Efficient Response Surface Modeling using Sensitivity

Semyung Wang† and Chwail Kim*

민감도를 이용한 효율적인 반응표면모델생성

왕세명† · 김좌일*

Key Words : Response Surface Method(반응표면법), Sensitivity(민감도), Moving Least Squares Method(이동최소자승법), Parametric Study(인자 연구)

Abstract

The response surface method (RSM) became one of famous meta modeling techniques, however its approximation errors give designers several restrictions. Classical RSM uses the least squares method (LSM) to find the best fitting approximation models from the all given data. This paper discusses how to construct RSM efficiently and accurately using moving least squares method (MLSM) with sensitivity information. In this method, several parameters should be determined during the construction of RSM. Parametric study and optimization for these parameters are performed. Several difficulties during approximation processes are described and numerical examples are demonstrated to verify the efficiency of this method.

기호설명

R_I	Size of an approximation region
sw_g	Weighting factor for gradient error
L_{new}	Newly defined least squares function

1. Introduction

The response surface method (RSM) [1] is a popular meta modeling technique to handle large and complex systems. Since this method has several advantages, it is applied for many engineering applications these days. Because of approximation errors, however, this approach may not give accurate solutions. The conventional RSM is based on the approximation of scattered position data, and it could be easily obtained using the least squares method (LSM), a global approximation method. The LSM is one of the major causes of large approximation errors in the conventional RSM.

† 광주과학기술원, 기전공학과 교수
E-mail : smwang@kjist.ac.kr
TEL : (062)970-2429 FAX : (062)970-2384
* 광주과학기술원, 기전공학과

In this research, moving least squares method (MLSM) [2,3], a local approximation method, is adopted to reduce the approximation errors, and sensitivity information is included for the purpose of high efficiency and accuracy. This paper mainly discusses how to construct RS models efficiently and accurately using the moving least squares method (MLSM) and sensitivity.

In this proposed method, there are several parameters that can be selected by designer such as a size of local approximation region, weighting factor for gradient, and weighting functions. These parameters determine the accuracy of the approximation. In this research, the variations of the accuracy with respect to the change of the parameters are simulated and studied. The correlation concept between the sampled data and the estimated data is adopted as a criterion of the accuracy. If a correlation coefficient is close to 1.0, the RSM is well fitted. The criterion function is defined as the sum of two correlations for functions and gradients, and this value should be maximized. During construction of RSM, the related parameters are optimized to get the best RS model using a global optimization procedure such as a genetic algorithm. However, there exists several problems during the selection of parameters and those problems are described. Finally, some numerical examples are given to verify the effectiveness of this method.

2. Moving Least Squares Method With Sensitivity Information

2.1 Concept of Moving Least Squares Method

An advanced method for regression is MLSM. This method can be explained as a weighted LSM that has various weights with respect to the position of approximation. Therefore, coefficients of a RS model are functions of the location and they should be calculated for each location. This procedure is interpreted as a local approximation [4], and Fig. 1 explains the main concept of LSM and MLSM.

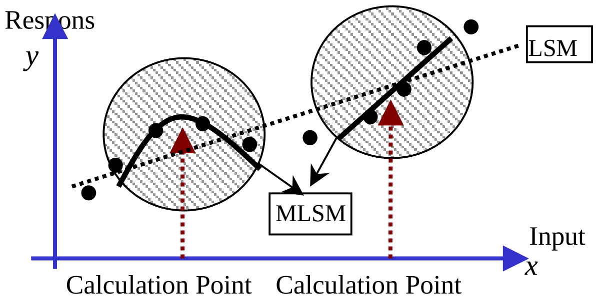


Fig. 1 Concept of LSM and MLSM

In the Fig.1, dotted curve is from the classical LSM. For the scattered data, only one best approximation curve can be obtained from the LSM. On the other hand in the case of MLSM, there exists an approximation function at a calculation point, which a designer wants to estimate, and there exists a different function at a different calculation point. Numerical derivation will be shown in the following section.

