Design and analysis of computer experiments

Design and analysis of computer experiments: The method of Sacks et al.

L.F.P. Etman

WFW reportnumber: 94.098

Design and analysis of computer experiments: The method of Sacks et al.

L.F.P. Etman
Stevin Centre, Eindhoven University of Technology

August 9, 1994

Abstract

In structural optimization, usually an approximation concept is introduced as interface between (FEM) structural analysis code and optimization algorithm. In some cases of optimum design, a global approximation concept can be effectively applied. Then, approximation model functions are built of all objective and constraint functions, whose values, for a certain design point, follow from the structural analysis calculations. In this way, the original optimization problem is completely replaced by an explicitly known approximate optimization problem.

Response-surface techniques are commonly applied to build global approximation models, especially when dealing with responses of physical experiments. For one response function, this means that a user defined model function is fitted to the response data calculated at the design sites of some experimental design. Errors between model function and experimental response values are assumed to be randomly distributed. However, in the structural optimization case, an experiment is a computer analysis with a deterministic response as a result. This rules out the statistical assumptions response-surface model building is based upon. Furthermore, it is often difficult to find model functions that approximate the true response behaviour within the desired accuracy.

Sacks and coworkers proposed a new model building strategy, that is especially suited for deterministic computer responses. Their basic assumption is that computer responses can be modelled as a realization of a stochastic process. This finally leads to a response prediction that exactly describes all calculated computer responses. Additionally, the mean squared error of a prediction can be calculated, which serves as a measure of accuracy of the prediction.

A review is given of the method of Sacks et al.. It has been implemented in MATLAB. Some analytical response functions and small design optimization problems have been used to test the model building capacities and the effectiveness of the method when applied as global approximation concept in structural optimization.

It is concluded that the strategy as proposed by Sacks and coworkers is not suited for implementation in a design optimization tool, mainly because of two reasons. Firstly, maximum likelihood parameter estimation is computationally expensive and not straightforward, while the quality of the parameter estimations is questionable. This computational effort is disadvantageous if many constraint functions have to approximated, which often occurs in structural optimization. Secondly, the mean squared error can not be unconditionally trusted. To have an independent accuracy measure several additional computer experiments (test points) have to be performed.

Main advantage of the approximations of Sacks et al., compared with response-surface models, is the flexibility to automatically adapt to the calculated response data. The same sort of flexibility also occurs for the moving least-squares variant of the least-squares parameter estimation. It is recommended to investigate whether this method can add flexibility to, and solve the model function selection problems of, the response-surface strategy without an unmanageable amount of computations.

Contents

Chapter 1

Introduction

In structural optimum design, optimization tools usually do not directly solve the optimization problem, but use a suitable approximation concept to interface structural analysis software and mathematical programming algorithm. The general optimization problem can be formulated as to find the set of n design variables $\mathbf{x}$ that minimizes objective function:

$$F_{obj}(\mathbf{x}) \tag{1.1}$$

subject to constraints:

$$g_j(\mathbf{x}) \leq c_j \qquad j = 1, \ldots, m \tag{1.2}$$

in the design space:

$$x_i^l \leq x_i \leq x_i^u \qquad i = 1, \ldots, n \tag{1.3}$$

Objective function and constraints are often not explicitly known: most objective function and constraint values of a certain design $\mathbf{x}$ follow from time-consuming structural analysis calculations. The introduction of an approximation concept avoids programming problems and is computationally more convenient (Haftka and Gürdal, 1992). The basic principle is to generate explicitly known approximations of objective function and constraints, which build an approximate optimization problem that can be easily solved by a mathematical programming algorithm.

Often, local approximations are used. A local approximation of objective function or constraint is based on function value and derivatives with respect to the design variables in a single point of the design space. Since such an approximation is only valid in the vicinity of this point, the search subregion of the approximate optimization problem has to be limited by additional constraints, so called movelimits. The optimum of the approximate problem serves as starting point of a new cycle of approximation and optimization. This sequential approximate optimization process is continued until an acceptable optimum is reached. Local approximation concepts are popular in structural optimization, because they are easy to use and they can handle design problems with a large number of design variables.

For some optimum design applications, it is useful to build global approximation models of objective function and constraints, that create an explicitly known approximate optimization problem in the complete design space or a large part of it. So, the region in which the approximations are valid is significantly larger than when using a local approximation concept. The price you pay is that the number of design variables is limited: the required amount of numerical experiments to build the approximation models grows exponentially for

increasing number of design variables. Additionally, the number of experiments also limit the model accuracy that can be obtained. Despite these rather severe limitations, global approximations can be used to get a rough idea of the influence of some important design variables on the response, and to search for promising starting points of a sequential approximate optimization process. Besides these preliminary investigations, global model building may also be useful after an optimum design has been found. In a region around this optimum design, approximation models of objective function and important constraints can be constructed. These models can give insight in, for example, the sensitivities of design variables on objective function and constraints. They may also prove to be useful if the design has to be combined with other designs, and small design variable or constraint boundary changes are still possible. An updated optimum design can be easily found, without the necessity to repeat the optimization.

To generate global approximations, often response-surface techniques are used (Schoofs, 1987), that were originally developed for the model fitting of physical experiments. Response-surface model building starts with postulating the approximate model functions. Then, an experimental design is constructed that contains the design points for which computer experiments are carried out. Finally, regression analysis is used to estimate the unknown parameters of the approximation models by fitting the numerical response data. Schoofs, Klink and Van Campen (1992), for example, applied the response-surface strategy to optimize a child's car seat. Schoofs (1987) and Roozen-Kroon (1992) designed a major third bell using regression models.

However, computer experiments are completely different from physical experiments. Repeating a computer experiment gives exactly the same response, this in contrary to the physical equivalent. This deterministic behaviour of computer responses is the main reason for Sacks, Welch, Mitchell and Wynn (1989b) to reject the response-surface method. An approximation model should be able to exactly predict the calculated responses, and therefore, they argue, one can not postulate model functions beforehand. The model should be flexible enough to adapt to the functional behaviour of the response. To accomplish this, Sacks and coworkers developed a new statistical method with interpolative features.

This report deals with the proposed statistical method for the design and analysis of computer experiments. It will be referred to as the method of Sacks et al.. The basic principles are discussed, and some simple analytical functions are used to test the method. The effectiveness in structural optimization is studied, and compared with the response-surface strategy. This finally leads to the conclusion that the present configuration of the method of Sacks et al. is not suitable for implementation in a design optimization tool.

Chapter 2

Theorem

2.1 Modelling, estimation and prediction

Starting point of the method of Sacks et al. is the modelling of the deterministic computer response $y(\mathbf{x})$ as a realization of a stochastic process Y:

$$Y(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}) \tag{2.1}$$

with:

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &= [f_1(\mathbf{x}), \ldots, f_k(\mathbf{x})]^T \\ \boldsymbol{\beta} &= [\beta_1, \ldots, \beta_k]^T \end{aligned}$$

The model is a sum of a linear regression model $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ and a random process $Z(\mathbf{x})$. This random process Z is assumed to have zero mean, and a covariance between $Z(\mathbf{x})$ and $Z(\mathbf{w})$ at design sites $\mathbf{x}$ and $\mathbf{w}$, that is the product of a process variance σ^2 and a correlation function $R(\mathbf{w}, \mathbf{x})$:

$$Cov(\mathbf{w}, \mathbf{x}) = \sigma^2 R(\mathbf{w}, \mathbf{x}) \tag{2.2}$$

The stochastic process Y is assumed to be Gaussian.

For a certain experimental design $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ computer experiments have been performed, and response data $\mathbf{y}_s = [y(\mathbf{s}_1), \ldots, y(\mathbf{s}_N)]$ is available. From these computer responses the unknown parameters $\boldsymbol{\beta}$ and σ^2 can be estimated:

$$\hat{\boldsymbol{\beta}} = \left(F^T R^{-1} F\right)^{-1} F^T R^{-1} \mathbf{y}_s \tag{2.3}$$

$$\hat{\sigma}^2 = \frac{1}{N}\left(\mathbf{y}_s - F\hat{\boldsymbol{\beta}}\right)^T R^{-1} \left(\mathbf{y}_s - F\hat{\boldsymbol{\beta}}\right) \tag{2.4}$$

with the regression design matrix F and the correlation matrix R defined by:

$$\begin{aligned} F &= [\mathbf{f}(\mathbf{s}_1), \ldots, \mathbf{f}(\mathbf{s}_N)]^T \\ R &= [R(\mathbf{s}_i, \mathbf{s}_j)]_{i,j} \qquad 1 \le i, j \le N \end{aligned}$$

However, before calculating $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$, first the unknown parameters of the correlation function have to be estimated. Using maximum likelihood, they result from the minimization of (Welch, Buck, Sacks, Wynn, Mitchell and Morris (1992)):

$$1/2(N \ln \hat{\sigma}^2 + \ln \det R) \tag{2.5}$$

which is a function of only the correlation parameters and the response data.

For these parameter estimations, the best linear unbiased prediction of the response is:

$$\hat{y} = \mathbf{f}^T(\mathbf{x})\hat{\boldsymbol{\beta}} + \mathbf{r}^T(\mathbf{x})\hat{\boldsymbol{\alpha}} \tag{2.6}$$

with the column $\hat{\boldsymbol{\alpha}}$ defined by:

$$\hat{\boldsymbol{\alpha}} = R^{-1}(\mathbf{y}_s - F\hat{\boldsymbol{\beta}}) \tag{2.7}$$

and with $\mathbf{r}$ being a column of correlations between the Z's at the design sites of S and the untried input $\mathbf{x}$:

$$\mathbf{r} = [R(\mathbf{s}_1, \mathbf{x}), \ldots, R(\mathbf{s}_N, \mathbf{x})]^T \tag{2.8}$$

The second part $\mathbf{r}^T(\mathbf{x})\hat{\boldsymbol{\alpha}}$ of formula (2.6) is in fact an interpolation of the residuals of the regression model $\mathbf{f}^T(\mathbf{x})\hat{\boldsymbol{\beta}}$. Therefore, all response data will be exactly predicted. In this report attention is restricted to a constant regression model β, because most authors did not find much advantage in using more complex regression models (e.g. see Welch et al. (1992) and Bernardo, Buck, Liu, Nazaret, Sacks and Welch (1992)).

