Experimental design and structural optimization

EXPERIMENTAL DESIGN AND STRUCTURAL OPTIMIZATION

A.J.G. SCHOOFS

EXPERIMENTAL DESIGN AND STRUCTURAL OPTIMIZATION

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN
DOCTOR AAN DE
TECHNISCHE UNIVERSITEIT EINDHOVEN,
OP GEZAG VAN DE RECTOR MAGNIFICUS,
PROF DR. F. N. HOOGE, VOOR EEN COMMISSIE AANGEWEZEN
DOOR HET COLLEGE VAN
DEKANEN IN HET OPENBAAR TE VERDEDIGEN OP
VRIJDAG 28 AUGUSTUS 1987 TE 16.00 UUR

DOOR

ALBERTUS JOHANNES GERARDUS SCHOOFS

GEBOREN TE BLADEL

Dit proefschrift is goedgekeurd door de promotoren:
Prof. dr. ir. D. H. van Campen
en
Prof. dr. ir. J. D. Janssen
Co-promotor:
Dr. A. Lehr

Febodruk, Enschede
ISBN 90-71382-19-2

Aan Anneke, Paul, Stan en Ivo

CONTENTS

Abstract

A widely used method for automated structural optimization consists of the combination of two important numerical methods, namely:

1. the finite element method (FEM) as a flexible and accurate tool for modelling and analysis of structures.
2. mathematical programming as a structured method for the search of an improved set of design variables.

Structural optimization problems almost always are essentially nonlinear, demanding an iterative solution method: the structure is analysed and, subsequently, improved values for the design variables are estimated. This process is repeated until, in general, a local optimum has been achieved. Here, our main interest is the final result; afterwards the intermediate iteration steps are of little importance.

If we accept that several FEM-analyses should be spent to solve the optimization problem, another approach can be used, too. With that approach the FEM-analyses to be carried out are a priori planned, both concerning their number and the values of the relevant design variables. From the outcomes of the computations an approximating mathematical model is derived by means of regression analysis. The derived models will be linear in the parameters; mostly polynomial models are used.

For the planning of the FEM-computations and for the analysis of the outcomes, the experimental design theory will be applied. This theory has been developed for the planning and analysis of comprehensive physical experiments. FEM-computations can be considered numerical experiments and can be used to establish a numerical experimental design, with the purpose to derive an efficient mathematical model of the structure under consideration. Such mathematical models can be used in that capacity, for example, in a design office, or they can be applied as a fast analysis module in optimization software.

In this thesis common strategies are discussed concerning structural optimization and experimental design. Furthermore, modifications of the experimental design theory will be treated, which are necessary and useful on behalf of numerical experimental designs. The integration of experimental design and structural optimization is argued.

Possible applications of the developed methods will be defined and guidelines are given for using them.

The methods have been tested and used extensively for the shape optimization of church and carillon bells, resulting in several new and musically very interesting bell types. Furthermore, the methods have been applied successfully to several mechanical engineering problems and to a biomechanical problem.

List of symbols

~	a tilde denotes a column matrix
•	a dot denotes differentiation with respect to time
-	an underscore denotes a stochastic variable
^	a circumflex denotes an estimated quantity
*	an asteriks denotes optimality (Ch. 2)

Operators

$\underset{\sim}{\nabla}$	column with gradient operators
∂	partial differentiation
A^T	transposed of matrix A
tr(A)	trace of matrix A
det(A)	determinant of matrix A
E($\underline{y}$)	expected value of stochastic variable $\underline{y}$
V($\underline{y}$)	variance of stochastic variable $\underline{y}$

A, AB, ...	factors or effects
B	damping matrix (Ch. 2)
	bias term of error (Ch. 3)
bce	equality behaviour constraint
bci	inequality behaviour constraint
C	normalized optimality criterion (Ch. 3)
c	frequency ratio in cents (Ch. 6)
dce	equality design constraint
dci	inequality design contraint
d($\underset{\sim}{x}$, $\underset{\sim}{n}$)	normalized response variance
D	design space
D_b	behaviour constrained space
D_d	design constrained space
D_f	feasible design space
e, $\underset{\sim}{e}$, $\underset{\simeq}{e}$	errors between observations and model
f	function or relationship
	eigenfrequency (Ch. 6)
f_i, $\underset{\sim}{f}$	model functions, regressor variables
F($\underset{\sim}{x}$)	objective function (Ch. 2)
	matrix containing model functions (Ch. 3)
$\underline{F}_i$	F-test value

$\underset{\sim}{G}$	gradient of objective function F
g	inequality constraint (Ch. 2)
$g(\underset{\sim}{x}, \underset{\sim}{\beta})$	interpolation function (Ch. 3)
H	Hessian matrix
I	identity matrix
J	number of active constraints (Ch. 2)
	criterion for robust designs (Ch. 3)
k	number of model parameters in $\underset{\sim}{\beta}$
k_j	idem, however with respect to response quantity j
K	stiffness matrix
KS_r	residual sum of squares
l_d	number of equality design constraints
l_b	idem, however concerning behaviour constraints
m	number of response quantities (Ch. 3)
	number of inequality constraints (Ch. 2)
md	idem, however with respect to design constraints
mb	idem, however with respect to behaviour constraints
M	mass matrix (Ch. 2)
	information matrix (Ch. 3)
n	number of design or control variables (Ch 1, 2, 3)
N	number of observations (Ch. 3)
N_j	idem however, under treatment j
$\underset{\sim}{N}$	column with numbers of observations
$N(0, \sigma^2)$	normal distribution with zero mean and variance σ^2
$\underset{\sim}{p}$	column containing prescribed loads
q	iteration step number
r	penalty factor (Ch. 2)
	number of candidate points (Ch. 3)
$\underset{\sim}{r}$	column with design variables concerning shape of actual reference curve of bell profile
r_i	radius of point i on reference curve of bell profile
S	diagonal matrix containing standard deviation of response errors
$\underset{\sim}{s}$	search direction column
$\underset{\sim}{t}$	stochastic variable following Student's t-distribution (Ch. 3)
t_i	wall thickness of bell at point i (Ch. 6)
$\underset{\sim}{t}$	column with design variables concerning wall thickness

t_n	nominal wall thickness
t_a	actual wall thickness
$\underset{\sim}{u}$	column with displacements (Ch. 2)
	column with response quantities (Ch. 3)
$\underset{\sim}{v}$	eigenvector (Ch. 2)
$\underset{\sim}{w}$	column with weighted response quantities (Ch. 3)
$W(\underset{\sim}{x})$	matrix containing weighted model functions (Ch. 3)
x_i, $\underset{\sim}{x}$	design variables or control variables (coded or standardized input variables) (Ch. 3)
x_A	variable describing effect A
X	design matrix
y, $\underset{\sim}{y}$	response quantities
$\underset{\sim}{z}_j$	column containing k_j zeros (Ch. 3)
α	stepsize parameter
β	objective function in Zoutendijk's method (Ch. 2)
β_i, $\underset{\sim}{\beta}$	model parameters (Ch. 3)
γ	exponent in penalty function
	parameters in model of frequencies as function of wall thicknesses (Ch. 6)
δx_i	perturbation of variable x_i (Ch. 2)
δ_{ij}	Kronecker δ (Ch. 3)
Δq	increment of stepsize α in step q
$\underset{\sim}{\eta}$	column with weighted errors
$\kappa(\underset{\sim}{\xi}, \theta)$	true physical relationship
λ	eigenvalue of information matrix (Ch 3)
μ	mean value or expected value of $\underset{\sim}{y}$
ξ	input variables
ϱ	correlation matrix for $\underset{\sim}{\beta}$
σ_{max}	maximum von Mises'stress (Ch. 2)
σ^2	variance of response quantity (Ch. 3)
θ	physical constant (Ch. 3)
	weighting factor (Ch. 2)
ψ	penalty function
χ	space containing candidate points (Ch. 3)
χ_i, $\underset{\sim}{\chi}$	candidate points
ω	eigenfrequency

CHAPTER 1: INTRODUCTION

1.1 Structural optimization and experimental design

The research field of structural optimization is concerned with the design of structures which can meet requirements better, e.g. with respect to the resistance of loads applied to them, or with respect to their performance. Designing consists of an iterative procedure of analysis and synthesis. By means of analysis the behaviour of the structure is evaluated, while synthesis is used to modify the structure in a way that it probably meets stated demands better. It is assumed that during the optimization process the behaviour and the performance can be described uniquely by a set of n design variables x_1, x_2, ..., x_n. The criterion by which the structure is judged to be better or worse than another one, is described by the so-called objective function and of course also depends on the design variables. In Chapter 2 we will discuss this in detail.

Modern research on numerical structural optimization was started during and after World War II in the aerospace industry due to the need for light-weight aircraft. During the last few decades optimization research has become important in almost any field of structural engineering. As reasons for this development one can see the diminishing energy and material resources, environmental problems concerning air pollution and noise, and last but not least the rapid development of computing facilities which have become a common tool of almost any engineer.

For the analytical task in a structural optimization problem mostly the finite element method (FEM) is used for two good reasons. The first reason is its nice and flexible modelling facilities, by which very different structures can be modelled in essentially the same way. Secondly there is the accuracy of a FEM analysis which can easily be controlled by means of the used element grid. But FEM analyses of actual engineering structures are often very time consuming, yielding a serious drawback for application of the method during structural optimization because of the iterative character of the optimization process.

In structural optimization problems the structure itself is subject of modifications, which is the reason why these problems are often essentially non-linear, and a closed-form solution for the optimum design seldom can be found. As already stated, optimization problems are mostly solved iteratively. In the next section we will briefly describe the process.

Experimental design, in the literature also referred to as "Design and analysis of experiments", or "Planning of experiments", provides methods for both the formulation of measurement programs and

the analysis of the measurement results. We emphasize that by "design" here is meant the design of a measurement program. In the scientific research cycle the following stages appear:

- theory and model building
- prediction
- experiment
- confrontation of prediction and experiment.

The last stage may be followed by model modification or refinement and the cycle is repeated until a satisfactory model is obtained. In many cases experimentation is time and cost consuming and the need for well organized experiments is clear. The experimental design theory provides tools for this organization.

In conjunction with optimization we are interested in factorial experiments which are designed to study some relationship, for instance:

$$y = f(x_1, x_2, \ldots, x_n)$$

showing a response variable y as a function of n independent variables $x_1, x_2, \ldots x_n$. In the experimental design theory the variables x_i are called "factors", hence the name "factorial experiments".

In the design and analysis of experiments the following questions have to be resolved:

- which factors play a role?
- should the factors be used in their original form, or should they be transformed or coded first?
- which is the range of interest for these factors?
- which form of the function f may be suitable to describe the searched relationship?
- in order to run the experiments, discrete values have to be chosen for the factors; how many discrete values, so-called levels, for each factor are needed?
- making a choice for a certain level of each factor represents a discrete design point in the space spanned by the factors; how many and which design points, in other words which measurement program, should be chosen?
- how can the relationship be estimated, and
- how can it be tested?

A very valuable aspect of the experimental design theory is that it guides experimentation in a structured way. When the founder of the experimental design theory, R.A. Fisher, in 1919 was appointed as a statistician at the Rothamsted Agriculture Research Station, he found data of experimentations over a period of 70 years. It proved that no confidential conclusions could be drawn from all these data; for Fisher it was the motive to develop the experimental design theory, the basis for modern experimentation. Answering the above questions is not possible all at once and it is certainly unwise trying to do so. Those questions constitute a circular problem: for instance to define the measurement program, the type of relationship has to be known, but the relationship can only be estimated after a

measurement program has been defined and the experimentation done. The solution to this problem is to start with a preliminary measurement program of moderate extent. Analysis of the data emerged from this program will provide better answers to the mentioned questions and a more adequate measurement program can be defined. Box et al. (1978) give a 25% rule, which states that at the outset of an experimental investigation, not more than 25% of the experimental effort (budget) should be invested in a first measurement program. The whole procedure is repeated until finally all questions have been resolved in a satisfactory way.

In Chapter 3 we will discuss parts of the experimental design theory in more detail.

