Physical and Polynomial Response Surfaces
Frédéric Duprat, Franck Schoefs, Bruno Sudret

To cite this version:

HAL Id: hal-01007519
https://hal.archives-ouvertes.fr/hal-01007519
Submitted on 27 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
7.1. Introduction

Generally, structural reliability analysis is based on the supply of mechanical and probabilistic models and a limit state function. In this chapter, we first define a mechanical model that describes structural behavior. In a general sense, the mathematical transfer function \( M \) allows us to evaluate the influence of loading with the knowledge of input parameters (or stimuli) that describe the structure and its environment. These parameters constitute the vector \( x \). The model response is denoted here as \( y = M(x) \).

Next, we define a probabilistic model for the input parameters that are considered to be poorly known or uncertain, even from a statistical analysis of data samples when they are available, or by expert judgment and a database [JCS 02]. This probabilistic model is characterized by the joint density of the input random variable \( X \), denoted \( f_X(x) \).

Thirdly, and finally, we define a limit state function that mathematically translates the failure criterion against which the structure must be justified. This function is written with the general form \( g(M(X), X') \) and is based on the effect of loading to which we fix limits, gathered in a vector \( X' \).
By denoting $f_{X,X'}$ the joint density of vectors $X$ and $X'$, the objective of the analysis is to evaluate the structural reliability through one estimate such as the probability of failure $P_f$ defined by:

$$P_f = \int_{\{x,g(M(x),x')\leq 0\}} f_{X,X'}(x,x')dx$$  \hfill [7.1]

Several methods exist for solving the problem [DIT 96], [LEM 09]. Among them, in this chapter, only a Monte Carlo simulation and FORM method are presented and used.

### 7.2. Background to the response surface method

Many scientific fields and trends of thought have contributed to the elaboration of the so-called Response Surface Methodology (RSM). The beginnings of the response surface approach appeared a few years before Box & Wilson’s developments [BOX 50]. Various scientific fields were involved:

– animal and vegetal biology, and the building of growth curves [REE 29], [WIN 32], [WIS 39];

– human sciences and the analysis of the response of a population to stimuli [BLI 35a], [BLI 35b], [GAD 33]; these works were based on those of the psychiatrist Fechner in 1860;

– agronomy and the study of soil fertilization [CRO 41], [MIT 30], [STE 51].

These approaches were based mainly on the basic assumption (mathematically justified by the Weierstrass’s theorem) that, under some conditions of regularity, a response can be represented by polynomials. Thus, within this context, in 1951 the chemists Box and Wilson developed the concept of a response surface, relying both on analytical regression techniques and the building of experiences. In particular, they had already carefully described the need to pay attention to the choice of stimuli variables, and to the allocation of their relative weight. The empirical models they developed have been enriched by the definition of observation periods [BOX 55] and by error computation [BOX 57]. With the increase of the number of potential models, selection criteria were provided [BOX 59a] such as the generalized variance minimization of variable estimation [BOX 59b]. The period 1950–1970 was a fruitful one with the appearance of three major scientific developments with probabilistic insights:

– research into optimal functional representation through stochastic approximation in the presence of outliers; an expansion to multivariate problems was proposed [KIE 52], [ROB 51];
– the comparison of growth curves in biometry [ELS 62], [RAO 58], where response functions come from projections on a family of orthogonal polynomials. Their coefficients are then used for forecast studies;

– the theory of optimal models under constraint in the case of linear models [KIE 60]. The optimization of a response function was linked to the minimization of the generalized variance of parameters.

This last trend of thought offers a well-structured theoretical contribution that was soon used as a reference by others. Numerous works about optimization appeared after it [ANS 63], [NEL 65], [POW 65]. The end of this period saw the emergence of non linear models. The most significant works were certainly carried out on inverse polynomials [NEL 66] and the statistical estimation of parameters from experimental processes [ATK 68]. Nevertheless, the increase in use of nonlinear models was really significant, due to the increase in computational capacity of computers. The criteria for model validation appear to be very specific to each application field; review papers in biometry [MEA 75] and in the nuclear field [HEL 93] can be cited as examples.

7.3. Concept of a response surface

7.3.1. Basic definitions

The term “response surface” denotes the wish to develop a formal representation based on geometrical ideas; it is surface building in the probabilistic space of the response of a physical process to stimuli. The property being studied, or response $Y$, is the result of a transfer function that characterizes the sensitivity of a system to input parameters. This response then varies with the variation of input parameters known as stimuli. These are modeled by random fields or variables, denoted $X_i$, $i=1,...,n$, and then characterized by a set of available statistical information, denoted $θ_j$, $j=1,..., p$, (independent or correlated probability density functions, normalized moments, etc.). These random variables (or fields) are called basic random variables (or fields). This transfer is represented in Figure 7.1.

Random fields or variables $X_i$  $\rightarrow$ Transfer function $\rightarrow$ Response $Y$

Figure 7.1. Response of a transfer function to stimuli, modeled by random fields or variables
The modeling of random fields is sometimes needed for the representation of spatio–temporal variations of uncertain input parameters. Modeling with random variables, which is simpler, is nevertheless sufficient for specific problems. For simplicity in the following sections, we use random variables to represent input parameters.