Sacks et al. (1989b) also gave a representation of the mean squared error:

$$MSE(\hat{\mathbf{y}}(\mathbf{x})) = \sigma^2 \left\{ 1 - \begin{bmatrix} \mathbf{f}^T(\mathbf{x}) & \mathbf{r}^T(\mathbf{x}) \end{bmatrix} \begin{bmatrix} 0 & F^T \\ F & R \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f}^T(\mathbf{x}) \\ \mathbf{r}^T(\mathbf{x}) \end{bmatrix} \right\} \tag{2.9}$$

It is remarked that the MSE is zero for $\mathbf{x}$ equal to a design site in S.

2.2 Correlation functions

In all literature concerning the method of Sacks et al., always a correlation function of the type $R(\mathbf{w}, \mathbf{x}) = R(\mathbf{w} - \mathbf{x})$ is selected, and within this family attention is restricted to the product correlation type:

$$R(\mathbf{w}, \mathbf{x}) = \prod_{j=1}^{n} R_j(d_j) \tag{2.10}$$

with:

$$d_j = w_j - x_j$$

Bernardo et al. (1992) and Welch et al. (1992) used:

$$R_j(d_j) = e^{-\theta_j |d_j|^{p_j}} \tag{2.11}$$

In Sacks, Schiller and Welch (1989a) $p_j = 2\ \forall j$ was inserted to get realizations that are infinitely differentiable. They argued that this kind of correlation is especially suitable for computer experiments depending smoothly on the input variables. Parameter p_j should have smaller values than 2, for applications with more erratic responses. Instead of correlation function (2.11) in this report:

$$R_j(d_j) = e^{-|\frac{d_j}{\theta_j}|^{p_j}} \tag{2.12}$$

is used to get a better conditioned minimization problem (2.5).

Currin, Mitchell, Morris and Ylvisaker (1991) and Sacks et al. (1989b) mentioned other correlation functions, like a linear function:

$$R_j(d_j) = \begin{cases} 1 - \frac{|d_j|}{\theta_j} & \text{if } |d_j| < \theta_j \\ 0 & \text{if } |d_j| \geq \theta_j \end{cases} \tag{2.13}$$

and a cubic correlation function:

$$R_j(d_j) = \begin{cases} 1 - 6\left(\frac{d_j}{\theta_j}\right)^2 + 6\left(\frac{|d_j|}{\theta_j}\right)^3 & \text{if } |d_j| < \frac{\theta_j}{2} \\ 2\left(1 - \frac{|d_j|}{\theta_j}\right)^3 & \text{if } \frac{\theta_j}{2} \leq |d_j| < \theta_j \\ 0 & \text{if } |d_j| \geq \theta_j \end{cases} \tag{2.14}$$

These functions are non-negative. Currin et al. (1991) also gave a linear and cubic correlation function that can become negative.

Generally, hardly any attention is paid to criteria to select a proper correlation function. Currin et al. (1991) stated that simplicity is an important guiding principle. In any case, the number of parameters of the correlation function to be estimated should be reduced as much as possible, because of the computational expense of the maximization of the log-likelihood. For every evaluation of equation (2.5) β , σ^2 and $\det(R)$ have to be calculated, which becomes computationally expensive for growing number of design sites. Additionally, for increasing number of correlation parameters to be estimated, the required number of evaluations to minimize (2.5) will also increase. Usually, for higher functional dimension, the number of design points will grow exponentially, which is combined with an increase of the number of correlation parameters if every design variable direction is given it's own parameter values.

So, without caution a high increase of computational costs will occur for growing number of design variables. This is the main reason why Welch et al. (1992) tried to screen all correlation parameters that need a value of their own, and to find out which parameters can be equally valued. They also tried to combine the parameter screening with a design variable reduction. If θ_j of correlation function (2.11) becomes zero during the parameter estimation, the j-th design variable can be removed. I do not agree with this kind of design variable deletion, because (analytical) functions can be found for which θ_j can become almost zero, in which case the j-th design variable *does* influence the response. This is further illustrated in example 3.2.

2.3 Experimental design

Mainly optimal design strategies are used by Sacks et al. (1989a), Sacks et al. (1989b) and Currin et al. (1991) to build an experimental design. Sacks et al. (1989a) computed the experimental design S from the minimization of the mean squared error in the design space χ :

$$\int_{\chi} MSE(\hat{y}(\mathbf{x}))dx \tag{2.15}$$

Another criterion is to minimize the maximum mean squared error:

$$\max_{x \in \chi} MSE(\hat{y}(\mathbf{x})) \tag{2.16}$$

This criterion is expected to behave rather discontinuous, and to have many local minima, because of the max-value operator. A third optimal design strategy was applied by Currin et

al. (1991). They used an entropy criterion that finally boils down to the maximization of the determinant of the matrix R:

$$\det R \tag{2.17}$$

All optimal design strategies suffer from two serious drawbacks. Firstly, they are computationally expensive, like the minimization of equation (2.5) to estimate the parameters. But now the computational burden is even larger, because of the high dimensional optimization problem. Therefore, often a set of candidate points is selected from which design S has to be selected. The optimal design strategy is comparable with optimal design of experiments applied in response-surface analysis, where $\det(F^T F)$ is maximized instead of $\det R$.

A second drawback is that the unknown parameters of the correlation function have to be set on some fixed values to be able to calculate the optimal design criterion. Therefore, one has to search for a robust set of parameters before starting the optimum design calculations. The way to find these robust values and the effect on the final prediction is not clear.

To avoid these problems, one can use classical designs instead, like a full or fractional factorial design. Bernardo et al. (1992) and Welch et al. (1992) found Latin hypercube sampling especially suitable for computer experiments. Latin hypercubes can be generated by giving each design variable equally spaced values, but in different random order. As a result, a relative uniform covering of the experimental region will be found. A special feature of a Latin hypercube design is that every design variable level is used only once. So, if a design variable hardly influences the response and is deleted from the set of design variables, the experimental design is still a Latin hypercube design with no design sites coinciding. This in contrary to for example a full factorial design, that will lead to many redundant analysis calculations.

During the model building it may appear that the current experimental design is not sufficiently large to properly predict the response. Then, a larger experimental design is desired, which can be established by adding design points to the present experimental design. When dealing with classical designs, adding design points is not straightforward and often rather ad hoc. In contrary, optimal design techniques seem to be especially suited to add design sites to an existing experimental design.

Addition of design points to an existing experimental design is not mentioned in the literature concerning the method of Sacks et al. Welch et al. (1992) for example applied a Latin hypercube design of 30 and 40 design points, and reported that these designs proved to be not sufficient to model the response. For their problem they found 50 runs to be successful. But they waisted the preceding model building steps of 30 and 40 design points, and did not use the response values of the corresponding design points anymore. So, instead of 50, 120 calculations were made before a satisfactory model was found, while the model is based only on the last 50 calculations. I think that it is important for a successful and efficient model building to have the opportunity to add design points to an existing experimental design.

2.4 Accuracy and design variable effects

To have a measure of the accuracy of the prediction ($BLUP$), some authors calculate the mean squared error MSE of the response by means of equation (2.9). For a circuit-simulation example, Sacks et al. (1989b) computed the standard residuals at 100 random points $\mathbf{r}_i$;

$$\frac{y(\mathbf{r}_i) - \hat{y}(\mathbf{r}_i)}{\sqrt{MSE(\hat{y}(\mathbf{r}_i))}} \tag{2.18}$$

to see whether the MSE of the $BLUP$ is a meaningful indicator of uncertainty in the prediction. They showed that these standardized residuals were approximately standard normal and therefore concluded that the MSE did provide a valid estimate of the error for their example. However, Currin et al. (1991) concluded that their Bayesian predictor performed well, but that the most disturbing note was the failure of the 95 % probability intervals for the Bayesian predictors to cover the true values consistently well, except for the intervals produced by the linear and nonnegative linear correlations. This means that the calculated MSE values were for most correlation functions too small.

MSE predictions can be checked by comparing the overall root mean squared error ($RMSE$) with the empirical root mean squared error ($ERMSE$) calculated for a certain amount N_p of (random) design sites $\mathbf{r}_i$. The $RMSE$ and $EMRSE$ are respectively defined by:

$$RMSE = \sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p} MSE(\hat{y}(\mathbf{r}_i))} \tag{2.19}$$

and:

$$ERMSE = \sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p} (\hat{y}(\mathbf{r}_i) - y(\mathbf{r}_i))^2} \tag{2.20}$$

For N_p being sufficiently large the $RMSE$ value should be near the $EMRSE$ value. It will be clear that the $ERMSE$ calculations can only be done for analytical test examples, because of the computational expense of every experiment. To avoid additional design calculations Welch and Sacks (1991) used:

$$ERMSE_{-1} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{y}_{-1}(\mathbf{s}_i) - y(\mathbf{s}_i))^2} \tag{2.21}$$

with $\mathbf{s}_i$ the i-th experiment of the experimental design S, and $\hat{y}_{-1}(\mathbf{s}_i)$ the predictor of $y(\mathbf{s}_i)$ based on all the data except the observation $y(\mathbf{s}_i)$. In this way no extra design calculations are necessary. The correlation parameters have to be estimated based on all N observations, and have to be set at the same value for $\hat{y}_{-1}(\mathbf{x}_i)$ for all i, otherwise N times a new parameter estimation has to be performed. Still, N times an inverse of matrix R has to be computed. It is important that the prediction is based on sufficient data points, such that deletion of one design site has not large effects on the prediction.