1.2 Objective of the present research

A wealth of literature exists in the field of structural optimization based on FEM-analyses and also in the field of experimental design. In both research fields computer programs are available, aiming to release the engineer from tedious calculations, or without which it is at all impossible to obtain solutions. However, about the combination of these two research fields in the literature but few examples can be found (Schoofs (1984)). A literature search with on the one hand the key word "finite element" and on the other one of the following key words

- experimental design
- factorial design
- factorial analysis
- statistical analysis
- regression analysis
- parameter estimation
- parameter study
- surface fitting

yielded as a cross-section 13 articles. Only Pichumani (1975) and, to a less extent, Krishnamurthy et al. (1979) devoted explicit attention to the use of experimental design in combination with finite element analysis. In the other articles the use of experimental design and/or regression analysis is just mentioned, without consideration of procedural aspects.
Vanderplaats (1984) describes in a section "Formal approximations" some curve fitting techniques in optimization problems, which are similar to the procedures we will propose. We agree with him that the approach is competitive for problems with up to ten design variables and where computational cost is high. In contradiction with him we believe that the approach can also be effective in many structural optimization problems, due to application of experimental design procedures, which he does not mention.

Altogether, the use of experimental design in optimization appears to be a rather little elaborated research field.

The first motive for the author to consider structural optimization in combination with experimental design emerged from the master's degree study of Aerts (1979). This study describes

procedures using experimental design to "condense" a number of finite element analyses of a pin and hole joint to a handsome analysis model formed by a set of regression polynomials. Regarding structural optimization based on finite element calculations, the use of such regression models may have certain advantages. Due to the iterative character of the optimization procedure, a lot of finite element calculations are merely intermediate steps in the iteration process and are wasted when the optimization run is finished. Our idea is that, if a great number of finite element analyses has to be made, these analyses can also be used to derive an approximating regression model. In this way the results of every finite element calculation stay valuable.

In the present research we will consider experimental designs in which the experiments are of numerical nature, by applying finite element analyses. We will investigate the derivation of regression models and discuss the combination of experimental design and structural optimization, because we find these methods have much in common:

- the set of design variables in structural optimization plays the same role as the factors in experimental designs.
- in both methods the elementary process steps, that is in experimental design the measurements and in optimization the finite element analyses, are cost and time consuming and therefore their number has to be minimized.
- a finite element analyses of a structure can be considered as a numerical experiment.
- computer programs for structural optimization using the design variable concept are readily suited to do the "experiments" (finite element analyses) as indicated by the experimental design.
- a regression model can serve as a fast analysis module in a computer program for optimization.
- such a fast analysis module, allowes us to perform a large number of runs of the optimization program, every time emerging from a different starting point in the design space in order to find the global optimum for the structure; it is also possible to investigate outcomes of the optimization procedure using a number of different objective functions without being obliged to repeat a lot of finite element calculations.
- by applying a regression model an approximating optimum design can be found; this design can be used as initial design in an optimization based on direct FEM-analyses, thus resulting in a more accurate global optimum design.

In this thesis we will give procedural guidelines for the use of experimental designs in model building and structural optimization problems. We will develop special facilities in computer programs concerning both research fields and we will make an integrated use of these programs.

As one of the applications of these programs we designed a major-third carillon bell (Maas (1985)), by solving shape optimization problems for the bell geometry. Encouraged by this

success we decided to develop regression models of the eigenfrequencies for a wide class of different bell geometries. Achieving such a mathematical model for bells we consider a secondary, but nevertheless nice, goal of the present research.

CHAPTER 2: STRATEGIES IN STRUCTURAL OPTIMIZATION

2.1 Problem formulation

Numerical optimization of a mechanical structure requires the availability of a mathematical model of the structure. Such a model is characterized by a finite number of parameters. These model parameters may be of widely divergent types, such as parameters to specify physical properties, geometrical parameters and topological parameters to specify connections between structural components. The describing equations of the mathematical model can be derived in different ways, for instance, using a finite element method. For a linear, elastic, dynamically loaded structure this results in:

$$\underset{\sim}{M}\ddot{\underset{\sim}{u}} + \underset{\sim}{B}\dot{\underset{\sim}{u}} + \underset{\sim}{K}\underset{\sim}{u} = \underset{\sim}{p} \qquad (2.1.1)$$

and an appropriate set of initial conditions. Here M is the mass matrix, B the damping matrix, K the stiffness matrix, $\underset{\sim}{u}$ the column of unknown nodal displacements and $\underset{\sim}{p}$ the column of external, prescribed nodal loads. In the problems considered in this thesis often one or more of the parameters M, B and $\underset{\sim}{p}$ will be zero.

In a straightforward design all model parameters are chosen a priori, for instance based on prior investigations or on experience. In structural optimization, however, all or part of these parameters may depend on a finite number of a priori unknown design variables which have to be determined during the optimization process. Examples of design variables are the radius of a fillet, the cross-sectional area of a beam, the number of stiffeners on a panel and the thickness of a plate. In this thesis only scalar design variables are taken into account. They are denoted by $x_1, x_2, \ldots, x_n$ $(n \geq 1)$ and are considered the components of a column $\underset{\sim}{x}$:

$$\underset{\sim}{x}^T = [x_1 \; x_2 \; \ldots \; x_n] \qquad (2.1.2)$$

The design variables span an n-dimensional space, called the "design space" D. It should be emphasized that a design is represented by a point $\underset{\sim}{x}$ in D, even if the design is patently absurd (e.g. negative areas) or inadequate.

A design variable x_i may be discrete or continuous. In the second case, which occurs for instance if x_i is the radius of a fillet, x_i may be any real number in a given interval. The first case occurs, for instance, if x_i is the distance between equally spaced stiffeners on a given panel. A discrete design variable may be of type integer as will be the case if x_i is the number of stiffeners on a panel. Often discrete variables are treated as continuous ones during the optimization process. Then the final calculated value is

rounded off to the nearest discrete value a posteriori. In this thesis only continuous design variables are taken into account.

In general, design variables are subject to contraints. Here two types of constraints are distinguished:

Design_constraints

Design constraints are imposed on design variables for reasons of functionality, production, aesthetics etc. Examples are maximum track of a motor-car, minimum diameter of a hole, maximum and minimum slopes of a roof, etc. A very important characteristic of these constraints is that they can readily be evaluated without the need to solve the mathematical model of the structure.

Design constraints are explicit and/or simple, usually linear implicit relations in $\underset{\sim}{x}$. They can be represented by a set of md (md $\geq$ 0) inequalities

$$dci_j(\underset{\sim}{x}) \leq 0 \quad , \quad j = 1, 2, \ldots, md \tag{2.1.3}$$

and a set of ld (ld $\geq$ 0) equalities

$$dce_j(\underset{\sim}{x}) = 0 \quad , \quad j = 1, 2, \ldots, ld \tag{2.1.4}$$

The equality constraints (2.1.4) can be used to reduce the number of design variables, such that (2.1.4) is satisfied identically for the reduced set of variables. However, this is not always recommendable from the viewpoint of a general problem definition.

The constraints (2.1.3) and (2.1.4) define a subspace of the design space D, which will be called the "d-constrained design space" D_d. A design $\underset{\sim}{x} \in D_d$ is called a "d-constrained design" and it is assumed that such a design can be realized physically. The space D_d is of great importance in optimization algorithms because it is a proper subset of D: the design constraints reduce the set of points $\underset{\sim}{x}$ that might be of interest.

Behaviour constraints

Behaviour constraints are derived from requirements on the performance or the behaviour of the structure. Typical examples are limitations on maximum stresses and displacements, on eigenfrequencies and vibration modes etc. In general these constraints are nonlinear even for linear structures. Elaboration of the behaviour constraints requires the solution of the mathematical model of the structure. This is a characteristic, very important difference between behaviour and design constraints. Mathematically behaviour constraints are represented by a set of mb (mb $\geq$ 0) inequalities

$$bci_j(\underset{\sim}{x}) \leq 0 \quad , \quad j = 1, 2, \ldots, mb \tag{2.1.5}$$

and a set of lb (lb $\geq$ 0) equalities

$$bce_j(\underset{\sim}{x}) = 0 \quad , \quad j = 1, 2, \ldots, lb \tag{2.1.6}$$

Analogous to (2.1.4) the equalities (2.1.6) could be used to reduce the number of design variables. However, often the functions in (2.1.6) are complicated and implicit, and elimination of design variables may be practically impossible. Sometimes the model equations (2.1.1) are also incorporated in (2.1.6). Here (2.1.1) and (2.1.6) will be considered as separate sets of equations.

The constraints (2.1.5) and (2.1.6) define a subspace D_b of the design space D. This subspace will be called the "b-constrained design space". The set of all designs $\underset{\sim}{x}$, that satisfy both the design constraints and the behaviour constraints represents the "feasible design space" D_f. A design $\underset{\sim}{x} \in D_f$ is called a "feasible design". Mathematically D_f is defined as the intersection of the d- and the b-constrained design spaces

$$D_f = D_d \wedge D_b \tag{2.1.7}$$

The statement "some designs are better than others" implies the existence of a measure to compare designs. In mathematical optimization this measure is the objective function $F: D_d \rightarrow R$, which is defined on the d-constrained design space D_d. In many optimization formulations the objective function F is required to be semi-positive definite, i.e.

$$F(\underset{\sim}{x}) \geq 0, \quad \forall\, \underset{\sim}{x} \in D_d \tag{2.1.8}$$

A design $\underset{\sim}{x}^* \in D_f$ is considered to be a better design than $\underset{\sim}{x} \in D_f$ if and only if $F(\underset{\sim}{x}^*) < F(\underset{\sim}{x})$.

It will be clear that the outcome of the optimization process heavily depends on the choice of the objective function. Therefore this choice should be made with the utmost care. In some cases an obvious objective function exists. For example, in shape optimization of a fillet to reduce the stress concentration factor, it is trivial to choose $F(\underset{\sim}{x})$ equal to the maximum von Mises stress σ_{max} in the fillet. For transport systems often $F(\underset{\sim}{x})$ is chosen as the weight of the system since many design aspects of those systems are closely related to the weight. In other cases the objective function will be chosen as a weighted sum of functions which can represent totally different aspects, like costs of production and of exploitation, safety and environmental aspects. In such situations choosing an appropriate objective function is not trivial. A related problem occurs in situations where one tries to bring "everything" into account by means of terms in the objective function. If there is but little consensus about the importance of a certain aspect, that aspect should not be incorporated. The only result would be an increased "noise level" of the objective function.

The type of objective function may influence the flow diagram of the optimization process. It is assumed that $F(\underset{\sim}{x})$ is a computable function for each $\underset{\sim}{x} \in D_d$. In some problems evaluation of $F(\underset{\sim}{x})$ may require a complete elaboration of the mathematical model, for instance in the earlier mentioned case of stress optimization of a

fillet. If, on the other hand, the structural weight is chosen as an objective function $F(\underset{\sim}{x})$, evaluation of $F(\underset{\sim}{x})$ can be done using just the design variables $\underset{\sim}{x}$ describing a proposed new design. If the weight proves to be increased with respect to the current weight, analysis of the design $\underset{\sim}{x}$ makes no sense. In such a situation the process should be stopped or another design $\underset{\sim}{x}$ should be proposed.

In general the behaviour constraints and the objective function are highly non-linear functions of the design variables and the optimization problem must be solved iteratively. Usually, structural optimization problems converge to a local optimal design. The proof that this design also is the global optimum design can seldom be delivered. Hence, it is common practice to restart the optimization process with another initial design, and repeat this process until confidence is gained that the global optimum solution has been approached closely enough. Mathematically the optimization process can now be stated as follows:

$$\text{find } \underset{\sim}{x}^* \in D_f \text{ such that } F(\underset{\sim}{x}^*) \leq F(\underset{\sim}{x}) \text{ for all } \underset{\sim}{x} \in D_f \qquad (2.1.9)$$

After the previous discussions it will be clear that every phrase of this statement is essential in the optimization process. In other words, each of the steps

- find a feasible design $\underset{\sim}{x} \in D_f$
- find a (local) minimum for $F(\underset{\sim}{x})$ and
- get confidence that the solution is the global optimum,

generally are sub-problems of the same level of complexity.

2.2 Finite element method and structural optimization

In computer programs for structural optimization often the finite element method is used for modelling and analysis of the structure. This offers the possibility to model a great variety of structures in a general, accurate and flexible way. The accuracy can be controlled easily by an appropriate modification of the finite element model. In optimization problems often slight improvements of a structure are of interest, so a realistic model and accurate analysis results are of great importance. However, these advantages must be paid for with some serious drawbacks. First, the computing time is large. Second, the implementation of the design variable concept in finite element packages is non-standard and difficult. Last but not least, the computation of gradients of the objective function and of the behaviour constraints with respect to design variables is not trivial and difficult to implement.

For complicated linear structures finite element analyses are straightforward but can be very time consuming. If the structural problem is nonlinear, the finite element analysis has an iterative character and computing times may become tremendous. Furthermore, due to the iterative character of the optimization process several finite element analyses are required to obtain an optimal solution. In general this will be a local optimum and it is therefore recommended

to carry out several optimization runs. Hence, in problems where the finite element method is used, the number of iteration steps in the optimization process should be kept as small as possible.

