gradient information.

2) Compared with the Kriging approximation function, the dimension of correlation function matrix R for GBK approximate model is increased by n^2+2n times. Since the inverse matrix of correlation function matrix is required, the realization of GBK approximate model in high-dimensional design space is complex.

3) The proposed GBK approximate model is based on n-dimensional first order partial derivatives. When the first partial derivatives with respect to certain input variables are comparatively small, they can be ignored in the construction of the GBK approximate model. This means that the GBK model is based on the first order partial derivatives of the input variables, thus the dimension of correlation function matrix R is reduced, and the running efficiency of model is improved.

4) The proposed hybrid optimization method based on GBK approximate model is suitable for optimization design problems with complex high fidelity system analysis model and lower dimension design space.

1 Sacks J, Schiller S B, Welch W J. Design for computer experiments. Technometrics, 1989, 31(1): 41－47

2 Dyck D N, Lowther D A. Response surface modeling of magnetic device performance using function value and gradient. Int J Appl Electrom, 1998, 9(2): 241－248

3 Sacks J, Welch W J, Mitchell T J, et al. Design and analysis of computer experiments. Stat Sci, 1989, 4(4): 409－435

4 Cressie N A C. Statistics for Spatial Data. New York: Wiley, 1993

5 Montes P. Smoothing noisy data by Kriging with nugget effects. In: Laurent P J, Peters A K, et al., eds. Int Conf Curves and Surfaces on Wavelets, Images, and Surface Fitting. Wellesley: Adams-Blake Publishing, 1994. 371－378

6 Cox D D, John S. SDO: A statistical method for global optimization. In: Alexandrov N M, Hussaini M Y, eds. Proc ICASE/NASA Langley Workshop on Multidisciplinary Optimization. Hampton: Society for Industrial and Applied Mathematics, 1995. 315－329

7 Torczon V, Trosset M W. Using approximations to accelerate engineering design optimization. In: Proc 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization. Reston: AIAA, 1998. 738－748

8 Trosset M W, Torczon V. Numerical Optimization Using Computer Experiments. Technical Report, Department of Computational and Applied Mathematics, Rice University, Houston, TR97-02. 1997

9 Simpson T W, Mauery T M, Korte J J, et al. Kriging models for global approximation in simulation-based multidisciplinary design optimization. AIAA J, 2001, 39(12): 2233－2241

10 Han Y Z, Gao H S, Li L Z, et al. Kriging model-based multidisplinary design optimization framework for turbine blade. J Aerospace Power, 2007, 22(7): 1055－1059

11 Jeong S, Obayashi S, Yamamoto K. Aerodynamic optimization design with Kriging model. T Jpn Soc Aeronaut Space Sci, 2005, 48(161): 161－168

12 Keane A J. Wing optimization using design of experiment, response surface, and data fusion methods. J Aircraft, 2003, 40(4): 741－750

13 Wang X F, Xi G. Aerodynamic optimization design for airfoil based on Kriging model (in Chinese). Acta Aeronautica et Astronautica Sinica, 2005, 26(5): 545－549

14 Wang S G, Shi J H, Yin S J, et al. Theory on Linear Model (in Chinese). Beijing: Science Press, 2004

15 Lu D. Random Process with Applications (in Chinese). Beijing: Tsinghua University Press, 1986

16 Mckay M D, Beckman R J, Conover W J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 1979, 21(2): 239－245