Finally, insight in the effect of design variables on the response is valuable. Sometimes, design variables can be detected that hardly effect the response and can be removed. More response data will become available for the other design variables. Therefore, Sacks et al. (1989b) decomposed the response into an average, main effects for every design variable, two-variable interactions and higher order interactions. The average of the predicted response is defined by:

$$\mu_0 = \frac{1}{\prod_{h=1}^{n}(x_h^u - x_h^l)} \int \hat{y}(\mathbf{x}) \prod_{h=1}^{n} dx_h \tag{2.22}$$

the main term effects by:

$$\mu_p(x_p) = \frac{1}{\prod_{h\neq p}(x_h^u - x_h^l)} \int \hat{y}(\mathbf{x}) \prod_{h\neq p} dx_h - \mu_0 \tag{2.23}$$

and the interaction effects:

$$\mu_{pq}(x_p, x_q) = \frac{1}{\prod_{h \neq p,q}(x_h^u - x_h^l)} \int \hat{y}(\mathbf{x}) \prod_{h \neq p,q} dx_h - \mu_p(\mathbf{x}_p) - \mu_q(\mathbf{x}_q) - \mu_0 \tag{2.24}$$

In the case that the design variables are normalized between zero and one, the product terms before the integrals vanish and equations (2.22), (2.23) and (2.24) become equal to the expressions mentioned in, for example, Sacks et al. (1989b). If the prediction is a reasonable resemblance of the true response equations (2.22), (2.23) and (2.24) can give an indication of the different parameter effects. The product correlation type (2.10) is very helpful for an efficient calculation of the main term and interaction term effects: only one dimensional integrations have to be performed.

2.5 Sacks et al. versus response-surface in structural optimization

Building a global approximation model of an objective function or constraint can be divided into six distinct steps for both the response-surface strategy and the method of Sacks et al.:

1. Select model function.
2. Build a new experimental design or extend the existing design.
3. Perform (numerical) experiments.
4. Determine the unknown model parameters.
5. Estimate the accuracy of the approximation.
6. Restart at step 1 or 2 if the approximation is not satisfactory.

Step 1 is rather crucial for response-surface modelling. The accuracy that can be achieved mainly depends on the quality of the proposed approximation function. Whenever the model function is not satisfactory, new model terms have to be added (think for example of forward and backward regression techniques) or a completely different model function has to be chosen. This model selection is the most difficult and restricting part of response-surface model building.

The method proposed by Sacks et al. does not suffer from these model function selection problems. All experimental response data is exactly fitted. In fact, the model function relies on the response values and therefore is far more flexible. In step 1 'only' a proper correlation function has to be selected. This additional flexibility is beneficial in structural optimization, because the functional behaviour of objective function and constraints is usually not known. Based on physical insight some general design variable effects may be determined and incorporated into the model approximations by means of intermediate design variables and intermediate responses (see Barthelemy and Haftka (1993)). However, large parts of the true functional behaviour remain unknown.

Automatic model adaptation becomes almost necessary when many constraints have to be approximated. This often occurs. In that case response-surface modelling will only be manageable if the number of constraints to be approximated is small, or several constraints

can approximated by the same type of model function. This does not mean that the estimated parameter values need to be the same. Though, the number of model functions to be selected and adapted by the user should be as small as possible. In contrary, the models of Sacks et al. automatically adapt to the calculated response data and therefore all constraints can be dealt with separately.

However, parameter estimations in the case of Sacks et al. are far more expensive and less straightforward than for the response-surface variant. So for Sacks et al., the number of constraints is limited by the parameter estimation. For larger number of design variables and response values, the computational burden grows exponentially. The only way to increase the number of constraints is to have an other more simple way of parameter estimation. One of my conclusions will be that the statistical basis is questionable, and therefore one can argue against the maximization of the log-likelihood function to estimate the parameters. An alternative may be found in the moving least-squares method of section 2.6.

To complete the comparison between response-surface and Sacks et al., the different ways of calculating the accuracy of the approximation should be mentioned. In the response-surface case, the accuracy follows from the residuals between model prediction and response values at the design sites of the experimental design. Possibly, additional (random) experiments may be performed to have data which the models have not been fitted on. Accuracy in the case of Sacks et al. is determined by the mean squared error MSE of equation (2.9). However, the examples will show that the probability bounds based on this MSE should not be interpreted in an absolute statistical sense. They usually can only give some indication of the order of accuracy, so it is recommended to compute responses at some extra design sites to have an additonal independent measure of accuracy.

Finally, it is remarked that some special cases desire to have a specific behaviour of the approximations of the objective function and constraints, for example to apply a specific optimization algorithm, or to achieve a certain type of approximations. Then, the response-surface methodology is (of course) preferable, because you can select whatever model function you want. This in contrary to the method of Sacks et al. that heavily relies on the response data and therefore generates approximations which have a less explicitly known behaviour.

2.6 Moving least-squares approximations

Examination of equation (2.6) shows that the interpolation of the residuals is a sum of correlation functions stored in the column $\mathbf{r}(\mathbf{x})$. The elements of the parameter column $\hat{\boldsymbol{\alpha}}$ determine the individual heights of the correlation functions (see figure 2.1). Equation (2.7) implies that the parameter values of column $\hat{\boldsymbol{\alpha}}$ are calculated such that the residuals $(y_S - F\hat{\beta})$ are exactly predicted by equation (2.6). The correlation parameters θ_j in functions (2.12), (2.13) and (2.14) determine the width of the correlation functions. Parameters p_j of correlation function (2.12) influence the continuity of the function.

The correlation functions used by Sacks et al. are a kind of weight functions determining the influence of a response value of a certain design site on the response value of a design point at some distance. A same sort of usage of weight functions can be found in the moving least squares (MLS) methods (see e.g. Lancaster and Salkauskas (1981), or Belytschko, Lu and Gu (1994) who applied MLS in an element free Galerkin method). The moving least squares approximation of the response $y(\mathbf{x})$ is defined by:

$$\hat{y}(\mathbf{x}) = \mathbf{p}^T(\mathbf{x})\mathbf{a}(\mathbf{x}) \tag{2.25}$$

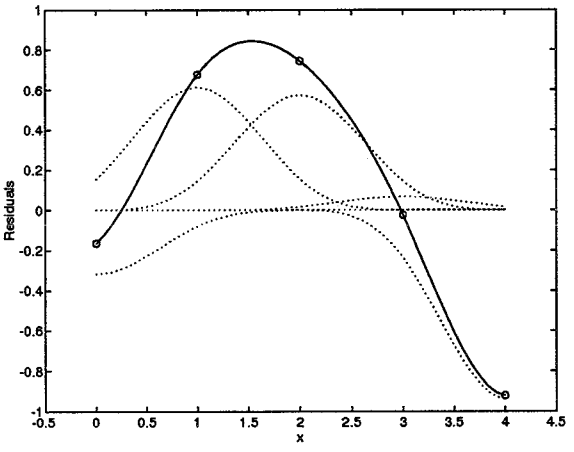

Figure 2.1: Prediction is a sum of correlation functions.

with $\mathbf{p}(\mathbf{x})$ a column of basis functions. For example, a linear basis in three dimensions gives:

$$\mathbf{p}^T(\mathbf{x}) = \begin{bmatrix} 1 & x_1 & x_2 & x_3 \end{bmatrix}$$

or a quadratic basis in two dimensions:

$$\mathbf{p}^T(\mathbf{x}) = \begin{bmatrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \end{bmatrix}$$

For a certain design $\mathbf{x}$, the coefficients $\mathbf{a}(\mathbf{x})$ follow from the minimization of the weighted norm:

$$J = \sum_{j=1}^{no} w(\mathbf{x} - \mathbf{x}_j) \left[\mathbf{p}^T(\mathbf{x}_j)\mathbf{a}(\mathbf{x}) - u_j\right]^2 \tag{2.26}$$

where no is the number of design points in the neighbourhood of $\mathbf{x}$ for which the weight function $w(\mathbf{x} - \mathbf{x}_j)$ is not zero, and $\mathbf{x}_j$ $j = 1, \ldots, no$ are the corresponding design sites of the experimental design $S : \mathbf{x}_j \in S$. Scalar u_j is equal to the response value at $\mathbf{x}_j$. Then:

$$\mathbf{a}(\mathbf{x}) = A^{-1}(\mathbf{x})B(\mathbf{x})\mathbf{u} \tag{2.27}$$

with matrices $A(\mathbf{x})$ and $B(\mathbf{x})$, and column $\mathbf{u}$ defined by:

$$A(\mathbf{x}) = \sum_{j=1}^{no} w(\mathbf{x} - \mathbf{x}_j)\mathbf{p}(\mathbf{x}_j)\mathbf{p}^T(\mathbf{x}_j) \tag{2.28}$$

$$B(\mathbf{x}) = \begin{bmatrix} w(\mathbf{x} - \mathbf{x}_1)\mathbf{p}(\mathbf{x}_1) & w(\mathbf{x} - \mathbf{x}_2)\mathbf{p}(\mathbf{x}_2) & \ldots & w(\mathbf{x} - \mathbf{x}_{no})\mathbf{p}(\mathbf{x}_{no}) \end{bmatrix} \tag{2.29}$$

$$\mathbf{u}^T = \begin{bmatrix} u_1 & u_2 & \ldots & u_{no} \end{bmatrix} \tag{2.30}$$

Weight function $w(\mathbf{x} - \mathbf{x}_j)$ determines which design sites $\mathbf{x}_j$ of the experimental design influence the response at design site $\mathbf{x}$ and how much influence every site actually has. If weight function $w(\mathbf{x} - \mathbf{x}_j)$ is chosen constant over the complete design space, the standard

least squares variant is obtained. Belytschko et al. (1994) relate the region of influence to the 'density' of the experimental design near $\mathbf{x}$. The region can be small if many points $\mathbf{x}_j$ lie near $\mathbf{x}$. If a rather uniformly distributed experimental design is applied (which will often be the case for a multi-dimensional problem), weight functions can be selected that have a constant region of influence everywhere in the design space. This corresponds with the correlation functions of Sacks et al., for which the region of influence is determined by correlation parameters that have the same value for any design site $\mathbf{x}$. Lu, Belytschko and Gu (1994) recommended to use orthogonal basis functions to have a numerically efficient implementation.