A major problem in structural optimization is the link between design variables and the finite element model. In several more or less dedicated structural optimization programs this link has been realized. In some general purpose finite element packages the design variable concept has been adopted only with the most simple types of variables, such as cross-sectional properties. The rapid development of computer graphics and solid modelling programs can become of great importance to overcome this shortcoming, because the input needed for an advanced solid modelling program is much more user oriented than a finite element model and the link between solid models and finite element models has already been established.

Implementation of the design variable concept in finite element packages will make computation of gradients much more easier. If such an implementation is not realized, gradients can only be computed using numerical differentiation, at the cost of an increased number of finite element analyses. However, if the design variables can be used within the finite element program, then computation of gradients can be carried out in a much more efficient way. This can be illustrated by some examples.

First, consider a linear, elastic, static problem described by

$$K\underset{\sim}{u} = \underset{\sim}{p} \tag{2.2.1}$$

Differentiation of (2.2.1) with respect to design variable x_i yields

$$K \frac{\partial \underset{\sim}{u}}{\partial x_i} = \frac{\partial \underset{\sim}{p}}{\partial x_i} - \frac{\partial K}{\partial x_i} \underset{\sim}{u} \tag{2.2.2}$$

The gradients on the right hand side can easily be calculated by a perturbation method. Assuming that the matrix K was decomposed during a previous analysis and that $\underset{\sim}{u}$ is known from that analysis, the calculation of $\partial \underset{\sim}{u}/\partial x_i$ is, in fact, straightforward.

Next a linear, elastic, undamped, dynamically loaded structure is considered. Then it can be shown that the gradients of eigenfrequencies are given by

$$\frac{\partial \omega_j}{\partial x_i} = \frac{\underset{\sim}{v}_j^T [\frac{\partial K}{\partial x_i} - \omega_j^2 \frac{\partial M}{\partial x_i}] \underset{\sim}{v}_j}{2\omega_j \underset{\sim}{v}_j^T M \underset{\sim}{v}_j} \tag{2.2.3}$$

where $\underset{\sim}{v}_j$ is the eigenvector coupled with the eigenfrequency ω_j. Again the gradients of K and M can be calculated easily. If the eigenfrequencies and eigenvectors are available from a previous analysis, then the evaluation of (2.2.3) is straightforward too.

2.3 The iterative optimization procedure

2.3.1 Specific versus general methods

The solution methods commonly used to obtain the optimum design may be divided into several categories. An important classification is the partition in specific and general methods.

Specific methods are used exclusively in structural optimization. They usually are categorized as optimality criterion methods (Morris (1982)). In these methods one tries to satisfy a well defined criterion and it is expected that by doing so some measure of the structure will become optimal. However, this measure is not used explicitly to control the optimization process. A typical example of an optimality criterion method is the well-known fully stressed design method. In this method one tries to match the stresses in the structural members to their allowable stresses (the criterion), with the implicit objective to minimize the weight of the structure. In the early stages of development of structural optimization specific methods enjoyed great popularity because they could solve special problems more efficiently than any general method. But the popularity of specific methods is decreasing because of their limitations.

General methods are those which are applicable and commonly are applied to optimization problems in several fields. They are based on linear and nonlinear mathematical programming methods, and efficient computer implementations have been developed. However, despite all powerful software it should be emphasized that good engineering intuition will remain of great importance in solving structural optimization problems.

In this chapter we will give an introduction to the general methods.

2.3.2 General optimization algorithm

Let $\underset{\sim}{x}_0 \in D_d$ be a given initial set of design variables. This will not always be a feasible design. Some algorithms exist of two stages: in the first stage a feasible design $\underset{\sim}{x} \in D_f$ is determined from $\underset{\sim}{x}_0 \in D_d$ and in the second the design is optimized. The algorithm is given by:

```
begin boolean CONV;
      {given: initial design x_0 ∈ D_d, convergence criteria, and
      max. number of iteration steps q_max}
      q: = 0
      x_q: = x_0;
      CONV: = false
      while not CONV and q < q_max do
      begin {analyse design x_q;
             evaluate behaviour constraints;
             evaluate objective function}
            {evaluate convergence criteria}
             if converged then CONV: = true
            {compute α_q, s_q}
            {compute x_q+1}
            q: = q+1
      end
end;
```

From the various steps in this algorithm only the optimization step will be considered now. In general the updating step can be formulated as:

$$\underset{\sim}{x}_{q+1} = \underset{\sim}{x}_q + \alpha_q \underset{\sim}{s}_q \qquad (2.3.1)$$

where q is the iteration number and $\underset{\sim}{s}_q$ a search direction column while α_q is a scalar stepsize factor. So the optimization step (2.3.1) consists of two parts:

- finding an appropriate search direction $\underset{\sim}{s}_q$. For this purpose the most effective optimization algorithms require partial derivatives of the objective function and of the constraints with respect to the design variables to be calculated.
- computing the scalar α_q^* such, that, moving in the direction $\underset{\sim}{s}_q$, the objective function is minimized.

A great number of mathematical programming techniques, linear and nonlinear, can be used for the optimization step. Some of these techniques are discussed in the sections 2.4, 2.5 and 2.6.

2.3.3 Partial derivatives and conjugate directions

Numerical optimization algorithms generally operate with approximations of the constraint functions and of the objective function. Such approximations can be formulated, using Taylor series expansions. In the sequel it is assumed that the functions are continuous and differentiable as often as necessary for all $\underset{\sim}{x} \in D_d$. The first and the second derivatives of the objective function F, i.e. the gradient and the Hessian matrix, are denoted by $\underset{\sim}{G}$ and H respectively:

$$\underset{\sim}{G} = \underset{\sim}{\nabla} F \tag{2.3.2}$$

$$H = \underset{\sim}{\nabla}(\underset{\sim}{G}^T) = \underset{\sim}{\nabla}(\underset{\sim}{\nabla} F)^T = H^T \tag{2.3.3}$$

where $\underset{\sim}{\nabla}$ is the gradient-operator, defined by:

$$\underset{\sim}{\nabla} = [\frac{\partial}{\partial x_1} \frac{\partial}{\partial x_2} \ \ldots \ \frac{\partial}{\partial x_n}]^T \tag{2.3.4}$$

In general it is not possible to give explicit relations for the gradient as a function of the design variables. An approximation for $\underset{\sim}{G}(\underset{\sim}{x})$ always can be determined by numerical differentiation, using the method of subsequent perturbations. This results in

$$G_i(\underset{\sim}{x}) \approx \frac{1}{\delta x_i} [F(\underset{\sim}{x} + \underset{\sim}{e}_i \, \delta x_i) - F(\underset{\sim}{x})] \tag{2.3.5}$$

where δx_i is a small perturbation and $\underset{\sim}{e}_i$ is the i^{th} unit column (i.e. component j of $\underset{\sim}{e}_i$ is equal to the Kronecker-delta δ_{ij}).

A second-order Taylor approximation $\hat{F}(x)$ of $F(\underset{\sim}{x})$ in the neighbourhood of $\underset{\sim}{x}_q$ now can be written as:

$$\hat{F}(\underset{\sim}{x}) \approx F_q + \underset{\sim}{G}_q^T(\underset{\sim}{x}-\underset{\sim}{x}_q) + \tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{x}_q)^T H_q(\underset{\sim}{x}-\underset{\sim}{x}_q) \tag{2.3.6}$$

where $F_q = F(\underset{\sim}{x}_q)$, $\underset{\sim}{G}_q = \underset{\sim}{G}(\underset{\sim}{x}_q)$ and $H_q = H(\underset{\sim}{x}_q)$

If in (2.3.6) $\underset{\sim}{x}_q = \underset{\sim}{x}^*$ is a real (local) minimum of $F(\underset{\sim}{x})$, then

$$\underset{\sim}{G}(\underset{\sim}{x}^*) = \underset{\sim}{0} \tag{2.3.7}$$

and $H(\underset{\sim}{x}^*)$ is positive definite, resulting in the quadratic form

$$F(\underset{\sim}{x}) = F(\underset{\sim}{x}^*) + \tfrac{1}{2} (\underset{\sim}{x}-\underset{\sim}{x}^*)^T H(\underset{\sim}{x}^*)(\underset{\sim}{x}-\underset{\sim}{x}^*) \tag{2.3.8}$$

Since higher order terms may be neglected in the neighbourhood of $\underset{\sim}{x}^*$, it may be deduced from (2.3.8) that quadratic forms are important for convergence considerations.

Using (2.3.6) optimization algorithms can be divided into:

- zero-order methods, using only function values
- first-order methods, using function values and gradients
- second-order methods, using function values, gradients and Hessian matrices.

An important class of optimization algorithms is based on the use of so-called conjugate directions. A set of n search direction columns $\underset{\sim}{s}_1, \underset{\sim}{s}_2, \ldots, \underset{\sim}{s}_n$ is said to be conjugate with respect to the n*n positive definite symmetrical matrix H if, for $\underset{\sim}{s}_i \neq \underset{\sim}{0}$ and $\underset{\sim}{s}_j \neq \underset{\sim}{0}$

$$\underset{\sim}{s}_i^T H \underset{\sim}{s}_j = 0 \quad \text{for } i \neq j \quad ; \quad i, j \in \{1, 2, \ldots, n\} \tag{2.3.9}$$

The use of conjugate search directions is based on the following lemma (Fox (1971)):

If a homogeneous quadratic function Q is minimized, using subsequent directions from a set of n conjugate directions, then:
- the exact minimum is reached at or before the nth iteration step.
- the choice of the initial solution does not matter.
- the sequence of search directions does not matter.

Although the functions to be minimized often are non-quadratic and iteration steps always show an inherent inaccuracy, conjugate search directions appear to be very useful in optimization methods.

2.4 Minimization of the objective function along a search direction

Starting from an initial design $\underset{\sim}{x}_0 \in D_d$ the optimal design $\underset{\sim}{x}$, being the solution of (2.1.9), is calculated iteratively. As stated before in (2.3.1), in iteration step q+1 ($q \geq 0$) a new estimate $\underset{\sim}{x}_{q+1} \in D_f$ is determined, using

$$\underset{\sim}{x}_{q+1} = \underset{\sim}{x}_q + \alpha_q \underset{\sim}{s}_q \tag{2.4.1}$$

For the moment it is assumed that the search direction $\underset{\sim}{s}_q$ is known and attention is focussed on the stepsize α_q. In some algorithms α_q is fixed a priori while in other algorithms α_q is determined from the requirement that the function $f_q = f_q(\alpha)$, defined by

$$f_q(\alpha) = F(\underset{\sim}{x}_q + \alpha \underset{\sim}{s}_q), \tag{2.4.2}$$

is minimized. Elaboration of this requirement yields that α_q is the solution for α of the (nonlinear) equation

$$\underset{\sim}{s}_q^T \underset{\sim}{G}(\underset{\sim}{x}_q + \alpha \underset{\sim}{s}_q) = 0, \tag{2.4.3}$$

However, this relation in general does not provide a practical starting point for the determination of the optimal stepsize α_q. The reasons are that many evaluations of the gradient may be necessary to solve (2.4.3) for α_q, that no explicit relations for the gradient as a function of the design variables are available and that each evaluation of the gradient will require considerable computational effort. Therefore, an approximation α_q of the optimal stepsize will be used. A practical method to determine such an approximation is based on the idea to fit the function $f_q = f_q(\alpha)$ with a polynomial $p_q = p_q(\alpha)$, for instance a quadratic one to start with:

$$f_q(\alpha) \approx p_q(\alpha) \quad ; \quad p_q(\alpha) = a_q + \alpha\, b_q + \alpha^2 c_q \tag{2.4.4}$$

The three coefficients a_q, b_q and c_q can be determined from the value of $f_q(\alpha) = F(\underset{\sim}{x}_q + \alpha\, \underset{\sim}{s}_q)$ for three values of α, often taken as $\alpha = 0$, $\alpha = \Delta_q$ and $\alpha = 2\Delta_q$ where Δ_q is a more or less arbitrary trial stepsize. As soon as a_q, b_q and c_q are known the stepsize α_q can readily be found from the requirement that $p_q = p_q(\alpha)$ is minimal for $\alpha = \alpha_q$. It is recommended to accept this stepsize only if $\alpha_q \in (0, 2\Delta_q)$.