2.2 Numerical Expression of MLSM

Suppose there are n-response values, y_i, with respect to the changes of x_{ij} , which denote the i^{th} observation of variable x_j. Assume that the error term ε in the model has $E(\varepsilon)=0$, $Var(\varepsilon)=\sigma^2$ and that the { ε_i } are uncorrelated random variables.

The following matrix form can express the relationship between the responses and the variables

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}_y \qquad (1)$$

where $\mathbf{y}$ is a vector of the observations, $\mathbf{X}$ is a matrix of the level of the independent variables, $\boldsymbol{\beta}$ is a vector of the regression coefficients, and $\boldsymbol{\varepsilon}_\mathbf{y}$ is a vector of random errors.

A least squares function $\mathrm{L_y}(\mathbf{x})$ could be defined like following equation which is the sum of weighted errors.

$$\mathrm{L_y}(\mathbf{x})=\sum_{i=1}^{n}\mathrm{w}_i\varepsilon_i^{\,2}=\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\boldsymbol{\varepsilon}=(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\,\mathbf{W}(\mathbf{x})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \qquad (2)$$

Now, note that the diagonal weight matrix, $\mathbf{W}(\mathbf{x})$, is not a constant matrix in the MLSM. In other words, $\mathbf{W}(\mathbf{x})$ is a function of location, and it can be obtained by weighting functions. There are several kinds of weighting functions like linear, quadratic, high order polynomials, and exponential functions. For example, polynomial-weighting function is defined by

$$w(\mathbf{x}-\mathbf{x}_I)=w(d)=\begin{cases}1-6\left(\dfrac{d}{R_I}\right)^2+8\left(\dfrac{d}{R_I}\right)^3-3\left(\dfrac{d}{R_I}\right)^4, & for\ \dfrac{d}{R_I}\le 1\\[2ex] 0, & for\ \dfrac{d}{R_I}>1\end{cases} \qquad (3)$$

where $\mathbf{x}$ is a vector of approximation point, $\mathbf{x}_I$ is a vector of I^{th} sampling (or experiment) point, d is the distance between $\mathbf{x}$ and $\mathbf{x}_I$.

A weighting matrix, $\mathbf{W}(\mathbf{x})$, can be constructed using the weighting function in the diagonal terms. And, minimizing $\mathrm{L_y}(\mathbf{x})$ gives coefficients of the RS model of the form

$$\mathbf{b}(\mathbf{x})=\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\mathbf{y} \qquad (4)$$

Note that a procedure to calculate $\mathbf{b}(\mathbf{x})$ is a local approximation and "moving" process performs a global approximation through the whole design domain.

2.3 Moving Least Squares Method with Sensitivity

If the sensitivity (gradient) of each sampling point can be calculated efficiently [5], that sensitivity information can be used to construct RSM as well as function (response) data. For sensitivity information $y^d_{i,xj}$, the i^{th} gradient of y with respect to x_j, Eq.(1) leads the following relation.

$$\mathbf{y}^d_{xj}=\mathbf{T}_{xj}\boldsymbol{\beta}+\boldsymbol{\varepsilon}_{g\,xj} \qquad (5)$$

Where

$$\mathbf{y}^d_{xj}=\begin{bmatrix}y^d_{1xj}\\ y^d_{2xj}\\ \vdots\\ y^d_{nxj}\end{bmatrix},\quad \mathbf{T}_{xj}=\begin{bmatrix}0 & 0 & \cdots 1 & \cdots\\ 0 & 0 & \cdots 1 & \cdots\\ \vdots & & & \\ 0 & 0 & \cdots 1 & \cdots\end{bmatrix},\quad \boldsymbol{\beta}=\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_k\end{bmatrix},\quad \boldsymbol{\varepsilon}_{g\,xj}=\begin{bmatrix}\varepsilon_{g1xj}\\ \varepsilon_{g2xj}\\ \vdots\\ \varepsilon_{gnxj}\end{bmatrix}$$

$\mathbf{y}^d_{xj}$ represents the vector of gradient of the response with respect to x_j, $\mathbf{T}_{xj}$ defines the transformation matrix which represents gradient vector, and $\boldsymbol{\varepsilon}_{g\,xj}$ means vector of gradient error.