To examine the equivalence of moving least squares and Sacks et al., all response values need to be exactly predicted. In the global design space, there are N design points $\mathbf{s}_i$ with corresponding response values y_i $(i = 1, \ldots, N)$. Therefore, instead of taking $u_j = y(\mathbf{s}_j)$, it is searched for all columns vectors $\mathbf{u}_i \in \mathbf{v}$ $(i = 1, \ldots, N)$ for which:

$$\hat{y}(\mathbf{s}_i) = y_i$$

which means that for every design point $\mathbf{s}_i$ follows:

$$\hat{y}(\mathbf{s}_i) = \mathbf{p}^T(\mathbf{s}_i) A_i^{-1}(\mathbf{s}_i) B_i(\mathbf{s}_i) \mathbf{u}_i = y_i \tag{2.31}$$

Assembling the N equations, a linear set of equations results, relating the responses $\mathbf{y}_s$ to the column vector $\mathbf{v}$:

$$\mathbf{y}_s = Q\mathbf{v} \tag{2.32}$$

from which $\mathbf{v}$ can be solved. Then, for any design site $\mathbf{x}$, the response $\hat{y}(\mathbf{x})$ can be predicted by using equations (2.25) and (2.27).

Now, suppose $\mathbf{p}(\mathbf{x}_j) = 1$, such that the approximation is only determined by the weight functions, like the approximation of Sacks only depending on the correlation functions. Then matrix $A(\mathbf{x})$ and $B(\mathbf{x})$ can be written as:

$$A(\mathbf{x}) = \sum_{j=1}^{no} w(\mathbf{x} - \mathbf{x}_j) \tag{2.33}$$

$$B(\mathbf{x}) = \begin{bmatrix} w(\mathbf{x} - \mathbf{x}_1) & w(\mathbf{x} - \mathbf{x}_2) & \ldots & w(\mathbf{x} - \mathbf{x}_{no}) \end{bmatrix} \tag{2.34}$$

Take weight function $w(\mathbf{x} - \mathbf{x}_j)$ equal to correlation function $R(\mathbf{x} - \mathbf{x}_j)$, and consider all N design points when calculating matrices $A(\mathbf{x})$ and $B(\mathbf{x})$, i.e. insert $no = N$. This means that matrix $B(\mathbf{x})$ is now equal to row vector $\mathbf{r}^T(\mathbf{x})$ of equation (2.8). Then, for equation (2.31) follows:

$$\hat{y}(\mathbf{s}_i) = \frac{1}{\sum_{j=1}^{N} w(\mathbf{s}_i - \mathbf{s}_j)} \mathbf{r}^T(\mathbf{s}_i)\mathbf{v} = y_i \tag{2.35}$$

leading to a total set of equations given by:

$$R\mathbf{v} = \tilde{\mathbf{y}}_s \tag{2.36}$$

with a modified response vector:

$$\tilde{\mathbf{y}}_s = \begin{bmatrix} y_1 \sum_{j=1}^{N} w(\mathbf{s}_1 - \mathbf{s}_j) \\ y_2 \sum_{j=1}^{N} w(\mathbf{s}_2 - \mathbf{s}_j) \\ \vdots \\ y_N \sum_{j=1}^{N} w(\mathbf{s}_N - \mathbf{s}_j) \end{bmatrix} \tag{2.37}$$

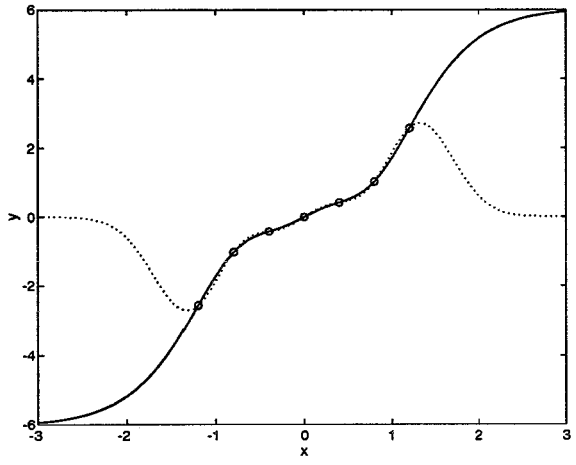

Figure 2.2: Moving least squares approximation (solid) and approximation according to Sacks et al. (dotted) using the exponential correlation function with $\theta = 0.6$ and $p = 2$.

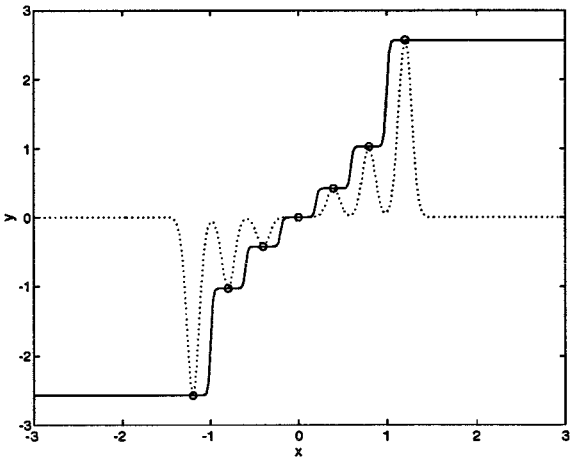

Figure 2.3: Moving least squares approximation (solid) and approximation according to Sacks et al. (dotted) using the exponential correlation function with $\theta = 0.1$ and $p = 2$.

The prediction can be written as:

$$\hat{y}(\mathbf{x}) = \frac{1}{\sum_{j=1}^{N} w(\mathbf{x} - \mathbf{s}_j)} \mathbf{r}^T(\mathbf{x}) R^{-1} \tilde{\mathbf{y}}_s \tag{2.38}$$

which has to be compared with:

$$\hat{y}(\mathbf{x}) = \mathbf{r}^T(\mathbf{x}) R^{-1} y_s = y_i \tag{2.39}$$

of Sacks et al., in the case that no regression model is present.

Both approximations (2.38) and (2.39) are summations of N weight functions. However, equation (2.38) has an additional product term that depends on $\mathbf{x}$ and a modified response vector $\tilde{\mathbf{y}}_s$, and therefore behaves quite different. This is illustrated by means of a one dimensional analytical response function $y = \tan(x)$ and an experimental design $S = \{-1.2, -0.8, \ldots, 1.2\}$. The weight function is chosen equal to correlation function (2.12) with $p = 2$. Figure 2.2 and 2.3 visualize the behaviour of the moving least squares approximation (2.38) and the approximation (2.39) of Sacks et al., for different values of θ. Approximation (2.39) is clearly a summation of correlation functions, apparent from the behaviour for the large θ value. Approximation (2.38) is much more an interpolation, with the width of the weight function determining the region in which the interpolation takes place. This different behaviour has also effect on the extrapolative character of the approximations.

Chapter 3

Examples

3.1 Response prediction and accuracy

Welch et al. (1992) considered the analytical test function:

$$y(x_1, x_2) = [30 + x_1 \sin(x_1)](4 + e^{-x_2}) \qquad 0 \leq x_1, x_2 \leq 5 \tag{3.1}$$

In figure 3.1 the contours of this function are plotted together with the location of 11 observations according to a Latin hypercube design.

Numerical optimization of the log-likelihood function yields, for correlation functions (2.12), (2.13) and (2.14), the correlation parameter estimations of table 3.1. In the third column of this table, the parameters are mentioned that are estimated, and which parameters have the same values. Using exponential correlation function (2.12) in case I and II, two optima in the minimization function (2.5) have been found. Welch et al. (1992) reported the local optimum estimation of case I (denoted by [I]), together with an estimated constant $\hat{\beta} = 129$ and a process covariance $\hat{\sigma}^2 = 341$. The present calculations have confirmed these values, and also the empirical root mean squared error ($ERMSE$) value of 3.18 on the $21 * 21$ grid $\mathbf{x} = (0.25i, 0.25j)$ for $0 \leq i, j \leq 20$. Welch et al. (1992) apparently did not find the optimum parameter values with the lower value of equation (2.5).

By means of table 3.2 the influence of the type of correlation function on the accuracy of the approximations can be compared. The three columns of the first block represent accuracy predictions calculated from the mean squared error (MSE) (equation (2.9)) on an equally spaced grid of $21 * 21$ test points $\mathbf{r}_j$ $(j = 1, \ldots, 421)$ on the design space. Column 'prob. bound.' refers to the percentage of function values of the test points between the predicted 95 % probability bounds. '$\max RMSE$' and '$RMSE$' are respectively the maximum root mean squared error: $\max_j \sqrt{MSE(\mathbf{r}_j)}$ and the overall root mean squared error defined in equation (2.19). The accuracy measures of the second block are based on the calculated function values of the $21 * 21$ grid, and can be used to check the accuracy predictions of the first block. '$ERMSE$' is the empirical root mean squared error according to equation (2.20), and '$\max error$' represents the maximum deviation between predicted and true values on the grid. Finally, in the third block (last column) an alternative empirical root mean squared error is given, defined by (2.21) that does not depend on the function values of the test points.

