AIAA JOURNAL
Vol. 43, No. 11, November 2005

Efficient Response Surface Modeling by Using Moving Least-Squares Method and Sensitivity

Chwail Kim* and Semyung Wang†
Gwangju Institute of Science and Technology, Gwangju 500-712, Republic of Korea
and
Kyung K. Choi‡
University of Iowa, Iowa City, Iowa 52242

The response surface method (RSM) has currently become one of the better-known meta-modeling techniques. However, its approximation errors have placed several restrictions on designers as classical RSM uses the least-squares method (LSM) to find the best-fitting approximation models from the given function data. We discuss how to construct RS models efficiently and accurately using the moving least-squares method (MLSM) combined with sensitivity information. The formulations for incorporating the sensitivity using the MLSM are derived. With this method, several parameters should be determined during the construction of response surfaces. The parametric study and optimization for these parameters are performed. However, because of the discontinuity problem of the optimization, a genetic algorithm is adopted. The correlation coefficient is used for the normalized comparison between the function and gradient errors. Also, the reciprocal condition number is applied to avoid ill-conditioned approximations. Several difficulties and their respective solutions during the approximation processes are described, and the numerical examples are then demonstrated to verify the accuracy and the efficiency of this method. If the sensitivity of each sampling point can be calculated efficiently by utilizing a cheap computation, the proposed method is recognized as very efficient and accurate.

Nomenclature

$\boldsymbol{b}$	=	least-squares estimator of $\boldsymbol{\beta}$ (coefficient vector of response surface method)
d	=	distance between $\boldsymbol{x}$ and $\boldsymbol{x}_I$
k	=	degree of freedom of regression
L_y, L_g, L_{new}	=	least-squares function of experiment (function), gradient, and total data
m_i	=	mean of a random variable X_i
n	=	number of experiment
p	=	number of parameter including the constant term $(k+1)$
R_I	=	size of an approximation region
$R_{X1\,X2}$	=	covariance of random variable $X1$ and $X2$
$r_{X1\,X2}$	=	correlation coefficient of random variable $X1$ and $X2$
sw_g	=	weighting factor for gradient error in L_{new}
$\boldsymbol{T}_{xj}$	=	gradient transformation matrix that represents gradient vector
$\boldsymbol{W}(\boldsymbol{x}), \boldsymbol{W}_g(\boldsymbol{x})$	=	weighting matrix for function and sensitivity at location $\boldsymbol{x}$
$w(d), w_g(d)$	=	weighting function of function and sensitivity data at distance d
$\boldsymbol{X}$	=	$n \times p$ matrix of the level of the independent variables
$\boldsymbol{x}$	=	vector of approximation location
$\boldsymbol{x}_I$	=	vector of Ith sampling (or experiment) point
x_j	=	jth independent variable
xn_j	=	jth normalized independent variable $(-1 \sim 1)$
$\boldsymbol{y}$	=	$n \times 1$ vector of the observations
$\boldsymbol{y}^d_{xj}$	=	vector of gradient of the response with respect to x_j
$\hat{\boldsymbol{y}}$	=	vector of fitted value from the response surface model
$\boldsymbol{\beta}$	=	$p \times 1$ vector of the regression coefficients
$\boldsymbol{\varepsilon}_y$	=	$n \times 1$ vector of the errors of responses
$\boldsymbol{\varepsilon}_{gxj}$	=	vector of the errors of gradients with respect to x_j
σ_i	=	standard deviation of a random variable X_i

Received 22 July 2004; revision received 23 May 2005; accepted for publication 26 May 2005.

*Postdoc, Department of Mechatronics, 1 Oryong-dong, Buk-gu.
†Professor, Department of Mechatronics, 1 Oryong-dong, Buk-gu; smwang@gist.ac.kr. Senior Member AIAA.
‡Professor and Director, Center for Computer-Aided Design.

Introduction

THE response surface method (RSM)[1] is a popular meta-modeling technique that has the ability to handle large and complex systems. Because this method has several advantages, it is commonly utilized in many engineering applications. Recently, multidisciplinary design optimization and reliability-based design optimization (RBDO) methodologies have become a central feature in engineering problems. However, these methods require much computational time, and the RSM is a good tool for reducing the computation time. Therefore, RS methods have been researched and applied to RBDO.[2,3]

However, the approximation errors inherent in RSM and other approximation approaches might not give accurate solutions, and therefore many researchers have contributed to reduce the approximation errors. Recently, gradient-enhanced meta-modeling methods, which use gradient (sensitivity) values as well as function values, have been studied in order to decrease these errors as well as the computation time. Sellar and Batill[4] and Liu and Batill[5] developed the gradient-enhanced neural-network training method. Lauridsen et al.[6] researched the gradient-enhanced response surface (GERS) building method and developed iterative weighted least squares. At the same time, Kim et al.[7] researched the RSM using the moving least-squares method (MLSM) to incorporate sensitivity, and defined a formulation for the weighted error function. Van Keulen and Vervenne[8] updated the previous research on GERS and proposed a two-step blending approach and an iterative single-step approach. Additionally, they discussed the effect of the design of experiment. Liu and Batill[9] incorporated sensitivity into the kriging model, and they discussed some additional issues related to their method. Kim et al.[10] updated their sensitivity-based method. From

the parametric study, they studied the behavior of the accuracy according to the parameters and performed the optimization of the parameters. However, because of the discontinuity problem of the optimization, a genetic algorithm was adopted. Also, additional important improvements in this research field were developed. The correlation coefficient was adopted for the normalized comparison between the function and gradient errors, and the reciprocal condition number was applied to avoid ill-conditioned approximations. More details are listed in the following sections.

As conventional RSM is based on the approximation of scattered position data, obtained using the least-squares method (LSM), a global approximation method. LSM is one of the major causes of the large approximation errors in the conventional RSM.

In this research, the MLSM,[11,12] a local approximation method, is adopted in order to reduce the approximation errors, and sensitivity information is included for the purpose of high efficiency and approximation accuracy. This paper discusses how to construct RS models efficiently and accurately using the MLSM and sensitivity.