If α_q is outside this interval the value of the trial stepsize Δ_q must be modified adequately and the calculation of α_q must be repeated. The quality of the approximation (2.4.4) can be tested by evaluating $f_q(\alpha_q)$ and comparing this value with $p_q(\alpha_q)$. If the approximation is unacceptable $f_q(\alpha_q)$, $f_q(0)$, $f_q(\Lambda_q)$ and $f_q(2\Lambda_q)$ can be used to approximate $f_q = f_q(\alpha)$ by a cubic polynomial. From this approximation again a value for the stepsize α_q can be determined.

Each of these methods to arrive at a value for the stepsize involves one or more evaluations of the objective function. It is stipulated here once more that, in structural optimization, such an evaluation generally will involve a complete finite element analysis of the structure. Therefore, the number of evaluations of the objective functions per iteration step must be held as small as possible. Finally it is mentioned that the calculations of both $\underset{\sim}{s}_q$ and α_q in constrained problems are complicated by the condition that the new estimate $\underset{\sim}{x}_{q+1} = \underset{\sim}{x}_q + \alpha_q \underset{\sim}{s}_q$ must be a feasible design, i.e. that $\underset{\sim}{x}_{q+1} \in D_f$.

In the next section some methods, especially for the calculation of the search direction, in unconstrained problems are discussed. These methods are important since many optimization problems can be formulated without constraints or may be considered to be unconstrained in certain stages of the optimization process. Furthermore, many methods for constrained problems are based on methods for unconstrained problems. Constrained problems are discussed in some detail in section 2.6.

2.5 Methods for unconstrained problems

As mentioned before, the commonly used methods for unconstrained optimization problems can be divided into zero-order, first-order and second-order methods. Usually these methods are not completely unconstrained. For instance, designs in the random search method and initial designs in other methods will be chosen using the design constraints.

2.5.1 Zero-order methods

In this subsection two zero-order methods are considered, being the random search method and Powell's method.

In the random search method a given number $m (m \geq 1)$ of designs $\underset{\sim}{x}_i \in D_d$, $i = 1, 2, \ldots, m$ is generated. The objective function F is evaluated for each of these designs and the design, associated with the lowest value of this function, is considered to be the best ("optimal") design. This method can easily be implemented and requires modest computer storage capacity. Furthermore, if m is large enough one may expect to find a reasonable approximation for the optimal design. However, the efficiency of this method is low since many function evaluations are necessary. Although the efficiency can be improved by means of some simple modifications this method is not appropriate in structural optimization, based on finite element analyses.

Powell's method (Powell, 1964) is an iterative method, starting from a given initial design $\underset{\sim}{x}_0$. Each iteration step q ($q \geq 1$) involves first the definition of n search directions $\underset{\sim}{o}_{q1}$, $\underset{\sim}{o}_{q2}$, ... $\underset{\sim}{o}_{qn}$ and second the subsequent line minimalization of the objective function, using these directions. This results in an approximation $\underset{\sim}{x}_q$ for the optimal design. The difference $\underset{\sim}{x}_q - \underset{\sim}{x}_{q-1}$ defines the q^{th} conjugate search direction $\underset{\sim}{s}_q$. In the next iteration step the search directions are modified. It is common practice to choose $\underset{\sim}{o}_{1i} = \underset{\sim}{e}_i$, i = 1, 2,...n in the first iteration step. In iteration step q+1 the direction $\underset{\sim}{o}_{q1}$ is rejected, the other search directions are renumbered $\underset{\sim}{o}_{(q+1)i} = \underset{\sim}{o}_{q(i+1)}$ for i = 1, 2, ...n-1 and the last search direction $\underset{\sim}{o}_{(q+1)n}$ is chosen equal to the most recent conjugate search direction $\underset{\sim}{s}_q$. This process is illustrated in Fig. 2.1.

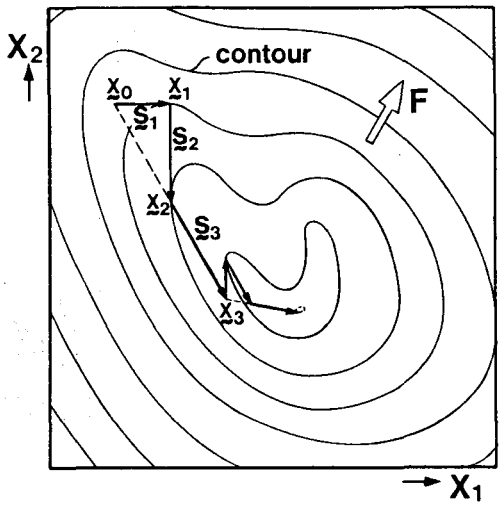


Fig. 2.1 Powell's zero-order method

Using Powell's method difficulties occur in two situations. First, if in a certain search direction no improvement can be made the method breaks down. Second, after some iterations, some of the search directions can become dependent. A simple remedy for these problems is to restart the whole process using the most recent solution as the initial design.

Powell's method is one of the most powerful and reliable zero-order methods, but still requires many function evaluations. To generate the conjugate directions n(n+1) searches are needed. If quadratic interpolation is used, within each search three function evaluations have to be done, giving a total number of $3n^2 + 3n$ evaluations. For non-quadratic functions this number may become as large as $5n^3$, which makes the method prohibitive for large numbers of design variables combined with a more than minor computational effort per evaluation.

2.5.2 First-order methods.

In this subsection two first-order methods are considered, namely the steepest descent method and the conjugate gradient method.

The steepest descent method is based on the idea that the best search direction is the direction of the greatest rate of decrease of $F(\underset{\sim}{x})$, which means:

$$\underset{\sim}{s}_q = - \nabla F(\underset{\sim}{x}_q) = - \underset{\sim}{G}_q \qquad (2.5.1)$$

Although the method may perform well in some problems, in other problems convergence may become surprisingly slow, because the minimization process may result in an n-dimensional zig-zag of small successive moves (see Fig. 2.2).

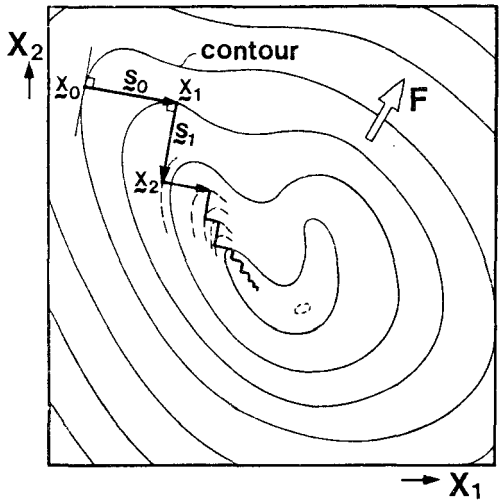


Fig. 2.2 Steepest decent method

Fletcher and Reeves (1964) developed a modification of the steepest descent method by which the convergence difficulties are greatly reduced. In this so-called conjugate gradient method the search direction is given by (see Fig. 2.3):

$$\underset{\sim}{s}_q = - \underset{\sim}{G}_q + \frac{\underset{\sim}{G}_q^T \underset{\sim}{G}_q}{\underset{\sim}{G}_{q-1}^T \underset{\sim}{G}_{q-1}} \underset{\sim}{s}_{q-1} \qquad (2.5.2)$$

The first term on the right-hand side represents the search direction of the steepest descent method. The second term makes $\underset{\sim}{s}_q$ a linear combination of $\underset{\sim}{s}_0, \underset{\sim}{s}_1, \ldots, \underset{\sim}{s}_{q-1}$. Fletcher et al. (1964) showed that the method generates a set of conjugate search directions. Compared to the steepest descent method, the convergence rate of the conjugate gradient method has been strongly improved, and yet the method remains rather simple. Computer implementation is quite simple and the algorithm requires only little computer storage. For a quadratic function the method converges in n steps or fewer, but for nonquadratic functions considerably more steps may be required and

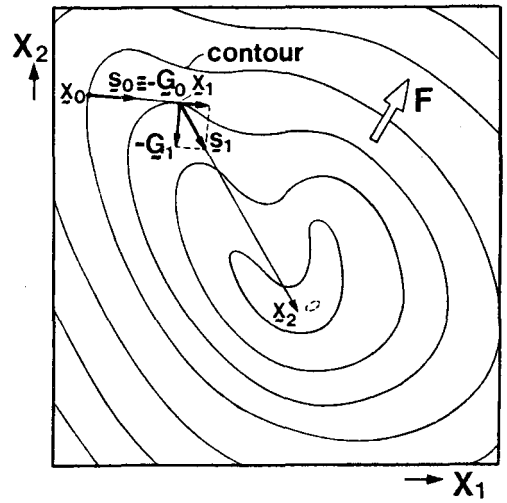


Fig. 2.3 Conjugate gradient method

ill-conditioning of the search directions may occur. It is recommended then to restart the process with $\underset{\sim}{s} = -\underset{\sim}{G}$. Fox (1971) gives the rule to restart if the process is not converged after n iteration steps.

2.5.3 Second-order methods; variable metric methods

Methods based on the second-order approximation (2.3.6) are called second-order or, more specific, Newton methods. In these methods the Hessian matrix H is explicitly used. Here the Newton-Raphson method is considered as an introduction to the quasi-Newton DFP-method.

According to (2.3.6) the approximation $\hat{F}(\underset{\sim}{x})$ for $\Gamma(\underset{\sim}{x})$ in the neighbourhood of $\underset{\sim}{x}_q$ is given by:

$$\hat{F}(\underset{\sim}{x}) = F_q + \underset{\sim}{G}_q^T (\underset{\sim}{x}-\underset{\sim}{x}_q) + \frac{1}{2} (\underset{\sim}{x}-\underset{\sim}{x}_q)^T H_q (\underset{\sim}{x}-\underset{\sim}{x}_q) \qquad (2.5.3)$$

This quadratic form has a minimum if H_q is positive definite. This minimum occurs for $\underset{\sim}{x} = \underset{\sim}{x}_{q+1}$, where $\underset{\sim}{x}_{q+1}$ satisfies:

$$\underset{\sim}{G}_q + H_q (\underset{\sim}{x}_{q+1} - \underset{\sim}{x}_q) = \underset{\sim}{0} \qquad (2.5.4)$$

Hence, in the Newton-Raphson method the new approximation $\underset{\sim}{x}_{q+1}$ for the optimal design is determined from:

$$\underset{\sim}{x}_{q+1} = \underset{\sim}{x}_q - H_q^{-1} \underset{\sim}{G}_q \qquad (2.5.5)$$

In structural optimization problems, the Newton-Raphson method has some serious drawbacks, for instance computation of the Hessian matrix is difficult, especially in large problems. In general, analytical relations are not available and numerical computation is

too time consuming. Furthermore, solution of $\underset{\sim}{x}_{q+1}$ from (2.5.4) may be prohibitive in large problems.

The DFP-method developed by Davidon (1959) and improved by Fletcher and Powell (1963), is a variable metric method. This method is very attractive, because it has some of the properties of a Newton-method without the need to compute the Hessian matrix. The method only uses gradients, so it is in fact a first order method. Fox (1971) showed that it can also be considered a conjugate direction method. To start the iteration process of the DFP-method an initial design $\underset{\sim}{x}_0$ and a positive definite matrix H, for instance the identity matrix, must be given. Then the process can be formulated as follows:

```
begin boolean CONV;
      G0: = G(x0);   x: = x0
      CONV: = false;
      while not CONV do
      begin s: = - HG0;
            {compute stepsize α which minimized F(x + αs)}
            x: = x + α s;
            Gn: = G(x);   y: = Gn - G0;   z: = Hy;     G0: = Gn;
            H: = H + α s sT(sTy)-1 - z zT(zTy)-1
      if converged then CONV: = true;
      end

end;
```

Although this method requires more computer storage than the conjugate gradient method it can be implemented quite straightforward. Due to the fact that the matrix H can represent the history of the iteration process much better than a single search direction, the method asks much less attention on breakdowns than the conjugate gradient method, resulting in a lower number of necessary restarts.

2.6 Methods for constrained problems

Methods for constrained problems can be based on those for unconstrained problems. The penalty-function methods, treated below, are typical examples. In this section also the feasible directions method and a sequential linear programming method are considered.