One can observe that the percentages of design sites between the 95 % probability bounds are smaller than 95 % for all considered cases, except for case V. This means that the MSE values are too small, which is confirmed by the $RMSE$ consequently being smaller than the

case	correlation	parameters	parameter estimations
I	(2.12)	$\theta_1, \theta_2, p_1, p_2$	$\hat{\theta}_1 = 12.3, \hat{\theta}_2 = 12.5, \hat{p}_1 = 1.51, \hat{p}_2 = 1.81$
[I]	(2.12)	$\theta_1, \theta_2, p_1, p_2$	$\hat{\theta}_1 = 2.81, \hat{\theta}_2 = 3.87, \hat{p}_1 = 2, \hat{p}_2 = 2$
II	(2.12)	$\theta = \theta_1 = \theta_2, p = p_1 = p_2$	$\hat{\theta} = 11.6, \hat{p} = 1.61$
[II]	(2.12)	$\theta = \theta_1 = \theta_2, p = p_1 = p_2$	$\hat{\theta} = 3.14, \hat{p} = 2$
III	(2.12)	$\theta = \theta_1 = \theta_2, p_1 = p_2 = 2$	$\hat{\theta} = 3.11$
IV	(2.13)	θ_1, θ_2	$\hat{\theta}_1 = 12.4, \hat{\theta}_2 = 22.7$
V	(2.13)	$\theta = \theta_1 = \theta_2$	$\hat{\theta} = 15.6$
VI	(2.14)	θ_1, θ_2	$\hat{\theta}_1 = 8.41, \hat{\theta}_2 = 12.3$
VII	(2.14)	$\theta = \theta_1 = \theta_2$	$\hat{\theta} = 8.72$

Table 3.1: Estimated correlation parameter values. The bracketed cases are parameter estimations corresponding to a local optimum of the log-likelihood function.

$ERMSE$. Linear correlation function (2.13) gives the best mean squared error predictions, exponential function (2.11) the worst. However, the $ERMSE$ indicates that response predictions based on the linear correlation function are worst, which is caused by the piecewise linear behaviour of the approximation.

Secondly, for this example, it appears that no profit is found in taking different correlation parameters for every design variable direction. In contrary, the $RMSE$ values are even somewhat larger. This confirms the strategy of Welch et al. (1992) to share common values of the correlation parameters. In my opinion, correlation parameters θ_j (and p_j) can share the same value in all design variable directions j independent of the response values y of the experimental design, if the design variables are scaled to have the same ranges, and if the density of the points of the experimental is largely the same in every design variable direction.

Finally, the $ERMSE_{-1}$ does not give a good estimate of the $ERMSE$. So, the alternative way of equation 2.21 to calculate the $ERMSE$ can not be used to avoid additional design calculations. For this example, $ERMSE_{-1}$ seems to behave as an upperbound, but this is not generally true.

For the single parameter cases III, V and VII of table 3.1, the influence of the correlation parameter value on the MSE prediction is investigated. The parameter estimations corresponding to these cases follow from the minimization of equation (2.5). This is visualized in figure 3.5. So, according to the method of Sacks et al., the minima represent the most likely parameter estimations.

$RMSE$ and $ERMSE$ values have been calculated for an equally spaced region of θ between 0.1 and 30.1, and are plotted in figures 3.2, 3.3 and 3.4. For correlation function (2.13) and (2.14) $ERMSE$ and $RMSE$ have the same type of curve. In the neighbourhood of the estimated parameter values $ERMSE$ and $RMSE$ curve are quite near each other, so the predicted MSE appears to be quite well. This does not occur for correlation function (2.12), for which $RMSE$ and $ERMSE$ curve are different. $RMSE$ predicts the $ERMSE$ correctly only for a few θ values. The minimum of equation (2.5) yields a parameter estimation with a $RMSE$ that is two times smaller than the $ERMSE$.

Generally, there is no guarantee that the estimated maximum likelihood parameters lead to a correct estimate of the mean squared error. For this example, a $4 * 4$ full factorial design

case	prob. bound	max $RMSE$	$RMSE$	max $error$	$ERMSE$	$ERMSE_{-1}$
I	86.6 %	4.05	2.40	13.4	3.74	5.17
[I]	75.7 %	4.07	1.67	15.1	3.18	7.65
II	93.2 %	4.35	2.71	12.1	3.29	5.23
[II]	84.4 %	4.75	1.96	14.7	3.08	7.02
III	84.4 %	4.75	1.96	14.7	3.08	7.02
IV	93.9 %	6.08	4.31	16.0	4.70	7.03
V	96.6 %	6.30	4.47	15.3	4.51	7.02
VI	90.9 %	4.56	2.16	12.9	2.70	6.42
VII	93.2 %	5.01	2.40	13.4	2.72	6.44

Table 3.2: Summary of the model prediction errors belonging to the parameter estimations of the previous table. The maximum range of response values of the 11-point Latin design is 49.0.

instead of the Latin design gives rise to large discrepancies between $ERMSE$ and $RMSE$ curves, also for correlation functions (2.13) and (2.14). Maximum likelihood parameters have been calculated for a 9, 16, 25, 36 and 49 point full factorial design using the cubic correlation function (2.14). The $ERMSE$ and $RMSE$ that result for these parameter values are plotted in figure 3.6. It is clearly visible that, apart from the $5 * 5$ design, rather large differences occur. My conclusion is that the mean squared error equation (2.9) can not be used to calculate absolute accuracy measures (e.g. 95 % probability bounds) of the model predictions. Only the order of magnitude may be used to have some sort of indication.

3.2 Maximum likelihood behaviour

Consider the following five one dimensional test functions visualized in figure 3.7:

$$y = \tan(x) \tag{3.2}$$

$$y = 0.5x^3 + x \tag{3.3}$$

$$y = \sin(x) \tag{3.4}$$

$$y = -0.15x^3 + x \tag{3.5}$$

$$y = x \tag{3.6}$$

Within the input variable range $x \in \{-1, 1\}$, test functions (3.2) and (3.3), as well as functions (3.4) and (3.5) have nearly the same functional behaviour. Using the experimental design given by $S = \{-1, -2/3, -1/3, 0, 1/3, 2/3, 1\}$ values of the maximum likelihood minimization function (2.5) have been calculated for a wide range of θ values. The minimization function values for correlation functions (2.12) (with $p = 2$), (2.13) and (2.14) as a function of θ are plotted in respectively figures 3.8, 3.9 and 3.10.

Figure 3.8 of the exponential correlation function (2.12) is rather astonishing. Although the response of test functions (3.4) and (3.5) are almost exactly the same, the behaviour of the minimization function is quite different. A somewhat larger, but still small, difference

between (3.2) and (3.3) even leads to a complete different behaviour of the minimization function: for test function (3.2) an optimum parameter estimation can be found at $\theta = 1$, while test function (3.3) has a maximum likelihood parameter estimation at infinity.

The cubic correlation function (2.14) is less sensitive for small changes in the response values than the exponential function. Though, the parameter estimations of functions (3.2) and (3.3) are still somewhat different. The linear correlation function seems to be completely independent of the response values, and to be only determined by the experimental design: all parameter estimations are the same for all test functions.

In the case of the linear response behaviour of function (3.6), one can expect parameter θ to go to infinity when using correlation functions (2.12) and (2.14). Otherwise, no linear approximation can be obtained. Only correlation (3.4) is able to reach a linear approximation for a finite value of the correlation parameter.

Using correlation function (2.11), Welch et al. (1992) stated that the j-th design variable is inactive if the corresponding correlation parameter θ_j becomes zero (infinity for (2.12)). From figure 3.2 can be concluded that this strategy is not correct: correlation parameter θ can go to infinity while the design variable x is certainly not inactive. They even use a test function with small linear terms in some design variables to illustrate that the unimportant design variable effects are detected. However, these design variables are found not because they have a small contribution to the total response, but because their contribution is linear.

Problems with correlation parameters going to infinity not only occur for correlation function (2.11) or (2.12). Test function $y = x_1^2 + x_1 x_2$ on $0 \leq x_1, x_2 \leq 2$ with S a $4 * 4$ full factorial design, shows that also correlation function (2.14) gives rise to infinite values of the correlation parameters, both when estimating θ_1 and θ_2, and in the case of a joint correlation parameter $\theta = \theta_1 = \theta_2$. Only the minimization function of correlation (2.13) still has a minimum.

3.3 Two bar truss

Svanberg (1987) described a two bar truss optimization problem. The truss structure, shown in figure 3.11, is parametrized by two design variables: x_1, the cross-sectional area (cm^2) of the bars, and x_2, half of the distance (m) between the two nodes 1 and 2. Objective is to minimize the weight of the truss structure, subject to stress constraints in both bars. Svanberg (1987) formulated this problem analytically as to minimize the objective function:

$$F_{obj}(\mathbf{x}) = x_1\sqrt{1+x_2^2} \tag{3.7}$$

subject to the stress constraints:

$$g_1(\mathbf{x}) = 0.124\sqrt{1+x_2^2}\left(\frac{8}{x_1} + \frac{1}{x_1 x_2}\right) \leq 1 \tag{3.8}$$

$$g_1(\mathbf{x}) = 0.124\sqrt{1+x_2^2}\left(\frac{8}{x_1} - \frac{1}{x_1 x_2}\right) \leq 1 \tag{3.9}$$

Here, the design space is set to: $1 \leq x_1 \leq 2$, $0.1 \leq x_2 \leq 1.6$. The second constraint will never become active, and is therefore not taken into account. Figure 3.12 visualizes the two bar truss optimization problem.

method	function	$RMSE$	max $error$	$ERMSE$
Sacks	F_{obj}	0.0116	0.0309	0.00549
Sacks	g_1	0.0308	0.349	0.0348
Response	F_{obj}	0.00170	0.0301	0.00380
Response	g_1	0.0751	0.488	0.0598

Table 3.3: Prediction errors of the two-bar truss problem for the method of Sacks et al. and the response-surface strategy

Global approximation models are built for objective function and first constraint, using both the response-surface strategy and the method of Sacks et al. Objective and constraint response values are calculated for the Latin hypercube design plotted in figure 3.12.