So the total weighted sum of squared errors of the gradient data can be written as

$$\mathrm{L_g}(\mathbf{x}) = \boldsymbol{\varepsilon}_{\mathrm{g}x_1}{}^{\mathrm{T}}\mathbf{W}_{\mathrm{g}}(\mathbf{x})\boldsymbol{\varepsilon}_{\mathrm{g}x_1} + \boldsymbol{\varepsilon}_{\mathrm{g}x_2}{}^{\mathrm{T}}\mathbf{W}_{\mathrm{g}}(\mathbf{x})\boldsymbol{\varepsilon}_{\mathrm{g}x_2} + \ldots + \boldsymbol{\varepsilon}_{\mathrm{g}x_{\mathrm{NDV}}}{}^{\mathrm{T}}\mathbf{W}_{\mathrm{g}}(\mathbf{x})\boldsymbol{\varepsilon}_{\mathrm{g}x_{\mathrm{NDV}}} \tag{6}$$

$\mathbf{W_g(x)}$can be constructed from the similar manner with the function case, but a different type of weighting function can be adopted.

Now, a new least squares function $L_{\mathrm{new}}(\mathbf{x})$, which contains the errors of gradient data as well as those of position data, can be defined by

$$\begin{aligned}\mathrm{L_{new}}(\mathbf{x}) &= (1-\mathrm{sw_g})\mathrm{L_y}(\mathbf{x}) + (\mathrm{sw_g})\mathrm{L_g}(\mathbf{x}) \\ &= (1-\mathrm{sw_g})\boldsymbol{\varepsilon}_{\mathrm{y}}{}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\boldsymbol{\varepsilon}_{\mathrm{y}} + (\mathrm{sw_g})\sum_{\mathrm{j=1}}^{\mathrm{NDV}}\boldsymbol{\varepsilon}_{\mathrm{gxj}}{}^{\mathrm{T}}\mathbf{W}_{\mathrm{g}}(\mathbf{x})\boldsymbol{\varepsilon}_{\mathrm{gxj}}\end{aligned} \tag{7}$$

where sw_{g} is a scale factor (or weighting factor) for gradient errors.

In order to minimize the new least squares function,

$$\left.\frac{\partial L_{new}}{\partial \boldsymbol{\beta}}\right|_{\mathbf{b}} = (1-sw_g)\left.\frac{\partial L_y}{\partial \boldsymbol{\beta}}\right|_{\mathbf{b}} + (sw_g)\left.\frac{\partial L_g}{\partial \boldsymbol{\beta}}\right|_{\mathbf{b}} = 0 \tag{8}$$

Substitution and rearrangement give

$$\begin{aligned}&\left((1-sw_g)\mathbf{X}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\mathbf{X} + sw_g\sum_{j=1}^{NDV}\mathbf{T}_{\mathbf{xj}}^{\mathbf{T}}\mathbf{W}_{\mathbf{g}}(\mathbf{x})\mathbf{T}_{\mathbf{xj}}\right)\mathbf{b} \\ &\quad= (1-sw_g)\mathbf{X}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\mathbf{y} + sw_g\sum_{j=1}^{NDV}\mathbf{T}_{\mathbf{xj}}^{\mathbf{T}}\mathbf{W}_{\mathbf{g}}(\mathbf{x})\mathbf{y}_{\mathbf{xj}}^{\mathbf{d}}\end{aligned} \tag{9}$$

Simplifying the above equation represents

$$\mathbf{A(x)b = c(x)} \tag{10}$$

And finally, the coefficients of the response surface model can be obtained of the form

$$\mathbf{b(x) = A(x)}^{-1}\mathbf{c(x)} \tag{11}$$

Through the sequences explained, a RS model that considers the gradient data as well as the function data can be obtained. Authors denote this RSM as sensitivity-based response surface model (SRSM). Note that the coefficients from the above sequences depend on the approximation location **x**. To verify the effectiveness of this proposed method, some examples will be demonstrated.