2.6.1 Penalty-function (PF) methods

In these methods the optimization problem with constraints is transformed to a problem without constraints and then solved using a method for unconstrained problems. Here only problems with inequality constraints are considered. Equality constraints can be handled as well, but then these methods are less effective and more complicated. The constraints (2.1.3) and (2.1.5) are taken into account by means

of a penalty term, which is added to the objective function F. This yields the penalty function ψ, generally defined as:

$$\psi(\underset{\sim}{x}, r) = F(\underset{\sim}{x}) + r \sum_{j=1}^{m} G(g_j(\underset{\sim}{x})) \qquad (2.6.1)$$

where g_j, $j = 1, \ldots, m$, represent the inequality constraints (2.1.3) and (2.1.5). The function G is chosen such that subsequent minimizations of ψ for a sequence of values for r, converge to the solution of the constrained problem. The factor r provides a weighting between the objective function value and the penalty term. There are several possibilities to choose G, each resulting in a particular PF-method. Here only the exterior PF-method is considered.

In this method the penalty function ψ is defined by:

$$\psi(\underset{\sim}{x}, r) = F + r \sum_{j=1}^{m} \langle g_j \rangle^{\gamma} \qquad (2.6.2)$$

$$\langle g_j \rangle = \begin{cases} g_j & \text{if } g_j > 0 \\ 0 & \text{if } g_j \leq 0 \end{cases}$$

where γ and r are positive numbers. Usually $\gamma = 2$ is chosen. From (2.6.2) it is clear that, with respect to F, ψ is raised outside the feasible region of the problem (Fig. 2.4). The penalty term increases rapidly with increasing violation of the constraints.

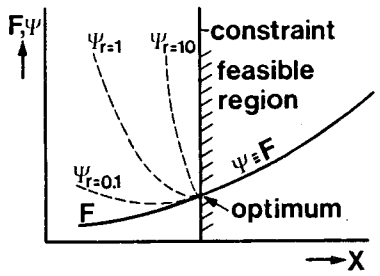


Fig. 2.4 Exterior PF-method

The process is started with a relatively small value for r, say r=1, and r is raised in the subsequent steps by a factor $C > 1$. This is done to keep ψ approximately quadratic in the neighbourhood of the current solution, thus making minimization of ψ easier. With a relatively large final value for r, say r = 1000, the minimum of ψ will be a good approximation of the minimum of F.

The initial design may be an infeasible design. Usually the solutions approach the feasible region from outside. However, the process cannot be stopped until a sufficiently converged solution is obtained.

PF-methods have become very popular and are implemented in several structural optimization packages as ACCESS (Schmit et al. (1975, 1976, 1979)) and more recently in NUW SUMT (Miura et al. (1979)). In the last package improved PF-methods using Lagrange multipliers (Imai (1978)) are implemented.

2.6.2 The feasible directions method

Whereas in PF-methods the constraints are taken into account indirectly, in the feasible directions method the constraints are explicitly used to guide the solution process (Avriel (1976), Gill et al. (1974), Murray (1976)). Again it is assumed that all equality constraints are eliminated. Each iteration step q($q \geq 1$) involves the determination of a search direction $\underset{\sim}{s}_q$, which has to fulfil two demands (see Fig. 2.5):

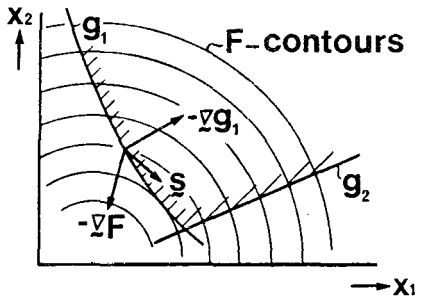


Fig. 2.5 Feasible and usable $\underset{\sim}{s}$

1. $\underset{\sim}{s}_q$ must be feasible. If none of the constraints $g_j < 0$, j = 1, 2, ... m is active at the current solution $\underset{\sim}{a}_q$ then any direction is allowed. Otherwise $\underset{\sim}{s}_q$ should point into the feasible region. Hence, each active constraint imposes a condition on $\underset{\sim}{s}_q$ and $\underset{\sim}{s}_q$ is called feasible if

$$\underset{\sim}{s}_q^T \nabla g_j < 0 \text{ for each active constraint } g_j \qquad (2.6.3)$$

2. $\underset{\sim}{s}_q$ must be usable in the sense that the objective function should decrease in the direction $\underset{\sim}{s}_q$. This results in the condition

$$\underset{\sim}{s}_q^T \underset{\sim}{G}_q < 0 \quad ; \quad \underset{\sim}{G}_q = \underset{\sim}{G}(\underset{\sim}{x}_q) \qquad (2.6.4)$$

Zoutendijk (1960) formulated the following linear programming problem to determine $\underset{\sim}{s}_q$:

maximize β ($\beta > 0$), such that

$\underset{\sim}{s}_q^T \nabla g_j + \theta_j \beta \leq 0$ for each active constraint g_j,

$\underset{\sim}{s}_q^T \underset{\sim}{G}_q + \beta \leq 0$ and

$||\underset{\sim}{s}_q||$ is limited

Here, $||\underset{\sim}{s}||$ represents a norm of the column $\underset{\sim}{s}$, for instance the maximum norm.
The coefficients θ_j are weighting factors between the individual constraints on the one hand and the active constraints and the objective function on the other. To solve the stated linear programming problem the well-known Simplex algorithm can be used. Once $\underset{\sim}{s}_q$ is found, an appropriate value α_q for the stepsize must be determined. Here two situations can occur (see Fig. 2.6). In the first case, the minimum of F in $\underset{\sim}{s}_q$-direction is found in the interior of the feasible region. Then α_q is found, using a one-dimensional search as described in Section 2.4. In the second case, the minimum of F in $\underset{\sim}{s}_q$-direction is outside the feasible region resulting in violation of one or more constraints. Then, starting in $\underset{\sim}{x}_q$ and moving in positive direction (i.e. $\alpha>0$) along the line $\underset{\sim}{x} = \underset{\sim}{x}_q + \alpha \underset{\sim}{s}_q$ it is

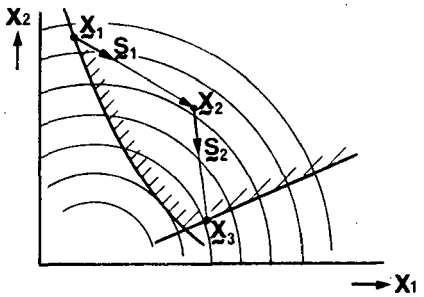


Fig. 2.6 Unconstrained ($\underset{\sim}{x}_2$) and constrained solution ($\underset{\sim}{x}_3$)

determined for which value of α the first constraint is violated. That value of α is taken as the stepsize α_q.

The feasible directions method turns out to be very well applicable for structural optimization. It is implemented, for instance, in the package CONMIN (Vanderplaats (1973)). The recently developed package ADS-1 (Vanderplaats et al. (1983)) offers a menu of optimization algorithms, some of them based on new feasible directions methods.

2.6.3 Sequential linear programming (SLP) method

Another aproach to solve nonlinear constrained optimization problems is solving a sequence of linear programming (LP) problems. Both the objective function and the constraints are linearized about the current solution $\underset{\sim}{x}_q$. The minimum of the objective function

generally is found in a vertex of the approximated feasible region. To formulate the linearized problem, first-order Taylor series expansions of the describing relations of the original problem are used. The already mentioned Simplex algorithm (Dantzig (1963)) can be used to solve the LP-problems.

In this form the SLP method will not work if the original problem is highly nonlinear, because then linearization over the whole feasible region will result in a very bad approximation. This problem can be tackled by introducing so-called move limits:

$$|\underset{\sim}{x}-\underset{\sim}{x}_q|_i \leq \Delta x_i \quad , \quad i = 1, 2, \ldots, n \qquad (2.6.5)$$

where Δx_i is a maximum stepsize for the individual design variable. Now the feasible region for the LP-problem is limited to the region described by (2.6.5). This region may be reduced further by one or more of the linearized constraints of the original problem, see Fig. 2.7.

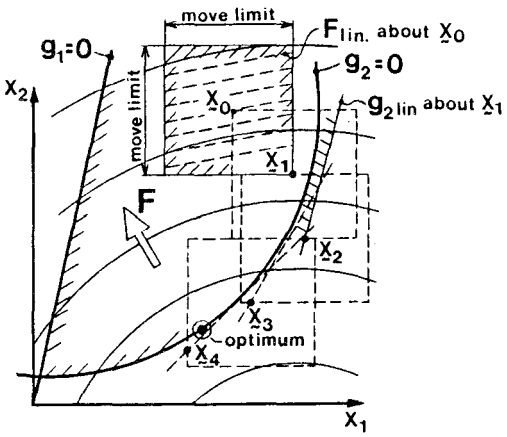


Fig. 2.7 SLP-method with move limits

The optimization process is started with relatively large move limits. As the solutions approach a (local) minimum they usually tend to oscillate between some solutions. At this point the move limits are reduced and the solutions usually continue to converge.

The SLP-method differs from other algorithms in the sense that it delivers a search direction and a stepsize at the same time (incorporated in the LP-solution). In other methods generally the search direction is determined first, followed by the stepsize. The SLP-method with use of move limits is applied rather successfully to a variety of structural optimization problems. It has been implemented in the structural optimization programs OPTIMA (Widdershoven (1980)) and DYNOPT (Van Asperen (1984)). The latter program has been used for the applications in Chapter 6.

2.7 Comparison of methods

In literature several investigations on the applicability and efficiency of optimization algorithms are given. Ragsdell (1984) gives an excellent review of several important and comprehensive investigations. Carpenter and Smith (1977) investigated the computational efficiency of several methods for structural optimization. They conclude that if, per iteration step, gradient values ask less computing time than function values, the SLP method is much more efficient than, for instance, the PF-methods and the feasible directions method. Our experience is that if methods are used such as described in section 2.3, computations of gradients are indeed more efficient than the computation of a function value.

Sandgren (1977) investigated a great variety of optimization algorithms, but did not take into account the implications of time consuming FEM-analysis in structural optimization. He collected 30 optimization problems of different levels of complexity. In addition he collected 35 codes of leading nonlinear programming methods. After a preliminary screening 23 problems and 24 codes were left. Sandgren forced all codes to operate with numerically computed gradients and carried out the tests on the same computer in order to get accurate measures for comparison. He subsequently used all codes and tried to solve as much test problems as possible with each code, recorded the required computing time and plotted the number of solved problems versus the required computing time, normalized as the fraction of the average time required for all codes. Fig. 2.8 shows a small selection from Sandgren's results. Table 2.1 indicates the used methods for the constrained problems and the algorithms which are used in the unconstrained sub-problems.

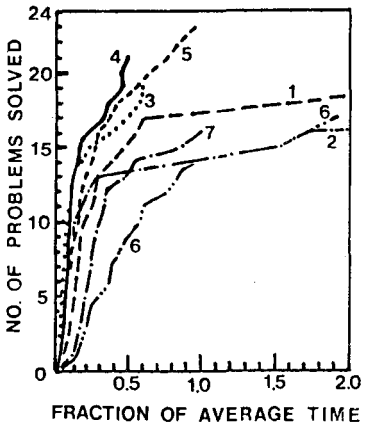


Fig. 2.8 Algorithm utility

	Method	
No.	Constrained	Unconstrained
1	GRG	DFP
2	SLP	-
3	GRG	BFS
4	GRG	CG
5	GRG	CG
6	Ext.PF	DFP
7	Ext.PF	DFP

Table 2.1 Algorithm types

The abbreviations in Table 2.1 have the following meaning:

GRG = general reduced gradient, a feasible directions method
SLP = sequential linear programming method
Ext. PF = exterior penalty function method
DFP = variable metric method (Davidon, Fletcher and Powell)
BFS = variable metric method (Broydon, Fletcher and Shanno)
CG = conjugate gradient method (Fletcher and Reeves).

Interpretation of Fig. 2.8 learns that:

- the methods 3, 4 and 5 are both efficient (a steep curve) and robust (a large number of solved problems)
- the methods 1 and 2 are efficient but less robust
- the methods 6 and 7 are less efficient and less robust than the methods 3, 4 and 5.

Another investigation is carried out by Schittkowski (1980). Significant differences with Sandgren's study are the very large number of test problems (180) and the incorporation of the most modern algorithms in the tests. It turned out that the new, so-called sequential quadratic programming methods (SQP) are very promising.

CHAPTER 3: EXPERIMENTAL DESIGN THEORY

3.1 Introduction; survey of the theory

Advanced scientific and technological research requires comprehensive and expensive experiments and the need for careful planning of the experiments is quite clear. On the one hand it is required to minimize the number of experiments, but on the other it is desired to gather as much information as possible about the relevant aspects of the system under consideration.

The experimental design theory (EDT) consists of two main parts. The first part, discussed in the sections 3.2, 3.3 and 3.4, concerns the planning of experiments and ends up with a list of experiments to be carried out. This list is called the experimental design, abbreviated to ED. In the second part the experimental results are analysed and fitted to some mathematical relationship. This is discussed in section 3.5.