In this proposed method, there are several parameters that can be selected by the designer such as the size of the local approximation region, the weighting factor for the gradient, and the weighting functions. These parameters combine to 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 for accuracy. If a correlation coefficient is close to 1.0, the RSM is well fitted. In this, the criterion function is defined as the sum of two correlations for functions and gradients, and this value should be maximized. During the construction of the response surface model, the related parameters are optimized to obtain the best RS model using a global optimization procedure such as the genetic algorithm. However, several problems exist during the selection of parameters, and those problems will be described in detail. Finally, some numerical examples are given to verify the effectiveness of this method.

MLSM with Sensitivity Information

Concept of MLSM

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, the coefficients of a RS model are functions of the location and thus should be calculated for each location. This procedure is interpreted as a local approximation,[13] and Fig. 1 explains the main concepts of LSM and MLSM.

In Fig. 1, the dotted curve is obtained from classical LSM. For the scattered data, only one best approximation curve can be obtained. Conversely, with MLSM there exists one approximation function at one calculation point, and there exists a subsequent function at each different calculation point. Therefore, the coefficients of the RS model are not constants but are variables of the calculation position. This locally weighted approximation can be performed from the consideration of effective data near the calculation location, and the data can be weighted according to the distance from the calculation location. The numerical derivation will be shown in the following section.

Derivation of MLSM

Suppose there are n-response values y_i with respect to the changes of x_{ij}, which denote the ith 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 $\{\varepsilon_i\}$ are uncorrelated random variables.

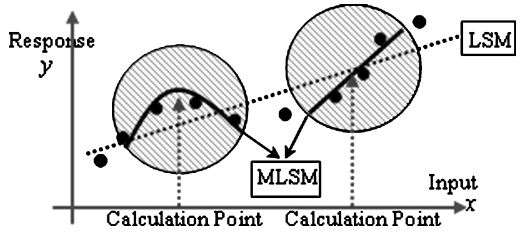


Fig. 1 Concept of LSM and MLSM.

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

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_y \tag{1}$$

The least-squares function $L_y(\boldsymbol{x})$ could be defined as in the following equation, which is the sum of weighted errors:

$$L_y(\boldsymbol{x}) = \sum_{i=1}^{n} w_i \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{W}(\boldsymbol{x})\boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W}(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{2}$$

Now, note that the diagonal weighting matrix $\boldsymbol{W}(\boldsymbol{x})$ is not a constant matrix in the MLSM. In other words, $\boldsymbol{W}(\boldsymbol{x})$ is a function of the location, and it can be obtained by utilizing weighting functions. There are several kinds of weighting functions such as constant, linear, quadratic, high-order polynomials, and exponential functions, which are defined by Eq. (3):

$$w(\boldsymbol{x} - \boldsymbol{x}_1) = w(d)$$

$$= \begin{cases} \text{if} \quad d/R_I \le 1, & \begin{cases} \text{constant} & 1 \\ \text{linear} & 1 - d/R_I \\ \text{quadratic} & 1 - (d/R_I)^2 \\ \text{4th-poly} & 1 - 6(d/R_I)^2 + 8(d/R_I)^3 - 3(d/R_I)^4 \\ \text{exponential} & \exp(-d/R_I) \end{cases} \\ \text{if} \quad d/R_I > 1, 0 \end{cases} \tag{3}$$

For example, a fourth-order polynomial weighting function is expressed by a bell-shaped figure. The weighting function has 1 (maximum value) at 0 normalized distance and 0 (minimum value) outside of 1 normalized distance, that is, $w(0)=1$, $w(d/R_I>1)=0$. Also the function decreases smoothly from 1 to 0.

Eventually, a weighting matrix $\boldsymbol{W}(\boldsymbol{x})$ can be constructed by using the weighting function in diagonal terms:

$$\boldsymbol{W}(\boldsymbol{x}) = \begin{bmatrix} w(\boldsymbol{x}-\boldsymbol{x}_1) & 0 & \dots & 0 \\ 0 & w(\boldsymbol{x}-\boldsymbol{x}_2) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & w(\boldsymbol{x}-\boldsymbol{x}_n) \end{bmatrix} \tag{4}$$

To minimize $L_y(\boldsymbol{x})$, the least-squares estimators must satisfy

$$\left.\frac{\partial L_y(\boldsymbol{x})}{\partial \boldsymbol{\beta}}\right|_b = -2\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{x})\boldsymbol{y} + 2\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{x})\boldsymbol{X}\boldsymbol{b} = 0 \tag{5}$$

Coefficients of the RS model $\boldsymbol{b}(\boldsymbol{x})$ can be obtained by the matrix operation:

$$\boldsymbol{b}(\boldsymbol{x}) = [\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{x})\boldsymbol{X}]^{-1}\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{x})\boldsymbol{y} \tag{6}$$

In this step, it is important to note that the coefficient $\boldsymbol{b}(\boldsymbol{x})$ is a function of the location or position $\boldsymbol{x}$. Note that a procedure to calculate $\boldsymbol{b}(\boldsymbol{x})$ is a local approximation, and "moving" processes perform a global approximation through the whole design domain.

Moving Least-Squares Method with Sensitivity Data

If the sensitivity (gradient) of each sampling point can be calculated efficiently,[14] then the sensitivity and function (response) data can be used to construct RS models. For sensitivity information $y^d_{i,xj}$, the derivative of y with respect to x_j at the ith sampling point, Eq. (1) leads us to the following relation:

$$\boldsymbol{y}^d_{xj} = \boldsymbol{T}_{xj}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{gxj} \tag{7}$$

where

$$\boldsymbol{y}_{xj}^{d} = \begin{bmatrix} y_{1x_j}^{d} \\ y_{2x_j}^{d} \\ \vdots \\ y_{nx_j}^{d} \end{bmatrix}, \qquad \boldsymbol{T}_{xj} = \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots \\ 0 & 0 & \cdots & 1 & \cdots \\ \vdots & & & & \\ 0 & 0 & \cdots & 1 & \cdots \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \qquad \boldsymbol{\varepsilon}_{gxj} = \begin{bmatrix} \varepsilon_{g1_{xj}} \\ \varepsilon_{g2_{xj}} \\ \vdots \\ \varepsilon_{gn_{xj}} \end{bmatrix}$$

$\boldsymbol{y}_{xj}^{d}$ is a sensitivity vector with respect to x_j. $\boldsymbol{T}_{xj}$ estimates the sensitivity vector and can be constructed from the matrix $\boldsymbol{X}$ in Eq. (1). Each component of $\boldsymbol{T}_{xj}$ is the derivative of each component of $\boldsymbol{X}$ with respect to x_j. Therefore, $\boldsymbol{T}_{xj}$ is dependent on the design of experiment and the basis model of RSM. For example, if the basis model of RSM is linear, $\boldsymbol{T}_{xj}$ is composed with 1 at the $(j+1)$th column and 0 elsewhere.