3. Parametric Study for Moving Least Squares Method with Sensitivity

3.1 Necessity of Parametric Study and Correlation Coefficient

When the sensitivity-based RSM is constructed, there are several parameters that can be selected by a designer such as RS model function (basis), type of weighting functions, size of an approximation region (R_I), and weighting factor for gradient (sw_g). These parameters are very important because they affect accuracy of the approximation. Parametric study is for checking how much these parameters affect a global accuracy of the approximation. Especially, the size of approximation (R_I) and weighting factor for gradient (sw_g) are the most important, and parametric studies of those parameters are performed.

As a measure of the accuracy for RSM, correlation coefficient is adopted. For two random variables $X1$, $X2$, a correlation coefficient is calculated by Eq.(14)

$$\begin{aligned}R_{X1X2} &= E\{(X1-m_1)(X2-m_2)\} \\ r_{X1X2} &= R_{X1.X2}/\sigma_1\sigma_2\end{aligned} \tag{12}$$

where $R_{X1\,X2}$ is a covariance of $X1$ and $X2$, m_{i} is the mean of X_{i}, σ_{i} is the standard deviation of X_{i}, and $r_{X1\,X2}$ is a correlation coefficient of $X1$ and $X2$.

The absolute value of the correlation coefficient can vary from 0 to 1. If it is 0, $X1$ and $X2$ are uncorrelated. If it is 1, $X1$ and $X2$ are perfectly correlated. In this research, a measure of the accuracy is the correlation coefficients of sampled data and estimated data from RSM.

There are several criteria for accuracy. An important reason why this correlation concept is adopted is a normalized (or equivalent) comparison between response error and gradient error. Since response and sensitivity values have different dimensions, direct comparison of those errors is impossible. However, by using the correlation concept, a normalized accuracy from 0 to 1 makes possible to compare the both errors.

Parametric studies for several mathematical functions are performed for several different conditions, and a few representative results will be shown in the following section. The following two sections show the results of parametric studies for a Rosenbrock test function, which is defined by

$$f(\mathbf{x}) = 100(x_1^{\,2} - x_2)^2 + (1-x_1)^2 \tag{13}$$

Each figure has 4 curves that represent

-Cor_Resp : Correlation coefficient of response values(15pts by LHC) for construction of RSM
-Cor_Sens : Correlation coefficient of sensitivity values(15pts by LHC) for construction of RSM
-(Cr+Cs)/2 : (Cor_Resp + Cor_Sens)/2
-Cor_TestPt : Correlation coefficient of response values(225 pts) for a test of global accuracy.

Cor_Resp and Cor_Sens are from the sampled data for construction of RSM, and Cor_TestPt is from the testing points that are many enough to represent a global accuracy. LHC means Latin Hypercube design.

3.2 Parametric Study for sw_g

The first parameter to study is sw_g. The following Fig.2 shows variations of correlation coefficient with respect to variations of sw_g. Since the trends of correlations are quite different for different weighting functions and other different conditions, a certain trend is not always the best. However, a few representative trends are found from the experiences.

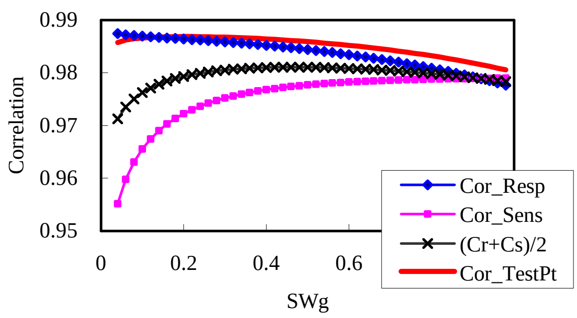


Fig.2 Correlation with respect to sw_g

Fig.2 is one of the general trends for sw_g. Larger *swg* (closer to 1) causes larger Cor_Sens and smaller Cor_Resp, because larger *swg* tries to minimize gradient errors more in Eq.(7). The important behavior is that the trend of (Cr+Cs)/2 is the most close to the trend of Cor_TestPt. This means that a maximum (Cr+Cs)/2 is the most close to the maximum global accuracy of RSM.