Using the model building strategy of Sacks et al., cubic correlation function (2.14) is selected, for both objective function and constraint. By means of minimization of the maximum likelihood the unknown parameters are estimated. The predictions of objective function and constraint that follow are visualized in figure (3.13). The root mean squared error ($RMSE$) accuracy prediction is calculated and compared with the empirical root mean squared error computed on a $41 * 41$ grid of points. These values are given in table 3.3, with additionally the maximum prediction error on the grid.

Full second order polynomials with third order main terms are used to fit the response data of objective function and constraint in the case of response-surface model building. Regression analysis is applied to estimate the ten parameters. The polynomial predictions are plotted in figure 3.14. For these approximations, root mean squared error ($RMSE$), empirical root mean squared error ($ERMSE$), and maximum prediction error are calculated as well, and tabulated in table 3.3. The response-surface variant of the $RMSE$ is defined by:

$$RMSE = \sqrt{\frac{1}{N-k}\sum_{i=1}^{N}(\hat{y}(\mathbf{s}_i) - y(\mathbf{s}_i))^2} \tag{3.10}$$

with k the number of parameters.

The approximations of the objective function are very accurate. The response-surface approximation is even more accurate than the prediction of Sacks et al. Differences become more apparent for the constraints. Figures 3.13 and 3.14 clearly visualize the different behaviour of both approximations. The model of Sacks tries to adapt to the calculated response data with a(n) (erroneous) local dip in the contour lines as a result. On the contrary, the polynomial approximation has difficulties in the (right) upper part of the design space and for small values of x_1, due to the selected polynomial model. Comparing the $ERMSE$ values leads to the conclusion that the constraint approximation of Sacks et al. is more accurate than the polynomial approximation. This is confirmed by the maximum prediction errors.

3.4 Modelling the acoustical damping of the hum of a bell

Roozen-Kroon (1992) constructed regression models for the frequencies and acoustical damping of the most important eigenmodes of the bell as a function of seven radii describing the

effect	3-4	3-3	3-5	1-3	1-4	1-6	5-5	1-5	2-3	1-2
percentage	27	25	19	16	15	14	12	11	11	10

Table 3.4: Important main and interaction term effects of the acoustical damping of the hum of a bell

bell shape (see figure 3.15). She concluded that the approximation models of the damping were quite inaccurate for some of the eigenmodes.

Etman (1992) considered the damping of the first partial (hum), for which the regression model has a reasonable accuracy, and tried to improve the approximation by means of the model building strategy of Sacks et al. Correlation function (2.11) was selected with one correlation parameter $\theta = \theta_1 = \theta_2 = \ldots = \theta_7$ and constant exponents $p_1 = p_2 = \ldots = p_7 = 2$. Using the same experimental design as Roozen-Kroon (1992), no improved model approximation could be found. For 41 random design points nearly the same empirical root mean squared error was found: $1.65\ 10^{-5}$ for the regression model of Roozen-Kroon (1992)[1] and $1.74\ 10^{-5}$ for the model generated by the method of Sacks et al., with about $1.6\ 10^{-4}$ for the total range ($\max_i y(\mathbf{s}_i) - \min_i y(\mathbf{s}_i)\ i = 1, \ldots, N$) of the damping data of the experimental design.

Again, for the hum an approximation model is built by means of the method of Sacks et al., but now correlation function (2.14) is selected. Then, an $EMRSE$ value of $1.76\ 10^{-5}$ is found which confirms the value of correlation (2.11). Additionally, main effects and interaction effects are calculated to get an idea of the important design variables describing the damping of the hum. Since the accuracy of the approximation model is reasonable (about 10 % of the total range), it is expected that equations (2.22), (2.23) and (2.24) give a good indication of the important main term and interaction effects.

The main term effects are plotted in figure 3.16. Clearly, radius 3 is of main influence on the acoustical damping of the hum of the bell. The contribution of an effect $\mu(\mathbf{z})$ to the total response is defined as:

$$\frac{\max_j \mu(\mathbf{z}_j) - \min_j \mu(\mathbf{z})_j}{\max_i y(\mathbf{s}_i) - \min_i y(\mathbf{s}_i)} \tag{3.11}$$

with $Z = \{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$ being the set of m sites for which the effect has been computed. Main effect 3-3 is calculated from equation (2.23) for 10 values of radius 3, varied from lower to upper bound. The contribution of this main effect according to (3.11) appears to be about 25 %. The same can be done for the interaction effects, but now a two dimensional scan ($10 * 10$) has to be performed for all interaction effects.

In table 3.4 the contributions of the different effects to the total response are given that are larger than 10 %. It becomes apparent that besides main effect 3-3 also interaction effects 3-4, 3-5 and 3-1 have an important contribution to the response. So, one can conclude that the most important design variables determining the hum of a bell are radii 3, 4, 5 and 1, and that radii 7, 2, and 6 are less important.

[1] Roozen-Kroon (1992) reported $3.3\ 10^{-5}$, but this value is two times the value that would follow from the response data plotted in a figure of her thesis. I have used the acoustical damping data that matches the thesis figures.

3.5 Ten bar truss

Haftka and Gürdal (1992) described a ten bar truss optimization problem with stress and displacement constraints. The truss structure, shown in figure 3.17, is designed for minimum weight with the cross-sectional areas of truss members as design variables. Constraints are defined by the maximum stress of 25 ksi in tension and compression, and by the maximum vertical displacement of 2 in of the free nodes. Length l is 360 in, and forces P are 100 $Kips$. The Young's modulus and the specific mass are respectively 10^4 ksi and 0.1 $lb\,in^{-3}$.

Instead of the cross-sectional areas, the reciprocals of the cross-sectional areas are taken as design variables to have a more linear functional behaviour between stresses and design variables. A random hypercube design of 100 experiments is constructed with the design variables ranging between $1/32$ in^{-1} and 1 in^{-1}. Stresses and displacements for this experimental design are calculated by means of a finite element program.

Global approximation models of stresses and displacements are built using the method of Sacks et al. with correlation function (2.14) and $\theta = \theta_1 = \ldots = \theta_{10}$. The models are tested for 100 additional random experiments. The maximum absolute prediction errors of the stresses lie in between 5 % to 15 % of the total variation of calculated stresses on the experimental design. The maximum displacement prediction errors are 5 % to 10 % of the corresponding displacement ranges. So, the approximations seem to be of a reasonable accuracy. The $ERMSE$ predicts fairly well the $RMSE$: more than 90 % of all calculated displacements and stresses at the random design sites are within the predicted 95 % probability bounds.

However, optimization of the design problem using the constructed approximation models, does not yield a design near the optimum reported by Haftka and Gürdal (1992). The displacement models are too inaccurate. For the calculated approximate optimum design, displacement prediction errors occur up to 100 % of the constraint boundary of 2 in. Though, related to the total range of displacements of node 2 of about 22 in this is only 10 %. So, the desired accuracy should be related to the constraint boundary value instead of the range of the response. Far away from the constraint boundaries, the model does not need to be equally accurate. This knowledge can not be taken into account in the method of Sacks et al.: every response value equally contributes to the response prediction. In the case of regression analysis, response values near the constraint boundary can and should be more heavily weighted than other response values.

Additionally, main term and interaction term effects are calculated, to find out which design variables have main influence on stresses and displacements. The stress in a bar appears to be mainly determined by the cross-sectional area of the bar it self, except for bar 5, for which cross-sectional areas of bars 7, 8, 3 and 1 are equally important. Cross-sectional areas of the bars to the left side of bar 5 mainly influence the stresses of the bars at the same side, and hardly the stresses of the bars on the other side. The same is true for the cross-sectional areas of the bars right to bar five. From these observations follows that for the stresses two substructures can be distinguished. Cross-sectional areas of bars 3, 1, 7 and 8 are of main influence on all displacements, which is physically correct. Despite the moderate accuracy, the main physical effects have been reconstructed from the approximation models.

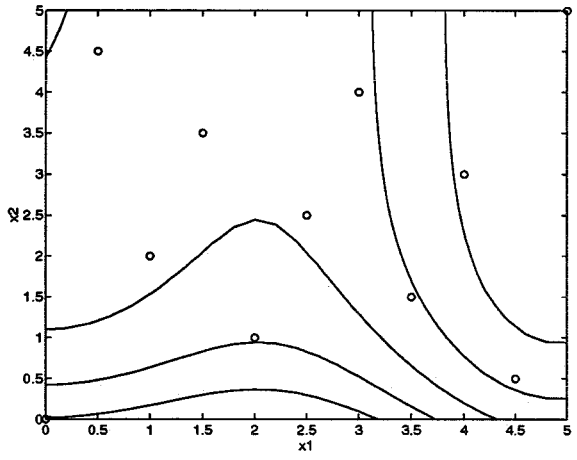

Figure 3.1: Contour plot of the analytical test function and the location of the 11-point Latin hypercube design.