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

$$L_g(\boldsymbol{x}) = \boldsymbol{\varepsilon}_{gx_1}^{T}\boldsymbol{W}_g(\boldsymbol{x})\boldsymbol{\varepsilon}_{gx_1} + \boldsymbol{\varepsilon}_{gx_2}^{T}\boldsymbol{W}_g(\boldsymbol{x})\boldsymbol{\varepsilon}_{gx_2} + \cdots + \boldsymbol{\varepsilon}_{gx_{\mathrm{NDV}}}^{T}\boldsymbol{W}_g(\boldsymbol{x})\boldsymbol{\varepsilon}_{gx_{\mathrm{NDV}}} \tag{8}$$

where $\boldsymbol{W}_g(\boldsymbol{x})$ can be constructed from the similar manner with Eq. (4), but a different weighting function can be adopted, if necessary.

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

$$\begin{aligned} L_{\mathrm{new}}(\boldsymbol{x}) &= (1 - sw_g)L_y(\boldsymbol{x}) + (sw_g)L_g(\boldsymbol{x}) \\ &= (1 - sw_g)\boldsymbol{\varepsilon}_y^{T}\boldsymbol{W}(\boldsymbol{x})\boldsymbol{\varepsilon}_y + (sw_g)\sum_{j=1}^{\mathrm{NDV}}\boldsymbol{\varepsilon}_{gxj}^{T}\boldsymbol{W}_g(\boldsymbol{x})\boldsymbol{\varepsilon}_{gxj} \end{aligned} \tag{9}$$

where sw_g is a scale factor (or weighting factor) for gradient errors.

To minimize the new least-squares function,

$$\left.\frac{\partial L_{\mathrm{new}}}{\partial \boldsymbol{\beta}}\right|_b = (1 - sw_g)\left.\frac{\partial L_y}{\partial \boldsymbol{\beta}}\right|_b + (sw_g)\left.\frac{\partial L_g}{\partial \boldsymbol{\beta}}\right|_b = 0 \tag{10}$$

Substitution and rearrangement gives

$$\begin{aligned} &\left[(1 - sw_g)\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{x})\boldsymbol{X} + sw_g\sum_{j=1}^{\mathrm{NDV}}\boldsymbol{T}_{xj}^{T}\boldsymbol{W}_g(\boldsymbol{x})\boldsymbol{T}_{xj}\right]\boldsymbol{b} \\ &\quad = (1 - sw_g)\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{x})\boldsymbol{y} + sw_g\sum_{j=1}^{\mathrm{NDV}}\boldsymbol{T}_{xj}^{T}\boldsymbol{W}_g(\boldsymbol{x})\boldsymbol{y}_{xj}^{d} \end{aligned} \tag{11}$$

In addition, if $\boldsymbol{X}$ matrix is composed with the normalized values of x_j, for example, $-1 \sim 1$, the gradient relation expression in Eq. (7) should be modified. By the chain rule,

$$\frac{\partial y}{\partial x_j} = \frac{\partial y}{\partial xn_j} \times \frac{\partial xn_j}{\partial x_j} \tag{12}$$

where $\partial xn_j/\partial x_j = 2/(xu_j - xl_j)$, where xu_j and xl_j are the upper and lower limit of x_j.

Equation (7) becomes

$$\boldsymbol{y}_{xj}^{d} = \boldsymbol{T}_{xj}\boldsymbol{\beta} \times \frac{\partial xn_i}{\partial x_j} + \boldsymbol{\varepsilon}_{gxj} \tag{13}$$

and Eq. (11) becomes

$$\begin{aligned} &\left[(1 - sw_g)\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{x})\boldsymbol{X} + sw_g\sum_{j=1}^{\mathrm{NDV}}\left(\frac{\partial xn_j}{\partial x_j}\right)^2\boldsymbol{T}_{xj}^{T}\boldsymbol{W}_g(\boldsymbol{x})\boldsymbol{T}_{xj}\right]\boldsymbol{b} \\ &\quad = (1 - sw_g)\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{x})\boldsymbol{y} + sw_g\sum_{j=1}^{\mathrm{NDV}}\frac{\partial xn_j}{\partial x_j}\boldsymbol{T}_{xj}^{T}\boldsymbol{W}_g(\boldsymbol{x})\boldsymbol{y}_{xj}^{d} \end{aligned} \tag{14}$$

After introducing the new notation, the preceding equation becomes

$$\boldsymbol{A}(\boldsymbol{x})\boldsymbol{b} = \boldsymbol{c}(\boldsymbol{x}) \tag{15}$$

Finally, the coefficients of the response surface model can be obtained in the form

$$\boldsymbol{b}(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x})^{-1}\boldsymbol{c}(\boldsymbol{x}) \tag{16}$$

Through the preceding sequences, a RS model, which considers the gradient and function data, can be obtained. Authors denote this RSM as a sensitivity-based response surface model (SRSM). Note that the coefficients from the preceding sequences are dependent on the approximation location $\boldsymbol{x}$.

The sensitivities at all of the sampling points are necessary as well as function data to apply the proposed method. If the sensitivities are calculated efficiently using a method such as the adjoint variable method (AVM),[14] the proposed method is computationally efficient. To verify the effectiveness of this proposed method, some examples will be demonstrated.