During the parametric study for *swg*, *swg* should larger than 0 and smaller than 1. 0 of *swg* means No sensitivity case (MLSM only) and 1 of *swg* means physically impossible case (no response data).

3.3 Parametric Study for R_I

R_I, a size of the local approximation region, is the second parameter to study. A small R_I makes an RSM approximation close to an interpolation which the RSM passes all sampled points. In this case, RSM can be very noisy and this noise phenomenon can lose a filtering effect which is one of the major advantages of RSM. Additionally, even though the number of data is not less than a minimum required number within an local approximation region that is determined by R_I, if R_I is too small, a matrix **A(x)** in Eq.(11) can be ill-conditioned. Eventually, ill-conditioned matrix operation causes very poor estimations. On the other hand, a large R_I makes the MLSM (local approximation) close to a conventional RSM (global approximation). Since R_I can affect the accuracy of approximation greatly, R_I is a very important parameter for the local approximation.

Figure 3 shows representative results of parametric studies for R_I. Minimum possible R_I leads maximum Cor_Resp, but generally that doesn't mean maximum global accuracy. As Fig.3 shows, consideration of both Cor_Resp and Cor_Sens can give good approximation because the profiles of (Cr+Cs)/2 and Cor_TestPt show a similar trend. Therefore, we have to maximize not Cor_Resp, but Cr+Cs in order to get a good RSM.

In other case, all 4 profiles can have the same trends. This result is good for approximation, because we don't have to worry about the accuracy criteria in this case. However, we have no idea whether the accuracy profiles will be like this or Fig.3.

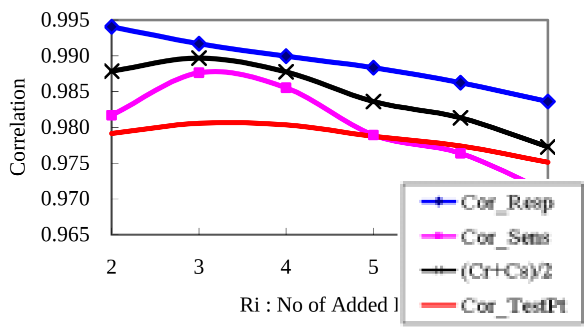


Fig.3 Correlation with respect to R_I

4. Optimization of Parameters for RS Modeling

During the construction of RSM, these parameters are optimized for the best RS modeling from an optimization procedure. Since a discontinuity problem can be occurred during this optimization procedure, a genetic algorithm [6] is adopted. One of the representative objectives of this parametric optimization is

maximize Corr_Resp + Corr_Sens
s.t. Approximation domain contains enough data (14)
(Matrix **A(x)** should not be singular)

5. Difficulties During the Approximation

There are several difficulties during the approximation.

5.1 Estimation at Near Boundary Points

At near a boundary, rack of the number of data within a local approximation region makes estimation poor or fails. Resizable approximation region or additional data near the boundary can solve this problem.

5.2 Near-Singular Problem

If a matrix **A(x)** in Eq.(10) is not singular but ill-conditioned, RSM estimation can be very poor. During matrix operations for calculating coefficients of RSM, an inverse matrix $\mathbf{A(x)}^{-1}$ in Eq.(11) becomes too sensitive, and this sensitive inverse operation cause the problem. In this research, a reciprocal condition number (Rcond) is adopted as a criterion to check the condition of **A(x)**. If a matrix **A** is well conditioned, Rcond(**A**) is near 1. If a matrix **A** is badly conditioned, Rcond (**A**) is near 0. For examples, Rcond of an identity matrix is 1, and Rcond of a singular matrix is 0.

In this research, if the reciprocal condition number (Rcond) of **A(x)** is smaller than a certain predefined value (for example, 0.0001), the parametric optimizer considers **A(x)** singular and finds larger Ri.

5.3 Selection of an Objective of Parametric Optimization

As the parametric study shows, the maximum of (Corr_Resp+Corr_Sens) doesn't guarantee maximum approximation accuracy. Therefore, a good selection of objective is important. Some possible alternatives are

Maximize *Cor_Sens s.t.Cor_Resp > Cor_Target(a given value)*
Maximize *Cor_Resp + Cor_Sens – (Cor_Resp - Cor_Sens)^2*
Minimize 1/(*Cor_Resp + Cor_Sens) + |Cor_Resp - Cor_Sens|*

However, more research is required to find other better objectives that can represent the globally maximum accuracy.