Generally, knowledge of a transfer function as an explicit form of basic random variables is not available. We therefore look for an approximation, called a response function, \( M \), which is often selected from amongst a family of usual functions, linear or not, characterized by random or deterministic parameters \( \chi_k, k=1,\ldots,l \). These parameters are deduced from the fitting of the response to the experimental data. The geometrical representation, with a curve, a surface, or a hypersurface, is called a response surface. The introduction of geometrical tools such as contour lines onto this response surface can then be used as frontiers of the safety domain. To build a response surface, we must provide:

– \( X = \{X_1, \ldots, X_n\} \), a ranked set of representative random variables;
– \( \theta = \{\theta_1, \ldots, \theta_p\} \), a set of statistical information about \( X \) (independent or correlated probability density functions, normalized moments, etc.);
– \( M(X/\theta) \), an approximation of the response \( Y \), formulated as an explicit function of \( X \) knowing \( \theta \), and obtained by the fitting of the set of parameters \( \chi \);
– \(|.|\), a metric in the probabilistic space of basic random variables and responses.

The quality of fit of the approximation \( M \) to the response \( Y \) is then measured.

The response function can then be formally written as in Figure 7.2.

\[
M(X/\theta) = M(\chi, X/\theta)
\]

Figure 7.2. Formal writing of a response function

### 7.3.2. Various formulations

The choice of the type of formulation for the response function is made on the basis of specific criteria coming from the selected scientific methodology for
studying the phenomenon. The first questions are the conceivable level of complexity, the availability of a complementary experimental approach and the actual state of knowledge. Two methods are possible but the building of a response surface is more and more frequently based on a mixed solution:

– the matching of an approximated model of the transfer by using usual mathematical functions (especially polynomials) and a selected database. [SCH 96] presents and compares the usual models used;

– the use of underlying deterministic physical laws in which random variables are introduced to account for intrinsic variations (height and period of a wave, speed of the wind for instance) or for uncertainties on modeling parameters [LAB 96]. Two issues then govern the building of the response surface, the physical meaning of the deterministic models and the selection of the random variables, i.e. those that govern the variations of the studied quantities.

Finally, the difficulty of statistically characterizing the basic variables adds to the difficulty of selecting the analytical formulation of the transfer. In the case of response surfaces that describe limit states, this question conditions the reliability measure. For instance, in the field of unidirectional laminated composite materials loaded in the direction or orthogonal to the direction of fibers, we consider generally that a limit state can be deduced from three criteria: stresses, external loading and the geometrical size of the material. Various assumptions on the number of random variables and the typology of their distribution can lead to variations of more than 30% of the value of corresponding safety factors [NAK 95].

7.3.3. Building criteria

Building criteria are specific to each application field. Thus, in the following discussion, criteria are ranked according to their importance for problems relative to the safety of buildings and structures. These criteria nevertheless raise questions that can be extended to other fields. Further, it would be wrong to consider that a unique solution exists. The final choice is the result of an optimization under constraints that we have proposed.

We aim especially to take benefit of the increased power and computational capacities of computers which offer the ability to refine the mathematical representation and control the intrinsic uncertainties due to model fitting. However, one must always keep in mind the requirement of the physical meaning. Thus, the major elements in our approach are:

– the physical meaning of the representation;

– the effects of the choice of probabilistic modeling;
– the measure of the quality of the fit;
– a reduction of the level of complexity, consistent with acceptable computational costs.

7.3.3.1. Physical meaning of the representation

An understanding of the physical mechanism that underlies the physical phenomenon is fundamental when choosing the set of input variables and the approximation function. This criterion can lead to the necessity to base the formulation of the function on deterministic relationships. Intrinsic randomness is then introduced through random variables and we account for the model uncertainty through random parameters. The use of deterministic relationships and a careful selection of basic variables, if an analytical relationship is available, are shown to be more realistic than the use of fitted models with the usual mathematical formulations.

7.3.3.2. Effects of the choice of probabilistic modeling

The probability law of the system response depends on the probabilistic characterization of the input parameters (probability distribution, scatter, skewness, kurtosis, etc.). In a simple case where the transfer function \( M \) is a linear function of normally distributed random variables, the response is normally distributed too. In the general case, random variables are non-normally distributed and the transfer function is more or less nonlinear.

To control the effect of this degree of nonlinearity of the transfer function, several works suggest approaching this function with a polynomial approximation (generally linear, quadratics or cubic). A linear approximation is usually not sufficient [BOU 95]; cubic or fifth order approximations allow us to assess in some cases the moments of third and forth order in a satisfactory way. Polynomial approximations can be of high order and costly in terms of identification of their parameters.

Actually, to guarantee the transfer of distribution laws, it is necessary to control the good fit of the Jacobian matrix \([D(X)/D(Y)]\) [LAB 95], [SCH 08]. We then consider a response surface with a single variable of the form \( Y = M(X) \) with \( M \), a bijective monotonic function which can be derived, and \( X \) a basic variable. Knowing the probability density \( f_X \) of \( X \), we can compute \( f_Y \), the probability density of \( Y \). \( G \) is the cumulative function associated with the probability density function \( g \), and we know that:

\[
GY(y) = P(Y < y) \quad \text{and} \quad f_Y(y) = \frac{d}{dy} P(Y < y)
\]
\[ f_Y(y) = P[X < M^{-1}(y)] \text{ and finally } f_Y(y) = f_X[M^{-1}(y)] \cdot \frac{1}{M'[M^{-1}(y)]} \]

where \(|.|\) denotes the absolute magnitude.