Parametric Study for Moving Least-Squares Method with Sensitivity

Necessity of Parametric Study and Correlation Coefficient

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

As a measure of the accuracy for RSM, a correlation coefficient is adopted. For two random variables $X1$, $X2$, a correlation coefficient is calculated as shown by Eq. (17):

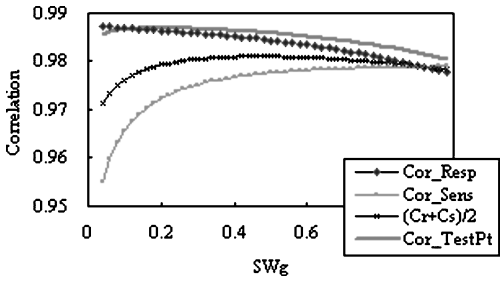
$$R_{X1X2} = E\{(X1 - m_1)(X2 - m_2)\}, \qquad r_{X1X2} = R_{X1,X2}/\sigma_1\sigma_2 \tag{17}$$

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, the measure of the accuracy is the correlation coefficients of sampled data and the estimated data from the RSM. Additionally, two kinds of data exist: the response and the sensitivity data. For the response data, the correlation coefficient can be calculated directly because the sampled and estimated response values are already in the two-vector form. However, for the sensitivity data, serialization can make the sensitivity data a vector, and then the correlation coefficient of the sampled and estimated sensitivity values can be obtained. For example, a system has three design variables and 20 sampling points, and then a correlation coefficient for sensitivities can be obtained from two vectors that have 60 elements (size of each vector is 60). Table 1 shows the serialization of the sensitivity data to calculate the correlation coefficient.

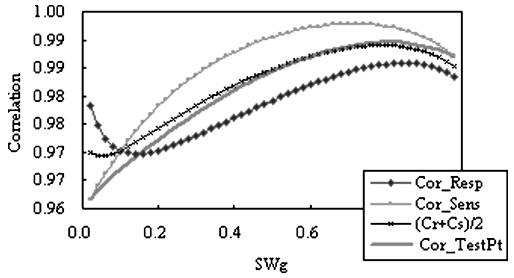
There are several criteria for ensuring accuracy. An important reason for adopting this correlation concept is that there is a normalized (or equivalent) comparison between the response error and gradient error. Because response and sensitivity values have different dimensions, a direct comparison of those errors is impossible. However, by using the correlation concept, normalized accuracy makes it possible to compare both errors.

Table 1 Example of the serialization of sensitivity data

Size of vector	No. of sampling points	Sampled sensitivity from a real system	Predicted sensitivity from RSM
1	1	$\partial y/\partial x_1$	$\partial \hat{y}/\partial x_1$
2	——	$\partial y/\partial x_2$	$\partial \hat{y}/\partial x_2$
3	——	$\partial y/\partial x_3$	$\partial \hat{y}/\partial x_3$
4	2	$\partial y/\partial x_1$	$\partial \hat{y}/\partial x_1$
⋮	⋮	⋮	⋮
58	20	$\partial y/\partial x_1$	$\partial \hat{y}/\partial x_1$
59	——	$\partial y/\partial x_2$	$\partial \hat{y}/\partial x_2$
60	——	$\partial y/\partial x_3$	$\partial \hat{y}/\partial x_3$



a) Case 1: Weighting functions for response/sensitivity = linear/exponential



b) Case 2: Weighting functions for response/sensitivity = constant/fourth polynomial

Fig. 2 Correlation with respect to sw_g.

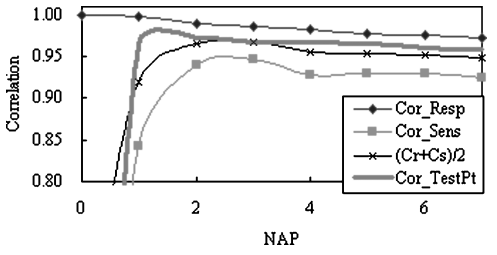
Parametric studies for several mathematical functions are performed for several different conditions, and a few representative results will be shown in the next section. The following two sections show the results of parametric studies for a Rosenbrock test function, which is defined by

$$f(\boldsymbol{x}) = 100\left(x_1^2 - x_2\right)^2 + (1 - x_1)^2 \tag{18}$$

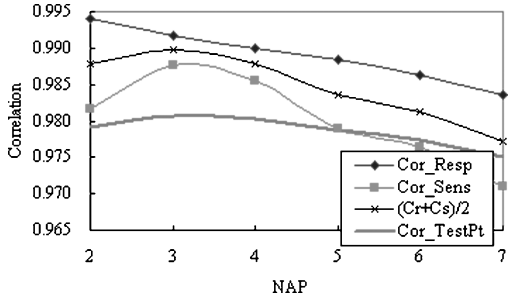
Figures 2 and 3 have four curves that represent the following: Cor_Resp, correlation coefficient of response values (15 pts by LHC) for construction of RSM; Cor_Sens, correlation coefficient of sensitivity values (15 pts by LHC) for construction of RSM; (Cr + Cs)/2, (Cor_Resp + Cor_Sens)/2; and Cor_TestPt, 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 a RS models, and Cor_TestPt is from the testing points demonstrating that there are many points to represent a global accuracy. LHC means Latin Hypercube design.

Parametric Study for sw_g

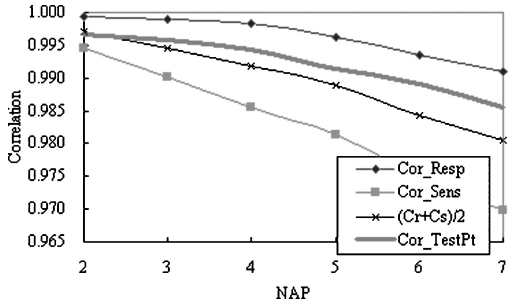
The first parameter to study is sw_g. Five weighting functions are tested, which are defined in Eq. (3). Each five weighting functions for response and, sensitivity are used, and, therefore, a total of 25 cases are tested. Because the trends of correlations are quite different for different weighting functions and other different conditions, a certain trend is not always the best. However, a few representative trends are shown among the 25 combinations of the weighting functions. Figure 2 shows two representative trends for sw_g.



a) MLSM only (without sensitivity): linear



b) MLSM with sensitivity: quadratic/fourth polynomial



c) MLSM with sensitivity: linear/fourth polynomial

Fig. 3 Correlation with respect to NAP.