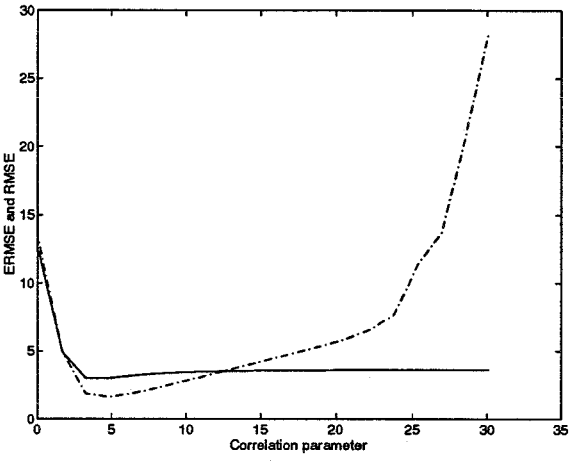

Figure 3.2: $RMSE$ (dash-dot) and $ERMSE$ (solid) using the exponential correlation function with $\theta = \theta_1 = \theta_2$ and $p_1 = p_2 = 2$ for the Latin design.

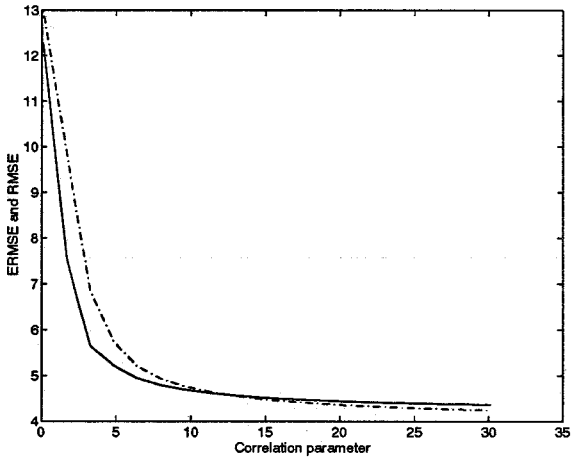

Figure 3.3: $RMSE$ (dash-dot) and $ERMSE$ (solid) using the linear correlation function with $\theta = \theta_1 = \theta_2$ for the Latin design.

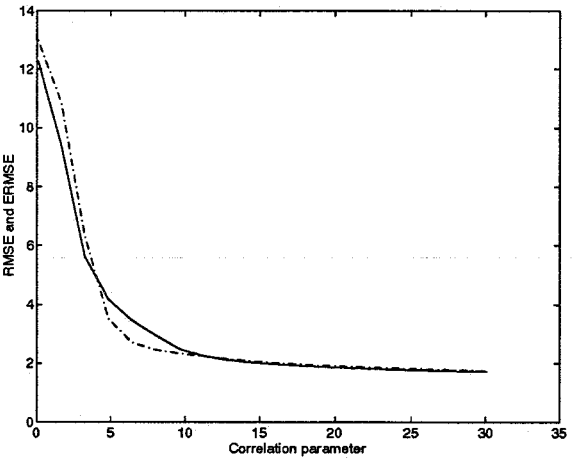

Figure 3.4: $RMSE$ (dash-dot) and $ERMSE$ (solid) using the cubic correlation function with $\theta = \theta_1 = \theta_2$ for the Latin design.

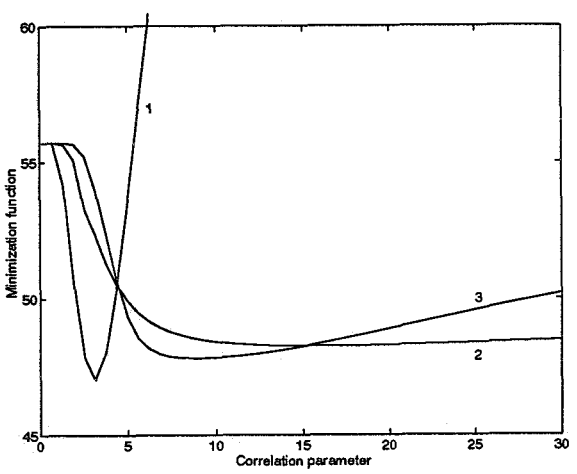

Figure 3.5: Maximum likelihood minimization function using the exponential (1), linear (2) and cubic (3) correlation function for the Latin design.

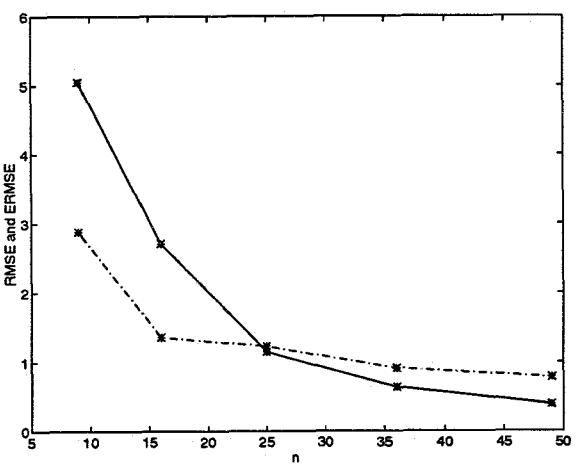

Figure 3.6: $RMSE$ (dash-dot) and $ERMSE$ (solid) using the cubic correlation function with $\theta = \theta_1 = \theta_2$ for an $n * n$ full factorial design.

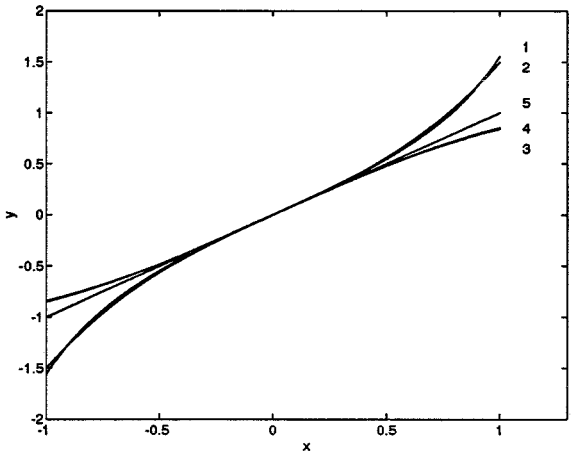

Figure 3.7: Test functions $y = tan(x)$ (1), $y = 0.5x^3 + x$ (2), $y = sin(x)$ (3), $y = -0.15x^3 + x$ (4) and $y = x$ (5).

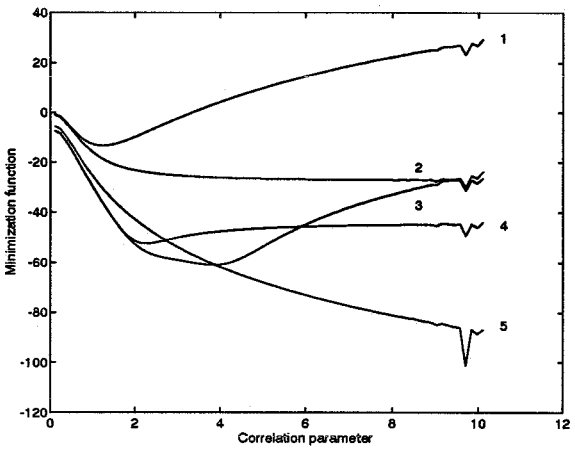

Figure 3.8: Maximum likelihood minimization function for the five test functions using the exponential correlation function.

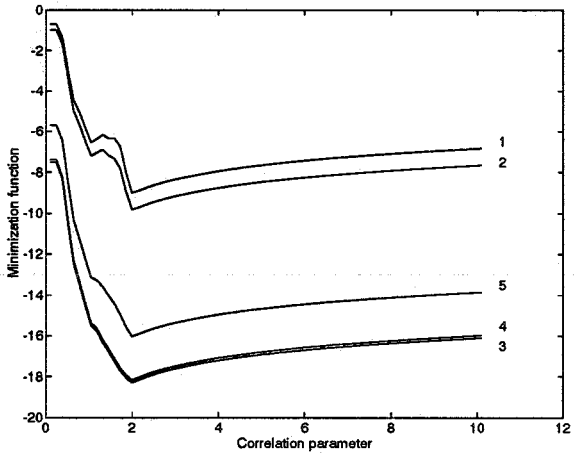

Figure 3.9: Maximum likelihood minimization function for the five test functions using the linear correlation function.

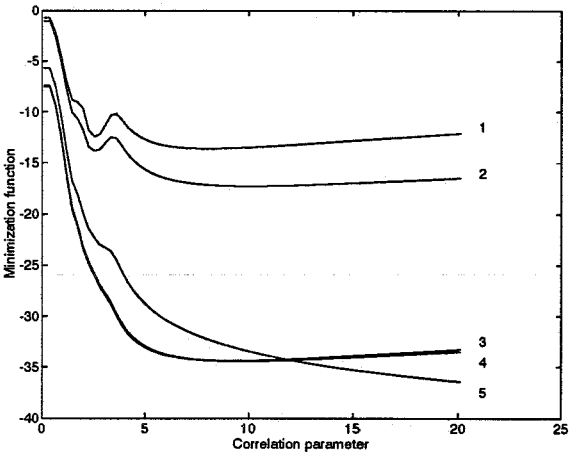

Figure 3.10: Maximum likelihood minimization function for the five test functions using the cubic correlation function.

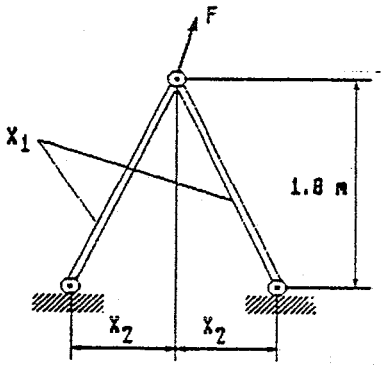

Figure 3.11: Two-bar truss structure.