In the more general case of a multivariate problem for which \(X\) and \(Y\) are vectors, probability density functions of random input and output are linked by the relationship:

\[ f_Y(y) = f_X(x) \cdot [M^{-1}(y)] \cdot \begin{vmatrix} D(X) \\ D(Y) \end{vmatrix} \]

Thus, to obtain a good approximation of \(P_Y\), a good fit of function \(M\) should be reached as well as a fitting of the partial derivatives. The linear, quadratic or cubic polynomial functions described before have, respectively, constant, linear and quadratic derivatives. The difference between these three functions is generally significant near the bounds of the studied domain (realizations of basic variables), and the perturbations on distribution tails can be significant and thus modify the results of reliability computation. Thus, the choice of low order polynomial, very convenient from a computational point of view, can lead to false probability functions for the response even they seem to correctly represent the trend. Such a choice is very sensitive to the effects of the choice of probabilistic modeling of input parameters.

7.3.3.3. Measurement of the quality of fit

We aim here to define a metric (a measuring tool) that gives a rational tool to quantify the quality of fit. Usual metrics, known as second order metrics, allow us to obtain the variables that are dominant in the response because we can quantify their influence on the variance of the response. They are thus not very effective when singular events external to the distribution functions of the input variables occur, and they only give an indication of the measure of uncertainty [BIE 83], [IMA 87]. To solve these shortcomings, metrics based on inter-quantile discrepancies [KHA 89] or on the measurement of the system entropy [PAR 94] are available.

Another approach consists of the use of regression metrics. Let us denote the response function for which we want to fit the parameters as \(f\), and the error as \(\varepsilon\). Let us consider, for example, a regression model such as:

\[ Y = M(X/\theta) + \varepsilon \]
When the error $\varepsilon$ is supposed to be normally distributed with 0 mean and a diagonal covariance matrix, then the fitting of $f$ with least square and maximum likelihood methods are identical. The $L_2$ metric (integration of the square of the residual $u$) is then the most efficient. The $L_1$ metric (integration of the norm of the residual) is more efficient for an exponential distribution of the error, which is then more scattered.

It has been shown in the previous section that the fitting of a Jacobian matrix of partial derivatives is very interesting. Thus, it also seems very interesting to choose a metric already available in variational theory:

$$
\|u\| = \sqrt{\|u\|_{L_2}^2 + \sum_{i=1,...,n} \left\| \frac{\partial u}{\partial X_i} \right\|_{L_2}^2}
$$

with $\|\cdot\|_{L_2}$, $L_2$ norm in the Sobolev space $H_1$:

$$
H_1 = \{ u \frac{\partial^\alpha u}{\partial X_i^\alpha} \in L^2 \text{ (second order integrable) for } \alpha = 0, 1 \text{ and } i = 1,...,n \}
$$

The underlying idea of this choice of metric is thus to prefer the better control of the distributions tails through successive transfers, in comparison to the control of the central part, by the fitting of the first moments. Every building of a response surface should be suggested with a metric that conditions the sense of the approximation and allows us to explain some limits in the representation.

7.3.3.4. Reduction of complexity level and tractability for computations

To gain accuracy, models of a high order could be interesting. This increase in the computational procedures (optimization algorithms for the fitting under constraint) must, however, be justified: this increase in the complexity level leads to an increase of the computational costs that should be kept as reasonable as possible.

For more details, this question is illustrated in [SCH 07] through several studies concerning wave–structure interaction, which look at the effects of the order of the Stokes kinematics model, of accounting for the inertia term in load computing, and the number of elements needed for the integration of distributed loading on the beam.
7.4. Usual reliability methods

7.4.1. Reliability issues and Monte Carlo simulation

Structural reliability aims to assess the failure probability of a mechanical system in a probabilistic framework according to a failure scenario. Structural reliability methods allow not only the probability of failure \( P_f \) to be computed for a single component of a system, or for a system as a whole then involving several interacting components, but also the sensitivity of this probability against each random variable of the problem to be determined [DIT 96], [LEM 09].

A failure criterion can be expressed thanks to a performance function \( g : \mathbb{R}^M \rightarrow \mathbb{R} \) conventionally described by parameters in such a way that \( D_f = \{ x : g(x) \leq 0 \} \) defines the *failure domain* and \( D_s = \{ x : g(x) > 0 \} \) defines the *safety domain*. The frontier \( \partial D = \{ x : g(x) = 0 \} \) is the *limit state surface*. The probability of failure is:

\[
P_f = \mathbb{P} \left( g(X) \leq 0 \right) = \int_{D_f} f_X(x) \, dx
\]  

where \( f_X \) is the probability density function of \( X \).