Figure 2 shows variations of the correlation coefficient with respect to variations of sw_g. Different cases mean different combinations of weighting functions. For instance, case 1 used the linear weighting function for the response data and the exponential weighting function for the sensitivity data. Case 2 used the constant weighting function for the response data and the fourth-order polynomial weighting function for the sensitivity data.

Case 1 of Fig. 2 is one of the general trends for sw_g. A larger sw_g (closer to 1) causes a larger Cor_Sens and a smaller Cor_Resp because the larger sw_g tries to minimize the gradient errors more in Eq. (9). Case 2 is not a general trend but demonstrates an important behavior similar to that of case 1. The important behavior is that the trend of (Cr + Cs)/2 is closest to the trend of Cor_TestPt. This means that the maximum (Cr + Cs)/2 is the closest to the maximum global accuracy of the RSM.

During the parametric study for sw_g, sw_g should be larger than 0 and smaller than 1. Zero of sw_g means "no sensitivity data" (MLSM only) and 1 of sw_g means "no response data" (physically impossible).

Parametric Study for R_I

R_I, the size of the local approximation region, is the second parameter to be studied. A small R_I makes an RSM approximation close to an interpolation in which the RSM passes all sampled points. In this case, the RSM can be very noisy, and this noise phenomenon can decrease the filtering effect, which is one of the major advantages of the RSM.

Additionally, if R_I is not a large enough size, the matrix $\boldsymbol{A}(\boldsymbol{x})$ in Eq. (16) can be singular or ill conditioned, and an ill-conditioned matrix operation causes very bad estimations. On the other hand, a

large R_I leads the MLSM (local approximation) closer to a conventional RSM (global approximation). Because R_I can greatly affect the accuracy of the approximation, R_I is a very important parameter for the local approximation.

Four weighting functions, which are constant, linear, quadratic and fourth-order polynomial, are tested for response and sensitivity data. Therefore, a total of 16 cases are tested to check the effect of R_I when sensitivity data are used to construct RSM. Among the 16 cases, two representative cases are shown in Figs. 3b and 3c. Figure 3b uses the quadratic weighting function for the response data and the fourth polynomial weighting function for the sensitivity data. Additionally, Fig. 3c uses the linear weighting function for the response data and the fourth polynomial weighting function for the sensitivity data.

When sensitivity data are not used to construct RSM, a total of four cases are tested. Among the four cases, one representative case is shown in Fig. 3a using the linear weighting function for the response data.

To perform the local approximation using MLSM, the minimum number of data required are the number of terms of the RSM. For example, a two-variable problem requires at least six data to construct a full quadratic RSM. In Fig. 3, the horizontal axis is number of added points (NAP), which is related to R_I. The NAP means the number of added points to the minimum number of data required to perform the local approximation. Therefore, NAP is proportional to R_I, but not linearly proportional.

As Fig. 3a shows, NAP (or R_I) affects significantly for the without-sensitivity case. In this case, note that the sensitivity is used only for the accuracy testing, not for the construction of an RSM. A minimum possible NAP (or R_I) leads to a maximum Cor_Resp, but that does not generally suggest a maximum global accuracy. As Fig. 3b shows, consideration of both Cor_Resp and Cor_Sens can give a good approximation because the profiles of (Cr + Cs)/2 and Cor_TestPt show similar trends. Therefore, we have to maximize Cr + Cs to get a good RSM.

In the case of Fig. 3c, all four profiles have the same trends. This result is good for suggesting an approximation because the accuracy criteria do not need to be considered in this case. However, whether the accuracy profiles will be like Figs. 3b or 3c is unknown.

Observations from the Parametric Study

From the parametric study, several observations that represent the behavior of the accuracy of the approximation can be summarized. An important fact is that the minimum possible R_I (size of local approximation), which maximizes the correlation of the response, is not good for the global accuracy. Both correlations for response and sensitivity should be maximized together because the trend of (Cr + Cs)/2 is the closest to the trend of the global accuracy (Cor_TestPt). Even though the maximums of both correlations do not always guarantee the global maximum accuracy, this is the best criterion for the current situation.

Optimization of Parameters for RS Modeling

During the construction of the response surface model, the parameters sw_g and R_I are determined for the best RS modeling from an optimization procedure. Because a discontinuity problem can occur during this optimization procedure, a genetic algorithm[15] is adopted. This parametric optimization is defined as follows.

Find:

$$sw_g, R_I$$

Maximize:

$$\text{Corr\_Resp} + \text{Corr\_Sens}$$

Subject to:

$$\text{Approximation domain contains enough data}$$

$$(\text{Matrix } \boldsymbol{A}(\boldsymbol{x}) \text{ should not be singular}) \qquad (19)$$

Difficulties and Solutions During the Approximation

There were several difficulties during the approximation and their solutions are listed.

Estimation at Near Boundary Points

At the boundary, the rack of the data points within a local approximation region made the accuracy of approximation decline. A resizable approximation region strategy or additional data near the boundary could potentially solve this problem.

Near-Singular Problem

If the matrix $\boldsymbol{A}(\boldsymbol{x})$ in Eq. (15) is not singular but ill conditioned, the RSM estimation could be very bad. During matrix operations for calculating coefficients of RSM, the inverse matrix $\boldsymbol{A}(\boldsymbol{x})^{-1}$ in Eq. (16) became too sensitive, and this sensitive inverse operation caused a problem. In this research, a reciprocal condition number (Rcond) was adopted as a criterion to check the condition of $\boldsymbol{A}(\boldsymbol{x})$. If matrix $\boldsymbol{A}$ were well conditioned, Rcond ($\boldsymbol{A}$) was near 1. If matrix $\boldsymbol{A}$ was badly conditioned, Rcond ($\boldsymbol{A}$) was near 0. For example, the Rcond of an identity and a singular matrix are 1 and 0, respectively.