In this thesis we are primarily interested in the use of (a special version of) EDT for the planning of numerical, FEM analyses. Such analyses can be regarded as numerical experiments: the investigator specifies a number of input parameters and as a result a number of response quantities emerge from the analysis program. The computations can be described as deterministic processes. Repeated computations using the same set of input parameters invariably result in exactly the same response quantity values. In general physical experiments show a stochastic character. Even if it were possible in physical experiments to adjust the input parameters to exactly the same values, repeated experiments will always show random variations of the response quantities. EDT has been developed for such stochastic processes. Use of the theory for the planning of deterministic numerical experiments is hardly mentioned in literature. We conform as much as possible to the common theory for stochastic processes, which is discussed first in the next four sections. A separate section, section 3.6, is devoted to the consequencies of the use of EDT in deterministic processes. It turns out that much of the common theory may still be used.

In this chapter special attention is given to the use of gradients in EDT. The use of gradients proves to be very advantageous in numerical experiments. The last section of this chapter gives information about a program for computer aided design and analysis of experiments. In the remainder of this section some definitions are given and the problem of experimental design is formulated.

We presume that there exists a true physical relationship between a response quantity u and input quantities $\underset{\sim}{\xi}$ via physical constants $\underset{\sim}{\theta}$:

$$u = \kappa(\underset{\sim}{\xi}, \underset{\sim}{\theta}) \qquad (3.1.1)$$

If the true physical relationship is known, then we only need to estimate the physical constants θ. Often relation (3.1.1) is unknown or too complex and (3.1.1) is approximated locally by some interpolating function $g(\underset{\sim}{x}, \underset{\sim}{\beta})$:

$$u \approx g\ (\underset{\sim}{x}, \underset{\sim}{\beta}) \tag{3.1.2}$$

where $\underset{\sim}{x}$ is a set of control variables, obtained by coding or standardizing the input quantities ξ and $\underset{\sim}{\beta}$ are the unknown coefficients in the interpolating function. The name "control variable" will be used throughout this chapter because it expresses very well our objectives: we want to control the behaviour of a system by means of manipulating the control variables. The concept of control variables is closely related to the concept of design variables in Chapter 2.

The control variables are assumed to be mutually independent. They are denoted by $x_1, x_2, \ldots, x_n$ ($n \geq 1$) and are considered the components of a column $\underset{\sim}{x}$

$$\underset{\sim}{x} = [x_1\ x_2\ \ldots\ x_n]^T \tag{3.1.3}$$

In this thesis only control variables will be considered which are continuous quantities. In EDT, however, it is common practice that each control variable is adjusted to a finite number of discrete values, called levels. A column $\underset{\sim}{x}$ containing one level for each of the control variables is called a treatment or experimental run. The set of all possible treatments, for which observations can be made, is denoted by χ and the number of elements of χ, i.e. the number of possible treatments, is denoted by r. In a so-called complete factorial experiment all treatments $\underset{\sim}{x} \in \chi$ are considered whereas in a fractional factorial experiment only treatments in a subset of χ are used.

If observations are repeated for a treatment $\underset{\sim}{x}$ then the observed results will vary about μ due to the stochastic nature of the experiments. We shall assume a normal distribution for the observed result, which we will call $\underline{y}$. The underscore denotes that y is a stochastic variable. The expected value of $\underline{y}$ equals μ, that is, $E(\underline{y}) = \mu$.

Because of the stochastic nature of the experiments and because of the approximation of the true relationship by an interpolating function, two types of errors will occur:

1. random error: $\underline{e} = \underline{y} - E(\underline{y})$,
2. systematic error, or bias: $E(\underline{y}) - g(\underset{\sim}{x}, \underset{\sim}{\beta})$.

Thus, the following relation exists:

$$\underline{y} = g(\underset{\sim}{x}, \underset{\sim}{\beta}) + \underline{e} + \text{bias} \tag{3.1.4}$$

We shall deal with bias in Section 3.3.6, where an experimental design technique is presented with which designs can be constructed that account for bias. For the present it is assumed that the

interpolating function exactly matches the true physical relationship and only random errors exist:

$$\begin{aligned} \underset{\sim}{y} &= g(\underset{\sim}{x}, \underset{\sim}{\beta}) + \underset{\sim}{e} \\ &= \underset{\sim}{f}^T(\underset{\sim}{x})\underset{\sim}{\beta} + \underset{\sim}{e} \\ &= \beta_1 f_1(\underset{\sim}{x}) + \ldots\ldots + \beta_k f_k(\underset{\sim}{x}) + \underset{\sim}{e} \end{aligned} \qquad (3.1.5)$$

where the components f_1, f_2, ..., f_k of the column $\underset{\sim}{f}$ are known, postulated, mutual independent functions of $\underset{\sim}{x}$ while the components β_1, β_2, ..., β_k of the column $\underset{\sim}{\beta}$ are unknown parameters. Often $f_1(\underset{\sim}{x})$ is chosen equal to 1 for all $\underset{\sim}{x}$. The postulated functions f_1, f_2, ..., f_k are called model functions or regressor variables. The relation (3.1.5) is the so-called regression equation. For the present only linear regression equations (i.e. linear in the components of $\underset{\sim}{\beta}$) are considered. In Chapter 4 some comments on nonlinear regression are given.

The unknown parameters $\underset{\sim}{\beta}$ have to be determined from the results of a series of N observations characterized by the treatments $\underset{\sim}{x}_1$, $\underset{\sim}{x}_2$ ..., $\underset{\sim}{x}_N$. All these treatments need not necessarily be distinct; observations may be repeated, i.e. carried out under the same treatment. It is the objective of the design of experiments to select the most suitable set of N treatments from the set χ of all candidate treatments such that the obtained results for $\underset{\sim}{\beta}$ are as accurate and reliable as possible for the given number N ($N \geq 1$) of treatments.

For the treatments $\underset{\sim}{x}_1$, ..., $\underset{\sim}{x}_N$ the sample model is:

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{e} \qquad (3.1.6)$$

where $\underset{\sim}{y} = [y_1\ y_2\ \ldots\ y_N]^T$ is the column of response quantities, $\underset{\sim}{e}$ is the column of errors and X is an (N*k)-matrix, the design matrix, which is given by:

$$X^T = [\underset{\sim}{f}(\underset{\sim}{x}_1)\quad \underset{\sim}{f}(\underset{\sim}{x}_2)\ \ldots\ \underset{\sim}{f}(\underset{\sim}{x}_N)] \qquad (3.1.7)$$

Unbiased estimates for $\underset{\sim}{\beta}$ can be computed if the following assumptions for $\underset{\sim}{e}$ are made:

1. the expected value $E(\underset{\sim}{e})$ is zero:
   $$E(\underset{\sim}{e}) = \underset{\sim}{0} \qquad (3.1.8)$$

2. the variance of the responses and hence of the errors are the same and equal to σ^2 for all treatments:
   $$E(\underset{\sim}{e}\ \underset{\sim}{e}^T) = V(\underset{\sim}{e}) = \sigma^2 I \qquad (3.1.9)$$

In general, the number of treatments in the experiment should exceed the number of the parameters $\underset{\sim}{\beta}$. Then $\underset{\sim}{\beta}$ can be estimated using

a least-squares technique. $\hat{\underset{\sim}{\beta}}$ is calculated from the requirement that the residual sum of squares, KS_r, defined by

$$KS_r = (\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})^T(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}) \tag{3.1.10}$$

is minimal. Here $\underset{\sim}{y}$ represents a column with N measured response quantity values. This results in a set of k so-called normal equations:

$$X^TX\hat{\underset{\sim}{\beta}} = X^T\underset{\sim}{y} \tag{3.1.11}$$

For each appropriate experimental design, i.e. each appropriate set of treatments $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_N$, the matrix X^TX will be regular and hence positive definite. Therefore $\hat{\underset{\sim}{\beta}}$ can be solved from (3.1.11), yielding

$$\hat{\underset{\sim}{\beta}} = (X^TX)^{-1}X^T\underset{\sim}{y} \tag{3.1.12}$$

For each treatment $\underset{\sim}{x}$ estimates $\hat{y}$ of the response variable y can be calculated from

$$\hat{y}(\underset{\sim}{x}) = \underset{\sim}{f}^T(\underset{\sim}{x})\hat{\underset{\sim}{\beta}} \tag{3.1.13}$$

This relation represents an operational regression model of the mathematical model (3.1.2).

A measure for the accuracy of the estimator $\hat{\underset{\sim}{\beta}}$ is the variance-covariance matrix $V(\hat{\underset{\sim}{\beta}})$:

$$V(\hat{\underset{\sim}{\beta}}) = E((\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})^T) = \sigma^2\,(X^TX)^{-1} \tag{3.1.14}$$

Furthermore, the correlations between the elements of $\hat{\underset{\sim}{\beta}}$ are determined by the correlation coefficient matrix ϱ, the elements of which are given by

$$\varrho_{ij} = \frac{V(\hat{\underset{\sim}{\beta}})_{ij}}{(V(\hat{\underset{\sim}{\beta}})_{ii}\; V(\hat{\underset{\sim}{\beta}})_{jj})^{1/2}} \quad \text{for } i, j = 1, 2, \ldots, k \tag{3.1.15}$$

For the response estimator $\hat{\underset{\sim}{y}}(\underset{\sim}{x})$ the variance $V(\hat{\underset{\sim}{y}}(\underset{\sim}{x}))$ is used as a measure for its accuracy. From (3.1.13) and (3.1.14) it follows:

$$V(\hat{\underset{\sim}{y}}(\underset{\sim}{x})) = \underset{\sim}{f}^T(\underset{\sim}{x})(X^TX)^{-1}\,\underset{\sim}{f}(\underset{\sim}{x})\,\sigma^2 \tag{3.1.16}$$

Often the variance σ^2 of the measured responses is unknown. An estimate for σ^2 can be calculated from

$$\hat{\sigma}^2 = \frac{1}{n-k}\,KS_r \tag{3.1.17}$$

Using this result estimates $\hat{V}(\hat{\underset{\sim}{\beta}})$, $\hat{\varrho}$ and $\hat{V}(\hat{\underset{\sim}{y}}(\underset{\sim}{x}))$ for $V(\hat{\underset{\sim}{\beta}})$, ϱ and $V(\hat{\underset{\sim}{y}}(\underset{\sim}{x}))$ can be calculated.

As stated before it is the objective of the design of experiments to select an optimal set of N treatments from the set χ, such that the obtained estimation $\underset{\sim}{\beta}$ for $\underset{\sim}{\beta}$ is as accurate and reliable as possible. In order to distinguish between candidate points and points actually present in the experimental design, candidate points are denoted by $\underset{\sim}{x}_j$, $j = 1, 2, \ldots, r$, whereas design points are denoted by $\underset{\sim}{x}_j$, $j = 1, 2, \ldots, N$. In a numerical experiment the selected treatments $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_N$ will all be different. In physical experiments however replicating observations may result in a more accurate estimation of $\underset{\sim}{\beta}$. Given the number N of treatments and given the set χ of candidate treatments the objective of the design of experiments is to

> determine for each candidate treatment $\underset{\sim}{x}_j \in \chi$, $j = 1, 2, \ldots, r$, the number N_j ($N_j \geq 0$) of observations such that $\underset{\sim}{\beta}$ is as accurate and reliable as possible and such that the total number of observations is equal to N, i.e.

$$N = \sum_{j=1}^{r} N_j \tag{3.1.18}$$

For this purpose several methods are available. We will treat two of these methods. In the next section a more or less classical method, resulting in so-called 2^n-designs, is discussed. The optimal design theory is the subject of the Sections 3.3 and 3.4.

3.2 The 2^n factorial design

3.2.1 The concept of 2^n designs

The design of a factorial experiment, where each of the n control variables is varied on two levels, is called a 2^n factorial design. Such designs are very popular and are quite suitable to develop regression models. Furthermore, they are useful in optimal design methods since they can constitute a set of candidate treatments.

There exists a wealth of literature on 2^n designs (e.g. Box et al. (1978), Doornbos (1984), Montgomery (1984)) and only the main topics of the method are discussed here.

In 2^n factorial designs a special notation is adopted. The control variables, i.e. the factors, are indicated by capitals: A, B, C ... etc., and can be adjusted to two levels: "high" and "low". "High" and "low" are just names; their specific meaning must be properly defined. A treatment is indicated by a string of lowercase letters. This string only contains those characters that correspond to the factors at the "high" level. The observation of the response quantity for a given treatment is denoted by a string of the same lowercase letters as the string denoting that treatment. For example, observation ab is the result of the treatment ab, where the factors A and B are both at the "high" level. Fig. 3.1 illustrates the possible treatments in a complete 2^3 experiment with the factors A, B and C.