When the performance function is analytically expressed, for instance when a response surface is implicated in the expression of \( g \), then the integral over the implicitly defined integration domain is numerically feasible from Monte Carlo simulation: \( N_{\text{sim}} \) simulations of the input vector \( X \) are supplied, and for each of them \( g \) is evaluated. If \( N_f \) is the number of simulations for which \( g \) is negative, the probability of failure \( P_f \) is approximated by the ratio \( N_f / N_{\text{sim}} \). The method is fairly simple but computationally very costly: about \( N_{\text{sim}} \approx 4.10^{k+2} \) samples are required if a 5% accuracy is expected when \( P_f \) is about \( 10^{-k} \). In usual practical applications, \( k \) lies from 2 to 6, which compromises the use of the method, except if a consistent response surface is available. Among methods which have been developed to circumvent this drawback, the First Order Reliability Method (FORM) is one of the most employed.

7.4.2. FORM

FORM provides an approximate value of the failure probability by recasting the problem in a *reduced centered Gaussian space*, where all random variables \( \xi \) are
Gaussian with zero mean and unit standard deviation. An iso-probabilistic transformation \( T \) is needed for this goal \( T: X \rightarrow \xi (X) \). For independent random variables with marginal cumulative probability function \( F_{X_i}(x_i) \), this transformation simply reads \( \xi_i = \Phi^{-1} \left( F_{X_i}(x_i) \right) \) where \( \Phi \) is the cumulative reduced centered Gaussian probability function. For correlated random variables, Nataf or Rosenblatt transformations can be resorted to [LEM 09], (Chapter 4). Equation [7.2] then becomes:

\[
P_f = \int_{\xi: G(\xi)=g(T^{-1}(\xi))\leq 0} \phi_M (\xi) \, d\xi_1 \ldots d\xi_M
\]

[7.3]

where \( G(\xi) = g \left( T^{-1}(\xi) \right) \) is the performance function in the reduced space and \( \phi_M \) is the reduced centered multinormal probability density function of dimension \( n \), defined by

\[
\phi_M (x) = \left( 2\pi \right)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2} \left( x_1^2 + \ldots + x_n^2 \right) \right].
\]

A maximum value of \( \phi_M \) is encountered at the origin of the reduced space, and \( \phi_M \) exponentially decreases with respect to the distance \( \| \xi \|^2 \) from the origin, all the more since the number of variables is high. In the failure domain, the points contributing most to the integral [7.3] are therefore those which are closest to the origin. The second stage of FORM is to determine the so-called design point \( P^* \) which is the point of \( D_f \) closest to the origin, and so the most probable failure point. This point is the solution of the optimization problem:

\[
P^* = \text{Arg min}_{\xi \in \mathbb{R}^n} \left\{ \frac{1}{2} \| \xi \|^2 / G(\xi) \leq 0 \right\}
\]

[7.4]

A suitable optimization under constraint algorithm can solve [7.4]. The reliability index \( \beta \) is then defined as the algebraic distance from the origin to the limit state surface \( \partial D : \beta = \text{sign} (G(0)) \| \xi^* \| \). Once \( \xi^* \) is determined, the limit state surface \( \partial D \) is approximated by a hyper-plan tangential at \( P^* \). The integral [7.3] is then reduced to \( P_f \approx P_{f, \text{FORM}} = \Phi(-\beta) \). The linearized limit state is expressed by \( \tilde{G}(\xi) = \beta - \alpha^T \cdot \xi \). The unit vector of direction, cosine \( \alpha \), which is perpendicular to the hyper-plan, allows the sensitivity factors to be computed by \( \alpha_i^2 \) for each independent random variable \( X_i \).
FORM is fruitful because it leads to a fairly good approximate value of $P_f$ with a reasonable computational cost. The approximation is better when $\beta$ is large, and is often relevant and satisfactory provided that the design point $P^*$ is consistent (i.e. with no local minima). Moreover, sensitivity measures and importance factors are given which are matters of interest for the designer.

Readers are invited to refer to [LEM 09] to investigate SORM (Second Order Reliability Method), based on a quadratic approximation of the limit state function. SORM is \textit{a priori} more costly than FORM, but also more accurate.

7.5. Polynomial response surfaces

Amongst other possible forms of response surfaces (RS), polynomial response surfaces have been widely used for reliability problems in mechanics. In the following section, only simple polynomial response surfaces are addressed, in contrast to those based on polynomial chaos in Chapter 8.

7.5.1. Basic formulation

If $X=\{X_i; i=1,\ldots,M\}$ is the random variables vector, the quadratic response surface $\hat{g}(X)$ of the true limit state function $g(X)$ is expressed by:

$$\hat{g}(X) = a_0 + \sum_{i=1}^{M} a_i X_i + \sum_{i=1}^{M} a_{ii} X_i^2 + \sum_{i=1}^{M} \sum_{j=1, j\neq i}^{M} a_{ij} X_i X_j$$  \[7.5\]

where $a=\{a_0, a_i, a_{ii}, a_{ij}\}^T$ is the vector of unknown coefficients. These coefficients are obtained by least square method from the Numerical Experimental Design (NED) \{x^{(k)}, k=1,\ldots,N\}, where the number of sampling points $N$ is at least equal to the dimension of $a$:

$$a = \arg \min_a \sum_{k=1}^{N} (y_k - \hat{g}(x^{(k)}))^2$$  \[7.6\]

where $y_k = g(x^{(k)})$ is the value of the true limit state function at point $x^{(k)}$. From

$$\hat{g}(X) = \begin{bmatrix} 1, X_i, X_i^2, X_i X_j \end{bmatrix}^T \begin{bmatrix} a_0, a_i, a_{ii}, a_{ij} \end{bmatrix}^T = B(X)^T a, \quad a \text{ is computed by} [FAV 89]$$

as:

$$a = (C^T C)^{-1} C^T y$$  \[7.7\]
where \( C \) is the matrix whose rows are the vectors \( B(x^{(k)})^T \) and \( y \) is the vector of components \( y_k \).