In this research, if the reciprocal condition number (Rcond) of $\boldsymbol{A}(\boldsymbol{x})$ was smaller than a certain predefined value (for example, 0.0001), the parametric optimizer considered $\boldsymbol{A}(\boldsymbol{x})$ singular and finds a larger R_I.

Selection of an Objective of Parametric Optimization

As the parametric study shows, the maximum (Corr_Resp + Corr_Sens) does not guarantee the maximum approximation accuracy. Therefore, a good selection of the objective is important. Some alternatives are as follows.

Maximize:

$$\text{Cor\_Sens s.t. Cor\_Resp} > \text{Cor\_Target (a given constant)}$$

Maximize:

$$\text{Cor\_Resp} + \text{Cor\_Sens} - (\text{Cor\_Resp} - \text{Cor\_Sens})\text{\^{}2}$$

Minimize:

$$1/(\text{Cor\_Resp} + \text{Cor\_Sens}) + |\text{Cor\_Resp} - \text{Cor\_Sens}|$$

However, more research is required to find more effective objectives that can represent the global maximum accuracy.

Numerical Examples

Function Test 1

The first mathematical example is a Rosenbrock function with two variables, which is well known for its banana-shaped domain. As the contour plot shows in Fig. 4a, this function has a long, narrow, parabolic-shaped flat valley. Sixteen points that evenly distributed through the four-level, full-factorial design were sampled for the experiments, and 100 points were selected for testing the accuracy of the RSMs. Sensitivities at each of the sampling points were obtained analytically.

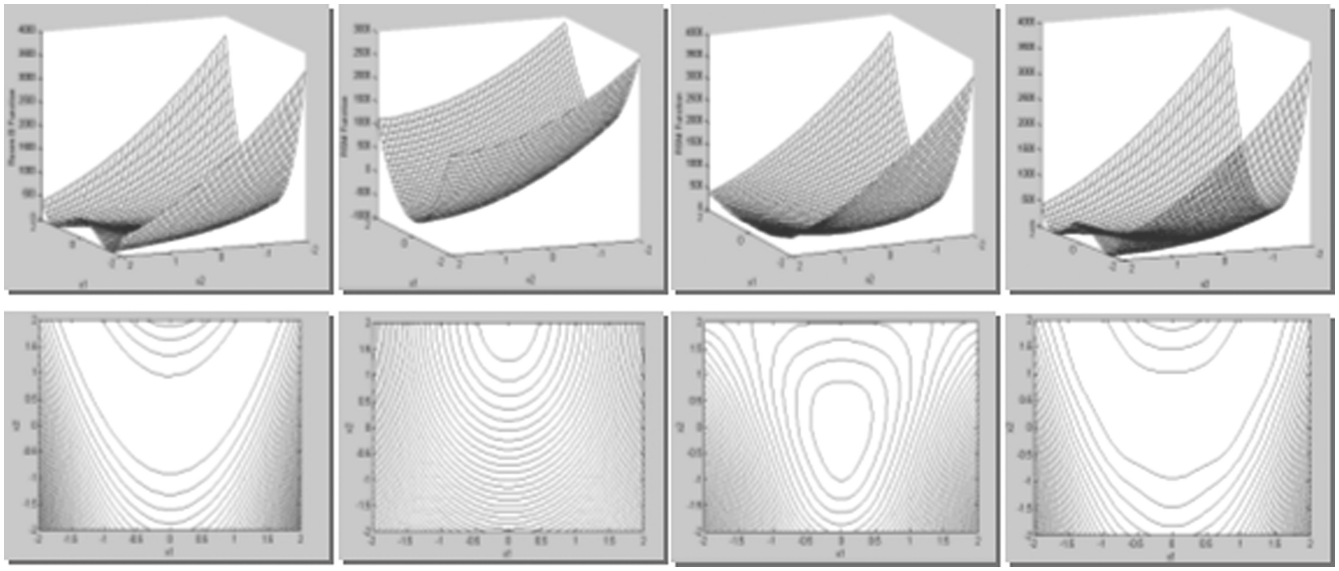
Figures 4b–4d show several of the constructed RS models using different methods such as the classical LSM, MLSM only, and MLSM with sensitivity-performing parametric optimization, respectively. In the case of Fig. 4d, a genetic algorithm was used for the parametric optimization, and the optimization problem was defined as in Eq. (19). The reciprocal condition number was also used to prevent a near-singular problem. If the reciprocal condition number (Rcond) of $\boldsymbol{A}(\boldsymbol{x})$ was smaller than a defined limit value (0.0001), the parametric optimizer considered $\boldsymbol{A}(\boldsymbol{x})$ singular and finds a larger R_I.

In the graphical point of view, Fig. 4d is very close to the original function, and its contour plot shows a V-shaped valley.

Table 2 shows the numerical accuracies of RSMs from several different methods. MLSM gives a much more accurate RSM than LSM for this example. Clearly, in the last case MLSM with