The treatment (1) in Fig. 3.1, represents the treatment with all factors at the "low" level.

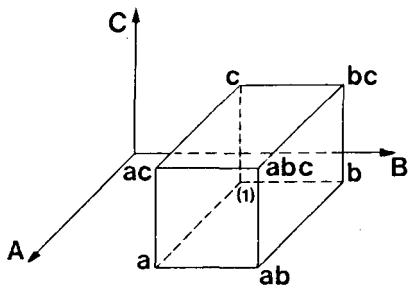


Fig. 3.1 Eight possible treatments in a 2^3 experiment.

A complete 2^3 experiment contains eight different treatments, which enables us to estimate up to a maximum of eight parameters in a regression equation, for instance:

$$y = \beta_1 + \beta_A x_A + \beta_B x_B + \beta_C x_C + \beta_{AB} x_A x_B + \beta_{AC} x_A x_C + \beta_{BC} x_B x_C + \beta_{ABC} x_A x_B x_C \qquad (3.2.1)$$

The parameters are often called effects and are usually indicated by the same upper case characters as the factors: the effect β_1 is indicated by I, β_A by A etc. Main effects are those effects which are indicated by just one letter, i.e. the effects I, A, B and C in the considered example. Effects indicated by more than one letter, for example AB, AC, etc., are called interactions or interactive effects.

In most cases it is possible to code the levels of the factors to +1 and -1. If for the 2^3 experiment at every treatment one observation is made, the design matrix X is given by

$$X = \begin{array}{c} \text{Effects} \\ \begin{array}{cccccccc} I & A & B & C & AB & AC & BC & ABC \end{array} \\ \begin{bmatrix} +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ +1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix} \end{array} \quad \begin{array}{c} \text{Observations } \underset{\sim}{y} \\ \begin{bmatrix} (1) \\ a \\ b \\ c \\ ab \\ ac \\ bc \\ abc \end{bmatrix} \end{array} \qquad (3.2.2)$$

and the parameters β_1 through β_{ABC} can be estimated using (3.1.12).

Another algorithm to estimate a certain effect can be explained with the help of the above mentioned example. For instance an estimate $\hat{AC}$ (= β_{AC}) of the effect AC can be calculated as follows:

1. summing all the observations with $x_A * x_C = +1$ ("high"):
   h = (1) + b + ac + abc

2. next summing all the observations with $x_A * x_C = -1$ ("low"):
   l = a + c + ab + bc

3. dividing the difference h-l by the total number of observations, results in the estimate

$$\hat{AC} = \frac{1}{8} \{(1) + b + ac + abc) - (a + c + ab + bc)\} \qquad (3.2.3)$$

For the other effects similarly simple relations can be derived. The right hand side of (3.2.3) is called the contrast associated with the effect AC. Yates developed a special algorithm for the analysis of 2^n-experiments (e.g. see Doornbos (1984))

3.2.2 Blocks and fractional 2^n experiments

In order to estimate all possible effects from a 2^n design a complete set of 2^n treatments should be used and under each treatment at least one observation should be made. For increasing n the measurement program will become impractically large and the number of measurements must be reduced. This is possible since in many problems the great majority of effects are interactions of little importance. To estimate the effects of interest usually much fewer observations are necessary than in a complete 2^n factorial design. To account for this a (small) number of chosen treatments is grouped in a so-called block and observations are made only for the treatments in this block. The crucial point in this procedure is the block definition, i.e. the choice of the treatments in the block. Since the number of observations is reduced, it is not possible to estimate all effects separately: some effects will be "confounded". The effects related to the block coincide with other effects and only the "sum" of those effects can be estimated. If observations are made only for one block, then the effects which are confounded with the block effect cannot be estimated (e.g. Box et al. (1978)).

3.2.3 Utilities for the definition of fractional factorial designs

In many factorial experiments the control variables are to be varied on more than two levels, for instance if the functions $f_i(\underset{\sim}{x})$ in (3.1.5) are nonlinear in $\underset{\sim}{x}$. Generalization of the methods for 2^n experiments to methods for more than two levels is somewhat cumbersome. However, in situations where the number of levels is a power of 2, the designs can be transformed into 2^n designs (v. Heck (1983)), and the techniques developed for 2^n designs can still be used.

For blocking and reduction of 2^n designs comprehensive tabulated factorial designs are available (Cochran et al. (1950), NBS (1957), Cox (1958), Box et al. (1978)). Furthermore, computer algorithms have been developed. The program DSIGN, developed by Patterson (1976), produces factorial designs for control variables at any

number of levels with a great variety of blocking structures. The designs are compared on the basis of their confounding patterns. The program CADE (Nagtegaal (1987)) incorporates besides algorithms for optimal design, also facilities for the generation of (fractional) 2^n designs.

3.3 Optimal design theory

3.3.1 Introduction

The central theme in the design of experiments is the formulation of experimental designs for the construction of regression models such that a suitable regression model can be determined with a minimum of experimental effort. In the optimal design theory direct evaluation criteria are used: the accuracy of the estimated parameters itself and/or the variance in the estimated response and iterative algorithms are used to search for optimal designs. Optimal design techniques are useful in those situations where classic designs are unsuitable or unavailable, that is when

- the experimental region is irregularly shaped due to constraints on the control variables,
- it is needed to augment or repair an existing design,
- the number of levels for the control variables varies considerable,
- designs have to be constructed for special models, i.e. other than polynomial models,
- designs have to be constructed for simultaneous observation of several responses.

The optimal design theory was initiated by Smith (1918). Significant progress was stimulated by the development of digital computers. Important contributions to the methods were made by Kiefer (1959) and Fedorov (1972). Nagtegaal (1987) gave some generalizations for the construction of experimental designs in case several regression models are being used simultaneously.

3.3.2 Optimality criteria

In the optimal design of experiments a number of N, not necessarily distinct treatments $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_N$ are determined from an a priori chosen and fixed set χ of r discrete candidate treatments $\underset{\sim}{\chi}_1, \underset{\sim}{\chi}_2, \ldots, \underset{\sim}{\chi}_r$. The control variables $x_1, x_2, \ldots, x_n$ are not allowed to vary continuously. It is assumed that somehow an appropriate choice for χ can be made. The main reason for this restriction is to symplify optimization algorithms resulting in less computing time for the determination of optimal EDs. It should be emphasized that in the sequel optimality of designs is always subject to these restrictions. The outcomes of optimization algorithms depend on the choice of the set of candidate points. For each of the treatments $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_N$, often called "design points" or shortly "points", an observation will be made so these points define an ED and the objective is to optimize that ED. The problem is now to choose N treatments from the r candidate treatments, resulting in the best N-point design.

Experimental designs can be evaluated, using the variances of the parameter estimator $V(\hat{\underset{\sim}{\beta}})$ or the variances of the response estimator $V(\hat{\underset{\sim}{y}}(\underset{\sim}{x}))$ as a measure, see (3.1.14) and (3.1.16) respectively. In both cases the quality of the ED is a function of the matrix $(X^TX)^{-1}$ and the objective is to determine that ED among all possible N-point EDs which makes $(X^TX)^{-1}$ minimal. However, the minimum of a matrix is not a well defined concept and a number of operational criteria have been developed. The most important of these criteria are:

- D-optimality, which is achieved if $\det (X^TX)^{-1}$ is minimal, i.e. if the product of the eigenvalues of $(X^TX)^{-1}$ is minimal.
- A-optimality, which is achieved if $\mathrm{tr}(X^TX)^{-1}$ is minimal, i.e. if the sum of the eigenvalues of $(X^TX)^{-1}$ is minimal.
- E-optimality, which is achieved if the largest eigenvalue of $(X^TX)^{-1}$ is minimal.
- G-optimality, which is achieved if the maximum over all candidate points of the estimated response variance $V(\hat{\underset{\sim}{y}}(\underset{\sim}{x}))$ is minimal.
- V-optimality, which is achieved if the estimated response variance, averaged over all candidate points, $\frac{1}{r}\sum_{j=1}^{r} V(\hat{\underset{\sim}{y}}(\underset{\sim}{x}_j))$ is minimal.

3.3.3 Definition of the experimental design

An N-point ED is completely specified by the choice of the candidate treatments and of the number N_j ($N_j \geq 0$) of observations under candidate treatment $\underset{\sim}{x}_j$ ($j = 1, 2, \ldots, r$), such that (3.1.18) holds. The numbers $N_1, N_2, \ldots, N_r$ are the elements of a column $\underset{\sim}{N}$:

$$\underset{\sim}{N} = [N_1 \ldots N_r]^T \qquad (3.3.1)$$

For each of the numbers $N_1, \ldots, N_r$ upper and lower bounds can be specified, apart from the trivial condition $0 \leq N_j \leq N$ for $j = 1, 2, \ldots, r$. Some typical conditions are:

$N_j \leq 1$ if point $\underset{\sim}{x}_j$ is allowed only once,
$1 \leq N_j$ if point $\underset{\sim}{x}_j$ should appear at least once,
$N_j^* \leq N_j$ if an existing ED has to be augmented and already N_j^* observations have been made in point $\underset{\sim}{x}_j$.

In practice the number r of candidate treatments and each of the numbers $N_1, \ldots, N_r$ are finite and integer. For numerical purposes, however, it may be advantageous to allow these numbers to be real. An ED with integer numbers is called an exact ED. Otherwise the ED is called discrete.

3.3.4 The information matrix and normalized optimality criteria

The matrix X^TX can be normalized, resulting in the information matrix M, where:

$$\underset{\sim}{M}(\underset{\sim}{N}) = \frac{1}{N} \sum_{j=1}^{r} N_j \, \underset{\sim}{f}(\underset{\sim}{x}_j) \, \underset{\sim}{f}^T(\underset{\sim}{x}_j) \qquad (3.3.2)$$

In order to determine an optimal ED we try to find a column $\underset{\sim}{N}^*$ for which a given criterion $C(\underset{\sim}{N})$ is minimal and for which the constraints are not violated. This criterion may be any of the earlier given criteria. Written in terms of M instead of X^TX these criteria take the following forms:

$$C(\underset{\sim}{N}) = \det(\underset{\sim}{M}^{-1}(\underset{\sim}{N})) \text{ for D-optimality} \qquad (3.3.3)$$

$$C(\underset{\sim}{N}) = \mathrm{tr}(\underset{\sim}{M}^{-1}(\underset{\sim}{N})) \text{ for A-optimality} \qquad (3.3.4)$$

$$C(\underset{\sim}{N}) = \lambda_{max}(\underset{\sim}{N}) \text{ for E-optimality} \qquad (3.3.5)$$

where $\lambda_{max}(\underset{\sim}{N})$ is the largest eigenvalue of $\underset{\sim}{M}^{-1}(\underset{\sim}{N})$

$$C(\underset{\sim}{N}) = \max_{j=1}^{r} d(\underset{\sim}{x}_j, \underset{\sim}{N}) \text{ for G-optimality} \qquad (3.3.6)$$

$$C(\underset{\sim}{N}) = \frac{1}{r} \sum_{j=1}^{r} d(\underset{\sim}{x}_j, \underset{\sim}{N}) \text{ for V-optimality} \qquad (3.3.7)$$

In the last two criteria $d(\underset{\sim}{x}, \underset{\sim}{N})$ is the variance of the response estimator in the point $\underset{\sim}{x}$, normalized for σ^2/N:

$$d(\underset{\sim}{x}, \underset{\sim}{N}) = \underset{\sim}{f}^T(\underset{\sim}{x}) \, \underset{\sim}{M}^{-1}(\underset{\sim}{N}) \underset{\sim}{f}(\underset{\sim}{x}) \qquad (3.3.8)$$

3.3.5 Optimization algorithms

Four classes of optimization algorithms can be distinguished. The first class contains algorithms which generate and evaluate all possible N-point designs. This is very time consuming and only feasible for small values of N. As a variant Welch (1982) developed a "branch and bound"-algorithm, in which a binary tree of minimization problems is generated. Not all designs are generated and evaluated. By exploiting bounds on the minimization only branches which might contain D-optimal designs are created. The algorithm guarantees global D-optimal designs, but computing costs are extremely high.