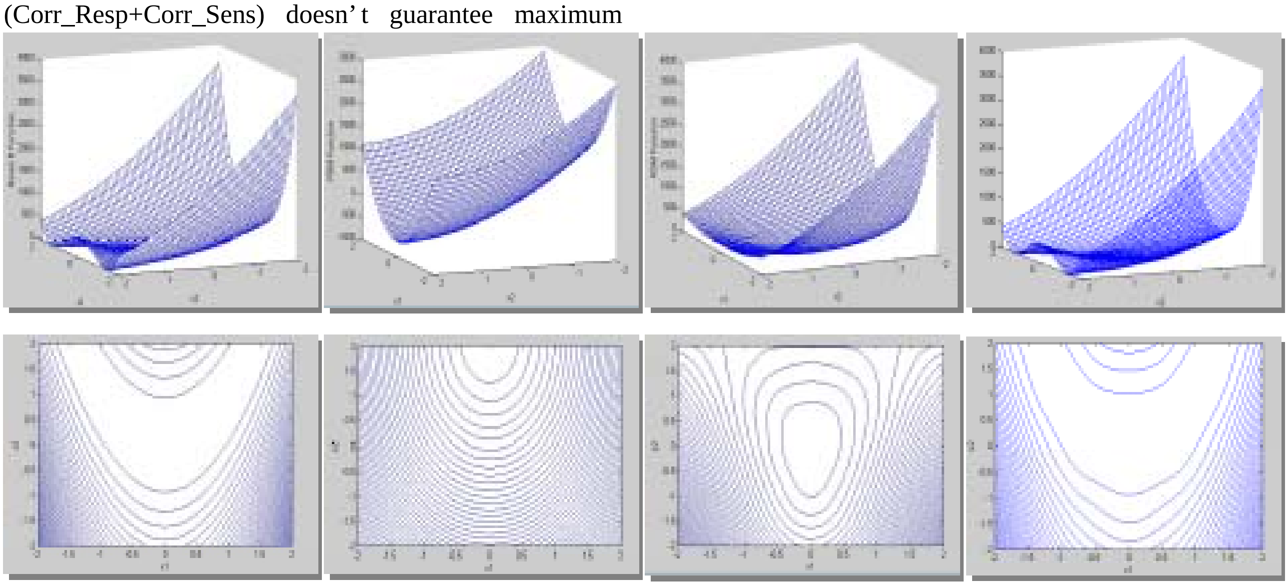
6. Numerical Examples

6.1 Function Test 1

The first mathematical example is Rosenbrock function with 2 variables famous for Banana function. As the contour plot shows in Fig.4 (a), this function has a long, narrow, parabolic shaped flat valley. Evenly distributed 16 points are sampled for experiments, and 100 points are selected for testing the accuracy of RSMs. Sensitivities at each sampling points are obtained analytically.

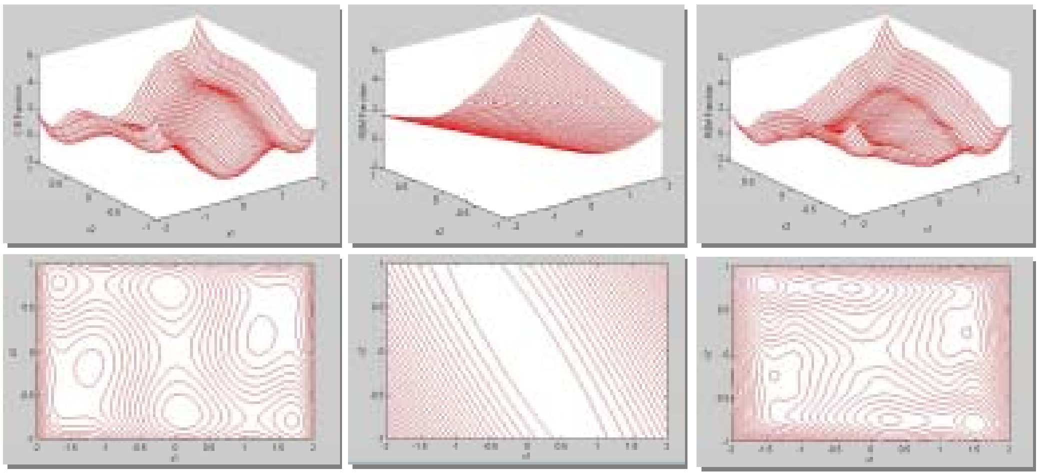
The Figs.4 (b~d) show several RS models constructed using different methods; the classical LSM, MLSM only, and MLSM with sensitivity performing parametric optimization, respectively. A genetic algorithm is used for the parametric optimization in the case of (d), and the objective is Eq.(14). Reciprocal condition number is also used to prevent a near-singular problem mentioned in the previous chapter.