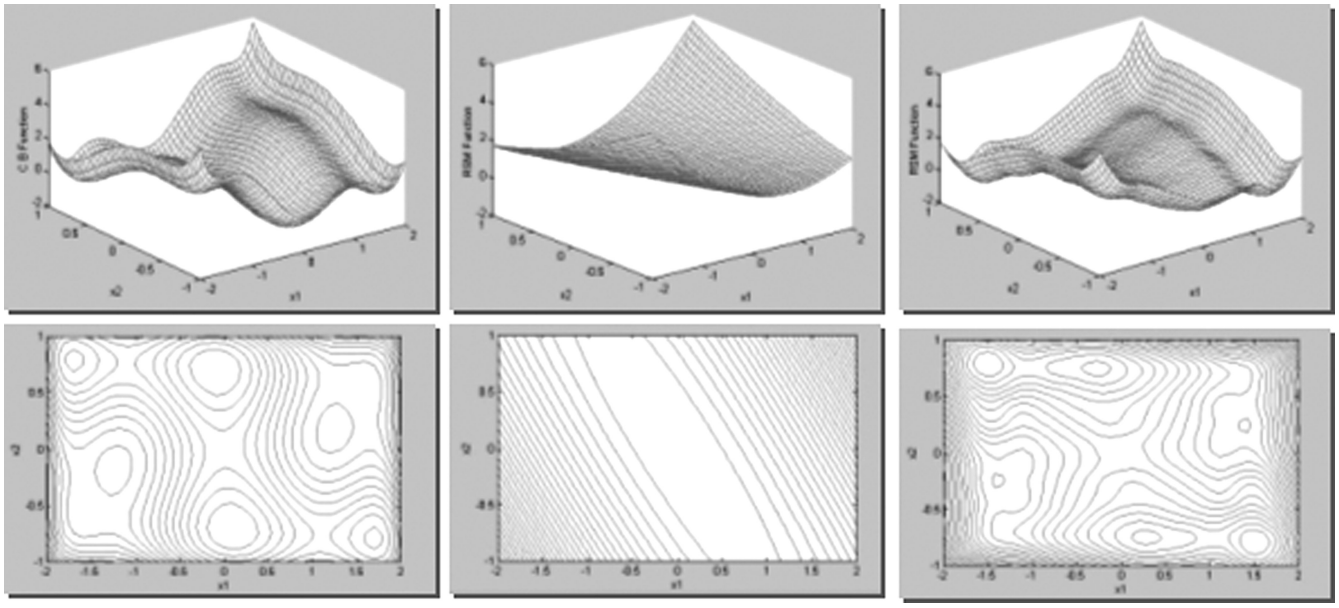
Table 2 Comparison of accuracies of RSMs by different methods (test function 1)

Case	RSM method	Computation time	Cor_Resp at 16 sampling pts	Cor_Sens at 16 sampling pts	Cor_TestPt at 100 test pts
b	LSM	16	0.892312	0.881685	0.872262
c	MLSM	16	0.999736	0.941945	0.980606
d	MSLM w. Sens	16 + 16*0.2	0.999857	0.999371	0.999849



a) Original function **b) Classical LSM** **c) MLSM** **d) MLSM with sensitivity**

Fig. 4 Function and contour plots for RSMs by different methods.



a) Original function **b) Classical LSM or MLSM** **c) MLSM with sensitivity**

Fig. 5 Function and contour plots for RSMs by different methods.

sensitivity-performing parametric optimization has the best accuracy from among the given models. Comparing case c (MLSM) and case d (MLSM with sensitivity) gives an interesting observation. Even though Cor_Resps of both RSMs are similar, a smaller Cor_Sens makes the RSM of case c less accurate globally (smaller Cor_TestPt). Obviously, case d uses more data than case c, so that the case d model should be more accurate than the case c model.

One assumption of this method was that the sensitivity used in this method could be obtained with cheap computations, for instance, by using the AVM.[14] The third column of Table 2 shows the estimated computation time by using the AVM. Conventionally, the computation time using AVM is about 10–20% of the function evaluation time. Because the computation time for RS modeling is negligible, the third column of Table 2 includes the time only for the sampling of function and sensitivity data. Therefore, when the sensitivity is calculated efficiently, this method is very effective. Finally, the parametric optimization leads to more enhancement of the RSM.

Function Test 2

The second function test is a two-dimensional six-hump camelback function, which is defined as

$$f(\boldsymbol{x}) = \left(4 - 2.1x_1^2 + x_1^4/3\right)x_1^2 + x_1x_2 + \left(-4 + 4x_2^2\right)x_2^2 \quad (20)$$

where

$$-2 \le x_1 \le 2, \qquad -1 \le x_2 \le 1$$

This function has four local optima and two global optima within the bounded region, as shown in Fig. 5a. The true global optimum points are located at (−0.0898, 0.7126), (0.0898, −0.7126), where $f = -1.0316$. In this example, the reconstruction of the original characteristics is very important, because the original system has very complex contours.

Sixteen evenly distributed points through the four-level, full-factorial design are sampled for the experiments, and 100 points are selected for testing the accuracy of the RSM. Figures 5b and 5c show the results of the construction of the RSM according to the different methods, and graphically case c can successfully describe the six optimum positions. Because only 16 points, which is too few to describe this complex system, are selected, the optimum positions are a little shifted and distorted. However, only the proposed RSM could represent all optimums. Because both cases of LSM and MLSM (without sensitivity) gave the same results, only one function and one contour plot (Fig. 5b) are represented.

Table 3 explains the errors of RSMs, and a special phenomenon occurs. Because the 16 sampling points satisfy a quadratic RS model accurately, the Cor_Resp at those 16 sampling points is 1.0 for the classical LSM and the MLSM (cases b and c). Those RS models seem well fitted, but the RS models contain large global errors caused by the large gradient errors. Case d, MLSM with sensitivity-performing parametric optimization, gives the most accurate approximation.

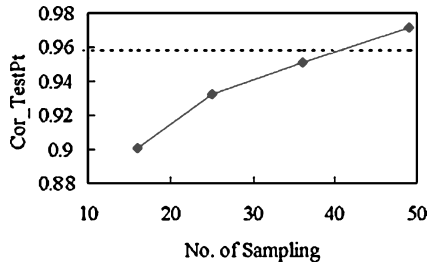


Fig. 6 Global accuracy of RS models according to the number of sampling (MLSM only).

When the MLSM is only used without sensitivity, variation of the global accuracy (Cor_TestPt) according to the number of sampling points is represented. This result verifies the efficiency of the proposed method. Figure 6 shows the accuracy of RSM using only the MSLM without sensitivity. From Fig. 6, MLSM needs over 40 point samplings to achieve the equivalent accuracy, Cor_TestPt = 0.958894, which is from the SRSM. Because the SRSM (case d in Table 3) used function and gradient data at 16 points, the total sampling (computation) time for the SRSM is 19.2(= 16 + 16*0.2). This result is for the situation when the sensitivity computation time using AVM is at most 20% of the function evaluation time. Therefore, the SRSM is much more computationally efficient than the conventional MLSM.