The basic formulation is simple. Nevertheless, some difficulties and controversial discussions arise when we consider the following points:

– working space. In order to assure the consistency of the NED, in such a way that the mechanical model behaves well, a physical space is preferable. Nonetheless, building the NED in a standardized space is easier. Indeed such a space is non-dimensional and allows the distance between sampling points to be efficiently controlled, which is useful to avoid the ill-conditioning of \( C \). Moreover, it is a natural space for determining the reliability index;

– number of sampling points. When increasing the number of sampling points the least squares regression is improved (F-statistics and variance of the unknown coefficients \( a \)) but not necessarily the quality of the approximation (adjusted \( R^2 \) or crossed \( Q^2 \) correlation coefficients), which is better if the number of sampling points is just equal to the number of unknown coefficients. Outside of the NED, the relevance of the RS (Response Surface) is improved when the number of sampling points is higher, but if reducing the computational cost is an aim (compared to the cost obtained when the true failure function is used), then the number of sampling points should be minimized;

– topology of the NED. When the NED is compact, the quality of the approximation is improved but the domain of suitability is reduced. The location of the sampling points should depend on the behavior (the sensitivity) of the true limit state function;

– validity of the RS. Depending on what is being sought (reliability index and/or assessment of the failure probability) the validity of the RS should be estimated locally or globally;

– adaptability of the RS. It is often necessary to rebuild the NED because the domain of final utilization of the RS (the region of the most probable failure point) is far from the mean point (in case of low failure probability), which usually plays the role of initial central point of the sampling grid. A sequential procedure is therefore needed to rebuild the NED with respect to criteria for the consistency of the RS and in conjunction with probabilistic results or procedures;

– the order of the polynomial should be less than or equal to the unknown degree of nonlinearity of the true failure function, in order to facilitate the solving of the linear system;

– the presence of mixed terms in the expression of the RS contributes to capturing the effect of interaction between variables.
7.5.2. Working space

In the framework of structural reliability two working spaces are possible: the physical space \( X \) (dimensional random variables with any distribution and possible correlation) and the standardized space \( U \) (reduced centered Gaussian and uncorrelated random variables). Both spaces are simultaneously required to compute the unknown coefficients only if the RS is built in a standardized space (i.e. if the mechanical model is indeed defined in a physical space). The transformation from \( X \)-space to \( U \)-space is nonlinear, and modifies the topology of the NED and the failure surface as well [DEV 97] (see Figure 7.3 in the case of a factorial design). Numerous authors have hence chosen to operate in a physical space [BUC 90], [GAV 08], [KAY 04], [KIM 97], [MUZ 93], [RAJ 93]. If the searching procedure for the reliability index is associated with the sequential procedure for building the NED, it is, however, clearly preferable to operate in a standardized space [DEV 97], [DUP 06], [ENE 94], [GAY 03], [GUP 04], [NGU 09].

![Figure 7.3. Effect of the space shift](image)

7.5.3. Response surface expression

Whatever the working space chosen, the objective is to provide a satisfactory estimate of the reliability and, if possible, with a lower computational cost than that resulting from the use of the true limit state function. To do this, if second order estimates of the reliability are sought, linear and quadratic polynomial RS are good candidates. Respectively, the expressions of the RS are:

\[
\hat{g}_I(X) = a_0 + \sum_{i=1}^{M} a_i X_i \\
\hat{g}_{II}(X) = a_0 + \sum_{i=1}^{M} a_i X_i + \sum_{i=1}^{M} a_{ii} X_i^2 + \sum_{i=1}^{M} \sum_{j=i+1}^{M} a_{ij} X_i X_j
\]

[7.8]
Nevertheless if Monte Carlo simulations are expected to be used to assess the failure probability in a more or less large zone around the most probable failure point, or if the degree of nonlinearity of the failure function is from evidence higher than 2, increasing the order of the polynomial could be required [GAV 08], [GUP 04]. One method proposed to select the best-fitted order of polynomial involves the use of Chebyshev polynomials, a statistical analysis of the Chebyshev polynomial coefficients, and a statistical analysis of high-order RS [GAV 08]. The increase in accuracy brought by high-order RS is, however, counteracted by the added computational cost of the selecting procedure.

7.5.4. Building the numerical experimental design

7.5.4.1. Number and layout of the sampling points

The minimum number of points is the number of unknown coefficients, namely $N = (M+1)$ for a linear RS, $N = (2M+1)$ for a quadratic RS without mixed terms and $N = (M+1)(M+2)/2$ for a quadratic RS with mixed terms. With such a minimum number of points, only an interpolation is provided and the statistical significance of the coefficients is very poor. On the other hand, the use of the RS beyond the frontier of the NED is not expected to be consistent. Even if saving runs of the true failure function is one of the objectives, it is not desirable to limit the number of points to the minimum, due to the resulting poor quality and suitability of the RS.