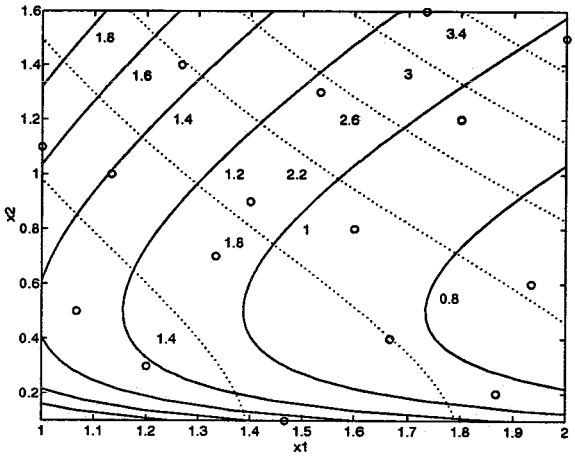

Figure 3.12: Two bar truss optimization problem and a 16-point Latin hypercube design.

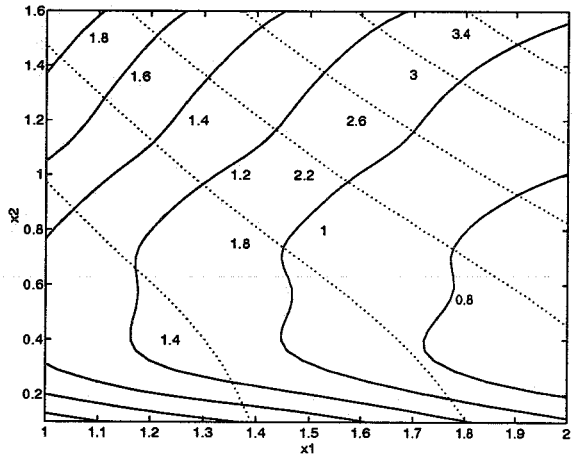

Figure 3.13: Objective function and constraint predictions of the two-bar truss by means of the method of Sacks et al.

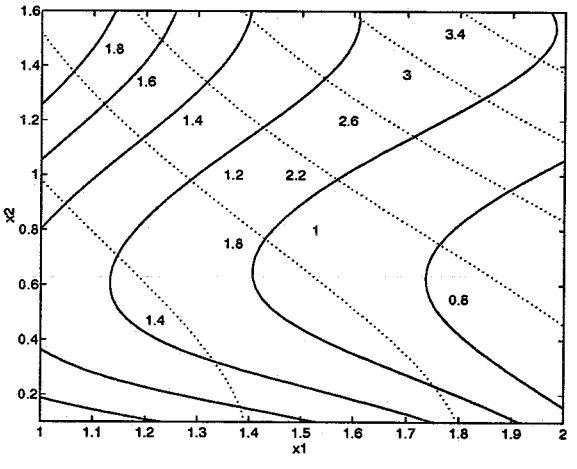

Figure 3.14: Objective function and constraint predictions of the two-bar truss by means of the response-surface strategy.

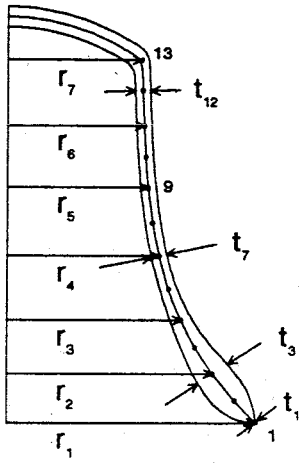

Figure 3.15: Bell geometry.

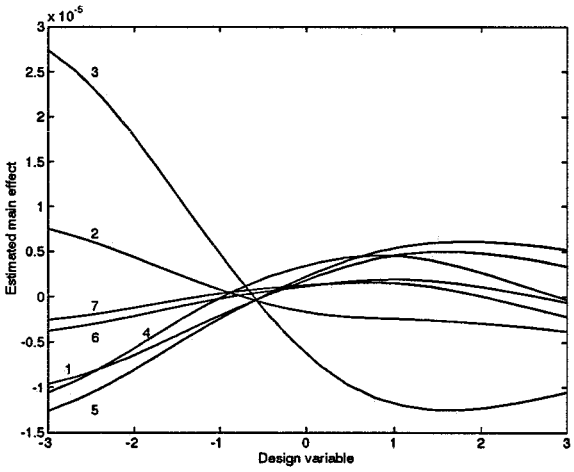

Figure 3.16: Main effects of the acoustical damping of the hum of a bell.

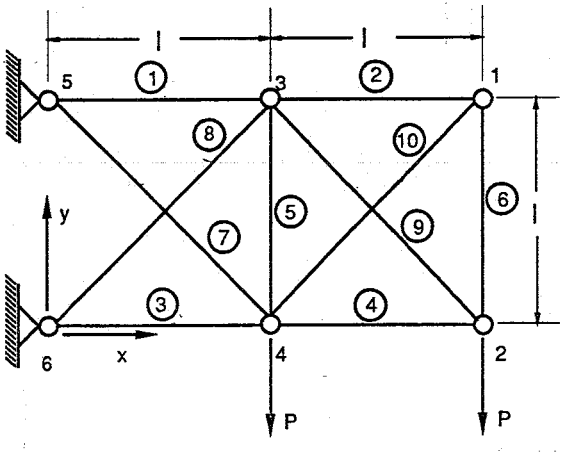

Figure 3.17: Ten bar truss

Chapter 4

Conclusion and discussion

Sacks and coworkers rejected the response-surface strategy to model responses of numerical experiments. They proposed a new statistical method especially suited for the design and analysis of computer experiments. However, I have not been convinced of the statistical value of the method. Examples have illustrated that the 95 % probability bounds often fail. The mean squared error has not proven to be an accuracy measure, that can be unconditionally trusted. Additionally, no evidence has been found that the maximum likelihood parameter estimations are indeed the most likely parameter values. For some correlation functions, the parameter estimations are sensitive for small response variable changes. I think that the basic statistical assumption of the response being a realization of a stochastic process is questionable.

Maximum likelihood parameter estimation is computationally expensive and far from straightforward. For the application in structural optimization, this is a serious drawback because often many constraint functions have to be approximated. Linear regression analysis of the response-surface strategy is computationally much cheaper and easier, especially if several constraints can be approximated by the same type of function.

Main advantage of the models of Sacks et al. is that they are much more flexible than response-surface models. All response data is automatically exactly predicted, this in contrary to response-surface models. Automatic model adaptation in the case of response-surface model building has to be done by means of forward and/or backward regression techniques that respectively add and/or remove terms of the regression model. Though, the flexibility of the models of Sacks et al. will probably never be reached.

Exact prediction of the response data of the experimental design is only permitted if all design variables are taken into account in the parameter estimation. Design variables can be removed if they have a small effect on the response. Then, these effects re-enter the model building problem as small (systematic) errors on the response data. So if, for example, all design variables are removed that are expected to have less than 10 % effect on the response, the approximation model does not need to exactly predict the response values but to stay just within 10 % deviation. For this case the strategy of Sacks should be adapted incorporating (systematic) errors in the response. Depending on the desired accuracy, the response-surface strategy with for- and backward regression may prove to be flexible enough, otherwise more flexible models are necessary.

The present configuration of the method of Sacks et al. is not suitable for implementation in a design optimization tool, mainly because of the parameter estimation and the unreliable

mean squared error. Maybe the statistical assumptions should be discarded. I think it is worthwhile to investigate whether there is an opportunity to add flexibility to regression models by means of the moving-least squares method mentioned in chapter 2. As far as I can see, a spline-like or a finite element approach in multi (> 2) dimensions seems to have too many degrees of freedom to be practically applicable in structural optimization.

Bibliography

Barthelemy, J. F. M. and Haftka, R. T. (1993). Recent advances in approximation concepts for optimum structural design, *Structural Optimization* **5**(3): 129–144.

Belytschko, T., Lu, Y. Y. and Gu, L. (1994). Element-free galerkin methods, *International Journal for Numerical Methods in Engineering* **37**: 229–256.

Bernardo, M. C., Buck, R., Liu, L., Nazaret, W. A., Sacks, J. and Welch, W. J. (1992). Integrated circuit design optimization using a sequential strategy, *IEEE Transactions on computer-aided design* **11**(3): 361–372.

Currin, C., Mitchell, T., Morris, M. and Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments, *Journal of the American Statistical Association* **86**(416): 953–963.

Etman, L. F. P. (1992). Some global and mid-range approximation concepts for optimum structural design, *Technical Report WFW 92.099*, Eindhoven University of Technology.

Haftka, R. T. and Gürdal, Z. (1992). *Elements of Structural Optimization*, 3 edn, Kluwer Academic Publishers.

Lancaster, P. and Salkauskas, K. (1981). Surfaces generated by moving least squares methods, *Math. Comput.* **37**: 141–158.

Lu, Y. Y., Belytschko, T. and Gu, L. (1994). A new implementation of the element free galerkin method, *Computer Methods in Applied Mechanics and Engineering* **113**: 397–414.

Roozen-Kroon, P. J. M. (1992). *Structural optimization of bells*, PhD thesis, Eindhoven University of Technology.

Sacks, J., Schiller, S. B. and Welch, J. (1989a). Designs for computer experiments, *Technometrics* **31**(1): 41–47.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989b). Design and analysis of computer exeriments, *Statistical Science* 4(4): 409–435.

Schoofs, A. J. G. (1987). *Experimental design and structural optimization*, PhD thesis, Eindhoven University of Technology.

Schoofs, A. J. G., Klink, M. B. M. and Van Campen, D. H. (1992). Approximation of structural optimization problems by means of designed numerical experiments, *Structural Optimization* 4: 206–212.

Svanberg, K. (1987). The method of moving asymptotes - a new method for structural optimization, *International Journal of Numerical Methods in Engineering* **24**: 359–373.

Welch, W. J. and Sacks, J. (1991). A system for quality improvement via computer experiments, *Commun. statist.-theory meth.* **20**(2): 477–495.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J. and Morris, M. D. (1992). Screening, predicting, and computer exeriments, *Technometrics* **34**(1): 15–25.