For example, when an RSM using MLSM is constructed from 36 points, Cor_TestPt of the RSM is close to that of SRSM. However, this RSM (MLSM with 36 computations) failed to describe all six of the local optimums because it has a large sensitivity error. Figure 7 shows the function and contour plot of the RSM, and only two optima are described. However, the SRSM used only 19.2 computations, but it successfully described the six optima as shown in Fig. 5c. Therefore, the consideration of sensitivity is good for the purpose of the computational efficiency as well as the accuracy.

Conclusions

The response surface method (RSM) has become one of the well-known approximation techniques for complicated systems. However, its significant approximation error is a major drawback of utilizing this approach. This paper discussed how to construct an RSM efficiently and accurately using sensitivity when accurate sensitivities were available by utilizing cheap computations. Also, the formulations for incorporating the sensitivity using the moving least-squares method (MLSM) are derived. During the approximation using the MLSM with sensitivity information, several parameters should be determined carefully. From the parametric study of the weighting factor for the gradient and the size of the local approximation region, the behavior of the accuracy depending on the parameters were studied. The optimization of the parameters was performed in order to increase the accuracy of the RSM, and a genetic algorithm was adopted because of the discontinuity problem of the optimization.

Further additional important improvements in this research were developed. The correlation coefficient was adopted for the normalized comparison between the function and gradient errors. Also, the reciprocal condition number was applied to avoid ill-conditioned

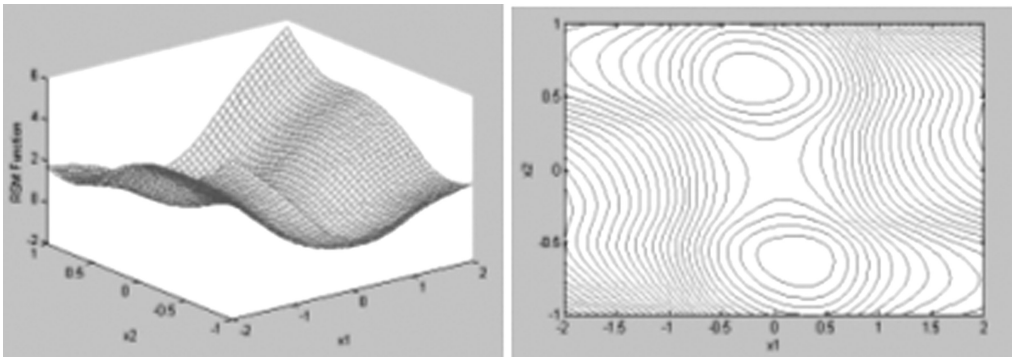


Fig. 7 RSM by using MLSM from 36 points.

Table 3 Comparison of accuracies of RSMs by different methods (test function 2)

Case	RSM methods	Computation time	Cor_Resp at 16 sampling pts	Cor_Sens at 16 sampling pts	Cor_TestPt at 100 test pts
b	LSM	16	1.00000	0.819928	0.900337
c	MLSM	16	1.00000	0.819928	0.900337
d	MSLM w. sens	16 + 16*0.2	0.999987	0.943883	0.958894

approximations. Several difficulties and their solutions, during the application of the proposed method, were described, and numerical examples were demonstrated. From these examples, the proposed methods gave not only accurate but also computationally efficient RS models.

Acknowledgment

This research was supported by the Center of Innovative Design Optimization Technology, Korea Science and Engineering Foundation.

References

[1]Myers, R. H., and Montgomery, D. C., *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley, New York, 1995.

[2]Choi, K. K., Youn, B. D., and Yang, R. J., "Moving Least Square Method for Reliability-Based Design Optimization," 4th World Congress of Structural and Multidisciplinary Optimization (WCSMO4), June 2001.

[3]Qu, X., and Haftka, R. T., "Probabilistic Safety Factor Based Response Surface Approach for Reliability-Based Design Optimization," 3rd ISSMO/AIAA Internet Conf. on Approximations in Optimization, Oct. 2002.

[4]Sellar, R. S., and Batill, S. M., "Concurrent Subspace Optimization Using Gradient-Enhanced Neural Network Approximations," AIAA Paper 96-4019, Sept. 1996.

[5]Liu, W., and Batill, S. M., "Gradient-Enhanced Neural Network Response Surface Approximations," AIAA Paper 2000-4923, Sept. 2000.

[6]Lauridsen, S., Vitali, R., Van Keulen, F., Haftka, R. T., and Madsen, J. I., "Response Surface Approximation Using Gradient Information," 4th World Congress of Structural and Multidisciplinary Optimization (WCSMO4), June 2001.

[7]Kim, C., Wang, S., and Choi, K. K., "Efficient Response Surface Modeling Using Sensitivity," 4th World Congress of Structural and Multidisciplinary Optimization (WCSMO4), June 2001.

[8]van Keulen, F., and Vervenne, K., "Gradient-Enhanced Response Surface Building," AIAA Paper 2002-5455, Sept. 2002.

[9]Liu, W., and Batill, S. M., "Gradient-Enhanced Response Surface Approximations Using Kriging Models," AIAA Paper 2002-5456, Sept. 2002.

[10]Kim, C., Wang, S., and Choi, K. K., "Parametric Study for Response Surface Modeling Using MLSM and Sensitivity," *5th World Congress of Structual and Multidisciplinary Optimization (WCSMO5)*, 2003, pp. 137, 138.

[11]Lancaster, P., and Salkauskas, K., "Surface Generated by Moving Least Squares Methods,"*Mathematics of Computation*, Vol. 37, No. 155, 1981, pp. 141–158.

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

[13]Atkeson, C. G., Moore, A. W., and Schaal, S., "Locally Weighted Learning," *AI Review*, Vol. 11, No. 1–5, 1997, (1–5), pp. 11–73.

[14]Haug, E. J., Choi, K. K., and Komkov, V., *Design Sensitivity Analysis of Structural Systems*, Academic Press, Orlando, FL, 1986.

[15]Gen, M., and Cheng, R., *Genetic Algorithms and Engineering Design*, Wiley, New York, 1997.

A. Messac
Associate Editor