A uniform layout of the sampling points around a central point is commonly adopted at the initial stage of building (central, central composite or factorial design). The number of sampling points of the NED depicted in Figure 7.3 is equal to $(2M+1+2M)$ or $(3M)$ depending on whether one or more points are located out of the axes. Along the axis $i$ the distance between points is stated as a function of the standard deviation of the variable:

$$x_i^{(k)} = X_{C,i} \pm h_i \sigma_i$$  \[7.9\]

where $x^{(k)}$ is the $k^{th}$ point of the NED and $X_C$ is the central point. Similarly, in a standardized space:

$$u_i^{(k)} = U_{C,i} \pm h_i$$  \[7.10\]

Except in the case of adaptive procedures, the factor $h_i$ is constant ($h_i = h$).

Too large or too small values of $h$ compromise the quality of the RS and lead to erroneous results. An example is given in [GUA 01] where 21 random variables are involved (for the reliability of a portal frame) and a quadratic RS without mixed
the failure probability shifts from $10^{-4}$ to 0.94 when $h$ varies globally from 0.1 to 5. When $h$ varies locally from 1.8 to 2.2, the failure probability and the reliability index, respectively, shift from $9.3 \times 10^{-5}$ to 0.37 and from 0.06 to 4.47.

In order to avoid the ill-conditioning of the system $C$ [1.7] the value of $h$ must be large enough. Values of $h$ lying in the range 1 to 3 are often suggested [BUC 90], [DEV 97], [DUP 06], [ENE 94], [KAY 04], [KIM 97], [MUZ 93], [NGU 09], [WON 05]. Minimum values of $h$ have been also proposed in the range 0.2 to 0.5 [DUP 06], [ENE 94].

7.5.4.2. Adaptive procedures

The main goal of adaptive procedures is to combine a satisfactory approximation of the true failure function, at least in the neighborhood of the most probable failure point, with a limited computational cost. The following items are part of the adaptation:

– shape of the RS. A first linear ($\hat{g}_I(X)$) or quadratic without mixed terms ($\hat{g}_{II}(X)$ with $a_{ij}=0$) RS is attractive due to the small number of unknown coefficients and the computational cost needed to determine the most probable failure point. A refinement can be then undertaken with a quadratic with mixed terms RS [GAY 03], [NGU 09];

– initial NED. If the role played by the random variables is known a priori (for instance from the engineer’s knowledge), a first central point $X_{C1}$ can be located at $X_{C1,i} = \mu_i - h_i \sigma_i$ if $X_i$ contributes to the non-failure and at $X_{C1,i} = \mu_i + h_i \sigma_i$ if $X_i$ contributes to the failure. When no information is available on the role of variables, a pre-selecting procedure can also be employed with a few runs of the true limit state function [GAY 03];

– second central point. This can be located with respect to the first most probable failure point thanks to an interpolation [BUC 90] or found at the same location provided that the latter is situated inside the first NED [DEV 97], [ENE 94];

– mesh of the sampling grid. The size of the mesh is generally reduced as iterations proceed, in order to concentrate the NED in the region of the most probable failure point [DEV 97], [ENE 94], [GAY 03], [KAY 04], [MUZ 93]. Some considerations can be added in order to refine the mesh according to the number of random variables [KIM 97], the sensitivity of the RS towards the variables [DUP 06], [NGU 09], or specific statistics (confidence interval on the coordinates of the most probable failure point) [GAY 03];

– weighting of the sampling points. The unknown coefficients are computed, introducing a weighting diagonal matrix in equation [7.7]. The weighting factors
depend on the closeness of points to the failure surface [KAY 04] or to the last most probable failure point [NGU 09];

– update of the NED. In order to limit the computational cost, it is of interest to keep in the NED all the points where the value of the true limit state function has already been computed. However, some of them, that could harm the quality of the RS, must be excluded according to suitable criteria [DEV 97], [DUP 06], [ENE 94], [GAY 03], [NGU 09].

7.5.4.3. Quality of the approximation

In order to verify the quality of the response surface, a classic measure of the correlation between the approximate and the exact value of the limit state function is the adjusted correlation factor:

\[
R^2 = 1 - \frac{(N - 1) \sum_{k=1}^{N} (\hat{g}(x^{(k)}) - y_k)^2}{(N - N_c - 1) \sum_{k=1}^{N} (y_{\text{mean}} - y_k)^2}
\]  

[7.11]

where \(y_{\text{mean}}\) is the mean value of the limit state function over the NED. If \(R^2\) is less than 0.9, the quality of the response surface has to be improved. The cross correlation factor \(Q^2\) is also employed (see Chapter 8).