In graphical point of view, Fig.4 (d) is very close to the original function and its contour plot shows the V-shaped valley. In numerical accuracy, case (d) also gave the most accurate solution.



(a) Original Function b) Classical LSM (c) MLSM (d) MLSM with Sensitivity

Fig.4 Function and contour Plots for RSMs using Different Methods



(a) Original Function (b) Classical LSM or MLSM (c) MLSM with Sensitivity

Fig.5 Function and contour Plots for RSMs using Different Methods

6.2 Function Test 2

The second test function is 2D six-hump camel back function, which is defined as

$$f(\mathbf{x}) = (4 - 2.1x_1^2 + x_1^4/3)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2 \quad (15)$$

Where $-2 \le x_1 \le 2 \, and \, -1 \le x_2 \le 1$

This function has 4 local optimums and 2 global optimums within the bounded region as shown in Fig.5 (a).

Evenly distributed sixteen points are sampled for experiments and one hundred points are selected for testing the accuracy of the RSM. Fig.5 (b~c) show the results of construction of RSM according to different methods, and graphically the case(c) can successfully describe the 6 optimum positions. The both cases of LSM and MLSM (without sensitivity) gave the same results.

7. Conclusions

The RSM became one of famous approximation and optimization techniques for complicate systems, however its approximation error is the major drawback of this approach. This paper mainly discussed how to construct RSM efficiently and accurately using sensitivity when the exact sensitivities were available. During the approximation using the moving least squares method (MLSM) with sensitivity information, several parameters should be determined carefully. Parametric study and optimization for these parameters, a weighting factor for gradient and size of local approximation region, were performed. Correlation coefficient and reciprocal condition number were adopted for better accuracy criteria, and a genetic algorithm was used for the parametric optimization. Several difficulties during applying the proposed method were described and numerical examples were demonstrated. From those examples, the proposed methods gave not only accurate but also efficient RS Models.

후 기

이 연구는 한국과학재단 지정 한양대학교 최적설계신기술연구센터의 연구비 지원으로 수행되었습니다.

참고문헌

(1) Myers, R. H., and Montgomery, D. C., 1995, "Response Surface Methodology: Process and Product Optimization Using Designed Experiments," John Wiley & Sons, New York.

(2) Lancaster, P., and Salkauskas, K., 1981, "Surface Generated by Moving Least Squares Methods," Mathematics of Computation, Vol. 37, No. 155, pp. 141-158.

(3) Viana, S. A., and Mesquita, R. C., 1999, "Moving Least Square Reproducing Kernel Method for Electromagnetic Field Computation," IEEE Transactions on Magnetics, Vol. 35, No. 3, pp. 1372-1375.

(4) Atkeson, C. G., Moore, A. W., and Schaal, S., 1997, "Locally Weighted Learning," AI Review, April 11:(1-5), pp. 11-73.

(5) Haug, E. J., Choi, K. K., and Komkov, V., 1986, "Design Sensitivity Analysis of Structural Systems," Academic Press, Orlando,.

(6) Gen, M., and Cheng, R., 1997, "Genetic Algorithms and Engineering Design," John Wiley & Sons, New York.