7.5.5. Example of an adaptive RS method

7.5.5.1. General description [NGU 09]

7.5.5.1.1. First iteration

The NED is centered at \(U_C\) and comprises \(U_C\) and one point along each axis located at:

\[
u_i^{(k)} = U_{C,i} + h_i \quad \text{with} \quad h_i = -\frac{h_0}{\|\nabla \hat{g}(U_C)\|} \frac{\partial \hat{g}(U_C)}{\partial U_i}
\]

[7.12]

where \(u^{(k)}\) is the \(k^{th}\) point of the NED and \(h_0\) lies basically between 1 and 3. For the first iteration, the RS is linear in the standardized space \(\hat{g}(U) = \hat{g}_j(U)\), the central point is the origin of the space \(U_{C_1}=U_0\) and a fictitious gradient of \(\hat{g}(U_{C_1})\) is considered, based on engineering knowledge in such a way that \(h_i=\pm 1\).
According to [KAY 04], the weighting diagonal matrix $W$ is introduced in equation [7.7] which becomes:

$$a = \left(C^T WC\right)^{-1} C^T W y$$  \hspace{1cm} [7.13]

where the weighting factors are expressed by

$$W_k = \exp\left(-\frac{g(U^{(k)}) - y_{min}}{g(U_0)}\right)$$  \hspace{1cm} [7.14]

In [7.14], $y_{min}$ is the minimum value of the true limit state function over the NED. Once the unknown coefficients $a$ of $\hat{g}(U)$ are computed by [7.13] the first most probable failure point $U^{*(1)}$ is determined by FORM.

7.5.5.1.2. Second iteration

For the second and further iterations, the RS has a quadratic form $\hat{g}(U) = \hat{g}_{II}(U)$ with $(M+1)(M+2)/2$ being unknown coefficients. As suggested in [BUC 90], the central point of the second NED is stated by:

$$U_{C2} = U_0 + \left(U^{*(1)} - U_0\right) \frac{g(U_0)}{g(U_0) - g(U^{*(1)})}$$  \hspace{1cm} [7.15]

which is closer to the true failure surface than $U^{*(1)}$ ($g(U^{*(1)}) \neq 0$ most of the time).

All the points of the first NED are maintained in the second and $((M+1)(M+2)/2 - (M+1))$ complementary points are added, in a half-star shape design around $U_{C2}$. Among the complementary points, $M$ points are located according to [7.12] which implies that points are situated towards the failure region with respect to the central point and at a distance from the latter proportional to the local sensitivity of the RS. The $((M+1)(M+2)/2 - 2(M+1))$ remaining complementary points are again generated from the $M$ previous ones, each of them playing the role of a new local central point. When applying equation [7.12], the axes are considered in descending order with respect to the components of the gradient vector $\nabla \hat{g}(U_{C2})$. The point $U^{*(1)}$ is kept in the second NED under the condition:

$$\|U_{C2} - U^{*(1)}\| \leq h_0$$  \hspace{1cm} [7.16]
The weighting factors are now computed by:

\[
W_k = \exp \left( -\frac{g(U^{(k)}) - y_{\min}}{g(U_0)} \right) \exp \left( -\frac{\|U^{*(iter-1)} - U^{(k)}\|^2}{2} \right)
\]  

[7.17]

in which the last term allows the closeness to the previous most probable failure point to be accounted for. If the condition [7.16] is not fulfilled, \(U^{*(1)} = U_{c2}\) in [7.17]. Once the unknown coefficients of \(\hat{g}(U)\) are computed by equation [7.13], the second most probable failure point \(U^{*(2)}\) is determined by FORM.

7.5.5.1.3. Further iterations (iter>2)

The NED is enriched with the point \(U^{*(iter-1)}\). The weighting factors and the coefficients of \(\hat{g}(U)\) are updated by equations [7.17] and [7.13] respectively. A new most failure point is determined. The convergence of the procedure is achieved when:

\[
\begin{align*}
\|\beta^{(iter)} - \beta^{(iter-1)}\| & \leq \epsilon_\beta \\
\|U^{*(iter)} - U^{*(iter-1)}\| & \leq \epsilon_P
\end{align*}
\]  

[7.18]

7.5.5.2. Examples

Two examples are reported here in order to show the interest but also the limit of the polynomial response surface method. The method described above is referred to as RSDW (Response Surface with Double Weighting).

7.5.5.2.1. Example 1

A simple explicit limit state function is considered in a standardized space:

\[
g(U) = \exp(0.4(U_2 + 2) + 6.2) - \exp(0.3U_1 + 5) - 200
\]  

[7.19]

The results are reported in Table 7.1 where \(N_r\) denotes the number of runs of the true limit state function. It can be seen that the values of the reliability index are very close to each other for all the methods under consideration. In the same way, the values of \(\left| g(U^*) / g(U_0) \right|\) are close to zero and express a good closeness of the most probable point with regard to the failure surface. The cumulative formation of the RS is depicted in Figure 7.4. The effect on the quality and efficiency of the initial grid size \(h_0\) can be seen in Tables 7.2 and 7.3. It is worth noting that the use of
a weighting system allows this effect to be mitigated and a satisfactory closeness to
the true failure surface to be obtained

<table>
<thead>
<tr>
<th>Method</th>
<th>β</th>
<th>(U^*_{1})</th>
<th>(U^*_{2})</th>
<th>(\frac{g(U^*)}{g(U_0)})</th>
<th>(N_r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive MC [KAY 04]</td>
<td>2.710</td>
<td>0.969</td>
<td>-2.531</td>
<td>3.36×10^{-5}</td>
<td>-</td>
</tr>
<tr>
<td>RS [KIM 97]</td>
<td>2.691</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RS [KAY 04]</td>
<td>2.686</td>
<td>0.820</td>
<td>-2.558</td>
<td>5.84×10^{-3}</td>
<td>8</td>
</tr>
<tr>
<td>RS [DUP 06]</td>
<td>2.710</td>
<td>0.951</td>
<td>-2.538</td>
<td>9.10×10^{-4}</td>
<td>21</td>
</tr>
<tr>
<td>RSDW ((R^2 = 0.997))</td>
<td>2.707</td>
<td>0.860</td>
<td>-2.567</td>
<td>8.48×10^{-4}</td>
<td>12</td>
</tr>
</tbody>
</table>

**Table 7.1.** Comparison between several RS methods (Example 1)

![Cumulative formation of the response surface](image)

**Figure 7.4.** Cumulative formation of the response surface (Example 1)

<table>
<thead>
<tr>
<th>(h_0)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta)</td>
<td>2.707</td>
<td>2.724</td>
<td>2.713</td>
<td>2.714</td>
</tr>
<tr>
<td>(\frac{g(U^*)}{g(U_0)})</td>
<td>8.48×10^{-4}</td>
<td>4.06×10^{-4}</td>
<td>8.02×10^{-4}</td>
<td>6.88×10^{-4}</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.997</td>
<td>0.999</td>
<td>0.999</td>
<td>0.930</td>
</tr>
<tr>
<td>(N_r)</td>
<td>12</td>
<td>10</td>
<td>11</td>
<td>13</td>
</tr>
</tbody>
</table>

**Table 7.2.** Influence of \(h_0\) with a weighting system \((W_k\neq 1)\) (Example 1)
\[
\begin{array}{|c|c|c|c|c|}
\hline
h_0 & 1 & 2 & 3 & 4 \\
\hline
\beta & 2.707 & 2.715 & 2.715 & 2.726 \\
\hline
\frac{g(U^*)}{g(U_0)} & 1.15 \times 10^{-3} & 1.45 \times 10^{-3} & 1.12 \times 10^{-3} & 3.16 \times 10^{-3} \\
\hline
R^2 & 0.997 & 0.999 & 0.999 & 0.885 \\
\hline
Nr & 14 & 11 & 11 & 15 \\
\hline
\end{array}
\]

**Table 7.3. Influence of \( h_0 \) without weighting system (\( W_k=1 \)) (Example 1)**

\[
\begin{array}{|c|c|c|c|}
\hline
u_{\text{max}} (\text{cm}) & 4 & 5 & 6 \\
\hline
\beta & \text{RSDW} & 2.19 & 2.86 & 3.28 \\
& \text{FORM} & 2.20 & 2.86 & 3.40 \\
\hline
\frac{g(U^*)}{g(U_0)} & \text{RSDW} & 5.39 \times 10^{-3} & 2.80 \times 10^{-3} & 1.06 \times 10^{-3} \\
& \text{FORM} & 6.03 \times 10^{-6} & 6.25 \times 10^{-4} & 1.72 \times 10^{-5} \\
\hline
Nr & \text{RSDW} & 256 & 258 & 255 \\
& \text{FORM} & 80 & 75 & 104 \\
\hline
R^2 & 0.993 & 0.994 & 0.993 \\
\hline
\end{array}
\]

**Table 7.4. Comparison between RS and direct FORM (Example 2)**

7.5.5.2.2. Example 2

An implicit limit state function is considered which involves the top displacement of a multi-storey and multi-span steel frame. This example is also presented in Chapter 8 (section 8.4.2) where the distributions of 21 correlated variables can be found. The results are reported in Table 7.4, where \( u_{\text{max}} \) denotes the threshold value of the displacement. As far as the convergence and closeness to the failure surface are concerned, it can be noted that RSDW is satisfactory. From a comparison with direct FORM (direct coupling between the Rackwitz–Fiessler algorithm and the true limit state function) it can be said that, on one hand the RS method supplies consistent values of the reliability index but, on the other hand, the computational cost is significantly higher. For such an example, RSDW is therefore less efficient than direct FORM.
7.6. Conclusion

The response surface method was first developed for any system whose response to stimuli could not be satisfactorily captured by explanatory models, and was employed for biological systems. When applying the response surface method to mechanical systems, the main goal is to reduce the computational cost resulting, for example, from the use of finite element explanatory models. Some specific developments of the response surface method have been carried out in the context of probabilistic reliability analysis. The examples reported in this chapter show that these developments are not fully efficient in terms of computational cost if the number of variables exceeds twenty. Nonetheless, the quality of the adaptive response surface presented above is sufficiently good for the probabilistic results to be very close to those obtained by the use of the true limit state function. Moreover, having an explicit approximate failure function in the region where the most failure point is located facilitates probabilistic post-processing, including assessment of the failure probability, sensitivity analysis, etc.

7.7. Bibliography


[SCH 96] SCHOEFS F., Surface de réponse des efforts de houle dans le calcul de fiabilité des Ouvrages, PhD thesis, ED 82-208, University of Nantes, France, 21 November 1996.