The second class of algorithms uses mathematical optimization techniques for a direct maximization of $\det(\underset{\sim}{M}(\underset{\sim}{N}))$, which is equivalent to minimization of $\det(\underset{\sim}{M}^{-1}(\underset{\sim}{N}))$. Due to the complexity of $\det(\underset{\sim}{M}(\underset{\sim}{N}))$ these methods are not successful.

The third class of algorithms neglects the integer character of the components of $\underset{\sim}{N}$ and, in consequence, results in a discrete ED. Subsequently this ED is rounded off to an exact ED. This may result in a good approximation if the number of points, N, is sufficiently higher than the number of parameters k. For this reason discrete

designs are sometimes called "approximate designs". Such a design can be used as an initial design in exchange algorithms. Algorithms for the construction of optimal discrete designs depend on the optimality criterion to be used (Fedorov (1972)).

Finally, the fourth class concerns exchange algorithms. Mitchell (1974 a, b) developed an efficient algorithm called DETMAX, which is the most popular of all exchange algorithms. The algorithm starts with an initial m-point ED; the final goal is an optimal N-point ED. During each iteration step that candidate point, which results in the largest increase of det(M), is added to the design, and subsequently that point, which results in the smallest decrease of det(M), is removed from the design. The number m of points in the initial design may be larger or smaller than N. If necessary the algorithm first adds (if m<N) or rejects (if m>N) points until the number of points in the ED is equal to N. In order to avoid local optima the algorithm is able to perform "excursions", in which several points are added at one go and subsequently the number of points is reduced to N. If the resulting N-point ED has not been improved, another excursion will be made from the same initial design. If the excursion is successful the resulting ED will be used as starting ED in a further attempt to maximize det(M). The algorithm terminates when, after several excursions, no better ED is found. The algorithm generates high quality EDs against relatively low computing costs.

3.3.6 Robust_experimental designs

An ED is called "robust" if parameters and responses can be adequately estimated, even when the basic assumptions, for the construction of the ED, are violated to some extent. We might, for example assume that the considered response quantity is stochastic with a $N(0,\sigma^2)$ distribution. In order to construct EDs we must specify a regression model in advance. However, the goodness of fit of the model can be checked only after the observations have been made. The model may prove to be biased and is said to be not exact.

Optimal EDs based on the criteria discussed earlier, show a dependency on the postulated regression model. They do not result in good estimates of the parameters when the exact model differs significantly from the assumed model.

In order to account for bias, the so-called J-criterion can be used (Box and Draper (1959)). This criterion is defined as the average mean squared error over all candidate points, normalized for σ^2/N

$$J = \frac{N}{\sigma^2} \frac{1}{r} \sum_{j=1}^{r} E[(\hat{y}(\underset{\sim}{x}_j) - y(\underset{\sim}{x}_j))^2] \qquad (3.3.9)$$

J can be expressed as the sum of a variance and a bias term:

$$J = V + B, \qquad (3.3.10)$$

where V is the normalized average variance:

$$V = \frac{1}{r} \sum_{j=1}^{r} d(\underset{\sim}{x}_j, \underset{\sim}{N}) = \frac{1}{r} \sum_{j=1}^{r} \underset{\sim}{f}^T(\underset{\sim}{x}_j) M^{-1}(\underset{\sim}{N}) \underset{\sim}{f}(\underset{\sim}{x}_j) \qquad (3.3.11)$$

and B is the normalized average squared bias given by:

$$B = \frac{N}{\sigma^2} \frac{1}{r} \sum_{j=1}^{r} (E(\hat{\underset{\sim}{y}}(\underset{\sim}{x}_j)) - \underset{\sim}{y}(\underset{\sim}{x}_j))^2 \qquad (3.3.12)$$

A problem is that B depends on σ^2 and on the unknown exact responses. However, Welch (1983) developed a method for the construction of robust designs based on the J-criterion. Nagtegaal (1987) generalized the method to the case of simultaneous observations of several response quantities.

Whether or not a postulated model is exact is seldom known beforehand, and one may be tempted to use robust designs in each case. This is not recommended, because in experiments where the model is nearly exact a robust design would increase the variance much more than it would reduce the bias. In this case a robust design will result in worse parameter- and response estimates, compared to designs based on one of the earlier mentioned criteria. Robust designs should be used only in those cases where the adequacy of the postulated model is seriously doubted.

A first choice of the regression model can be based on experience, on preliminary measurements or on theoretical considerations about the system under study. In general it is worthwhile to invest relatively much effort in such examinations.

3.4 Experimental designs in case of simultaneous observations of several quantities

3.4.1 Introduction

Sometimes more than one response quantity is relevant. Each of these quantities will have its own functional relationship with the control variables. Often they will be correlated, because they emerge from the same system.

A possible approach in the case of several responses would be to perform an experimental investigation for each response quantity separately. From economical point of view it is much more favourable to construct one optimal ED, which takes into account all response quantities simultaneously.

In Chapter 2 it was shown how partial derivatives of response quantities such as displacements and eigenfrequencies can be derived in FEM-formulations. Generalization to the case of several responses allows us to make use of these derivatives in experimental design and model building. The models for the partial derivatives of a response quantity with respect to the control variables follow by differentiation of the postulated model of that response quantity. In the following subsections some generalized definitions and optimization criteria are presented. Generalizations of optimization algorithms can be found in Fedorov (1972) (for D-optimality only) and Nagtegaal (1987).

3.4.2 Definitions

Let m be the number of response quantities $y_1, \ldots, y_m$ and let $\underset{\sim}{u}$ be the column with these variables as elements:

$$\underset{\sim}{u} = [y_1 \; y_2 \; \cdots \; y_m]^T \tag{3.4.1}$$

Since only linear models are used, response quantity y_j (j = 1, 2, ..., m) is assumed to depend linearly on k_j modelparameters, being the elements of a column $\underset{\sim}{\beta}_j$. Then the model equations are given by:

$$y_j(\underset{\sim}{x}) = \underset{\sim}{f}_j^T(\underset{\sim}{x})\underset{\sim}{\beta}_j \quad \text{for} \quad j = 1, 2, \ldots, m \tag{3.4.2}$$

where $\underset{\sim}{f}_j(\underset{\sim}{x})$ is a column whose elements are given functions, the so-called model functions or regressor variables.

The postulated models (3.4.2) can be represented in a compact form as follows. The columns containing the model parameters $\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, \ldots, \underset{\sim}{\beta}_m$ are filed in a column $\underset{\sim}{\beta}$ given by:

$$\underset{\sim}{\beta} = [\underset{\sim}{\beta}_1^T \; \underset{\sim}{\beta}_2^T \; \cdots \; \underset{\sim}{\beta}_m^T]^T \tag{3.4.3}$$

Column $\underset{\sim}{\beta}$ contains k elements, where k is given by:

$$k = \sum_{j=1}^{m} k_j \tag{3.4.4}$$

Next, each column $\underset{\sim}{f}_j(\underset{\sim}{x})$ is extended with m-1 columns $\underset{\sim}{z}_i$, i = 1, 2, ..., m, i ≠ j, each containing k_i zeros, to a column $\underset{\sim}{f}_j^*(\underset{\sim}{x})$ given by:

$$\underset{\sim}{f}_j^*(\underset{\sim}{x}) = [\underset{\sim}{z}_1^T \; \underset{\sim}{z}_2^T \; \cdots \; \underset{\sim}{z}_{j-1}^T \; \underset{\sim}{f}_j^T(\underset{\sim}{x}) \; \underset{\sim}{z}_{j+1}^T \; \cdots \; \underset{\sim}{z}_m^T]^T \tag{3.4.5}$$

Using these definitions the model equations (3.4.2) can be written as:

$$\underset{\sim}{u}(\underset{\sim}{x}) = F^T(\underset{\sim}{x})\underset{\sim}{\beta} \tag{3.4.6}$$

where $F(\underset{\sim}{x})$ is a (k*m) matrix, such that column j of $F(\underset{\sim}{x})$ is equal to $\underset{\sim}{f}_j^*(\underset{\sim}{x})$:

$$F(\underset{\sim}{x}) = [\underset{\sim}{f}_1^*(\underset{\sim}{x}) \; \underset{\sim}{f}_2^*(\underset{\sim}{x}) \; \cdots \; \underset{\sim}{f}_m^*(\underset{\sim}{x})] \tag{3.4.7}$$

In this thesis our attention is focused on the special case in which elements of $\underset{\sim}{u}(\underset{\sim}{x})$, say $y_i(\underset{\sim}{x})$, $1 \leq i \leq m-n$, are response quantities, while the elements $y_{i+1}, y_{i+2}, \ldots y_{i+n}$ are partial derivatives of $y_i(\underset{\sim}{x})$ with respect to the elements of $\underset{\sim}{x}$. Then only the column $\underset{\sim}{f}_i(\underset{\sim}{x})$ must be specified, since the columns $\underset{\sim}{f}_{i+1}(x) \ldots \underset{\sim}{f}_{i+n}(x)$ follow by differentiation of $\underset{\sim}{f}_i(x)$ with respect to the elements of $\underset{\sim}{x}$. In addition all the columns $\underset{\sim}{f}_i^*, \underset{\sim}{f}_{i+1}^* \ldots \underset{\sim}{f}_{i+n}^*$ are extended in the same way to columns $\underset{\sim}{f}_i, \underset{\sim}{f}_{i+1}, \ldots \underset{\sim}{f}_{i+n}$, given by:

$$
\begin{aligned}
\underset{\sim}{f}_i^* &= [\underset{\sim}{z}_1^T \cdots \underset{\sim}{z}_{i-1}^T \; \underset{\sim}{f}_i^T \; \underset{\sim}{z}_{i+n+1}^T \cdots \underset{\sim}{z}_m^T]^T \\
\underset{\sim}{f}_{i+1}^* &= [\underset{\sim}{z}_1^T \cdots \underset{\sim}{z}_{i-1}^T \; \underset{\sim}{f}_{i+1}^T \; \underset{\sim}{z}_{i+n+1}^T \cdots \underset{\sim}{z}_m^T]^T \qquad (3.4.8) \\
&\;\;\vdots \\
\underset{\sim}{f}_{i+n}^* &= [\underset{\sim}{z}_1^T \cdots \underset{\sim}{z}_{i-1}^T \; \underset{\sim}{f}_{i+n}^T \; \underset{\sim}{z}_{i+n+1}^T \cdots \underset{\sim}{z}_m^T]^T
\end{aligned}
$$

Each of the columns in (3.4.8) contains k elements, where k is given by:

$$
k = \sum_{j=1}^{i} k_j \quad + \quad \sum_{j=i+n+1}^{m} k_j \qquad (3.4.9)
$$

Since the regression equations of response quantity $y_i(x)$ and its partial derivatives contain the same unknown parameters $\underset{\sim}{\beta}_i$, the parameters in this special case can be filed in a column, given by:

$$
\underset{\sim}{\beta} = [\underset{\sim}{\beta}_1^T \cdots \underset{\sim}{\beta}_{i-1}^T \; \underset{\sim}{\beta}_i^T \; \underset{\sim}{\beta}_{i+n+1}^T \cdots \underset{\sim}{\beta}_m^T]^T \qquad (3.4.10)
$$

Using the definitions (3.4.7) through (3.4.10) the regression equations for this case can also be written in the same compact form of (3.4.6). We illustrate this by means of the following example:

$\underset{\sim}{u}(\underset{\sim}{x}) = [y_1 \; y_2 \; \cdots \; y_5]^T$ where, for instance:

$\underset{\sim}{x} = [x_1 \; x_2]^T$ and

$y_1 = \beta_1 + \beta_2 x_1$

$y_2 = \beta_3 + \beta_4 x_2$

$y_3 = \beta_5 + \beta_6 x_1 + \beta_7 x_2 + \beta_8 x_1 x_2$

$y_4 = \dfrac{\partial y_3}{\partial x_1} = \beta_6 + \beta_8 x_2$

$y_5 = \dfrac{\partial y_3}{\partial x_2} = \beta_7 + \beta_8 x_1$

Then $\underset{\sim}{u}(x) = F^T(\underset{\sim}{x})\underset{\sim}{\beta}$, where $F^T(\underset{\sim}{x})$ and $\underset{\sim}{\beta}$ are given by:

$$
F^T(\underset{\sim}{x}) = \left[\begin{array}{cc|cc|cccc}
1 & x_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & x_2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & x_1 & x_2 & x_1x_2 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & x_2 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & x_1
\end{array}\right]
$$

$$
\underset{\sim}{\beta} = [\beta_1 \; \beta_2 \mid \beta_3 \; \beta_4 \mid \beta_5 \; \beta_6 \; \beta_7 \; \beta_8]^T
$$

As in the previous section N, not necessarily distinct design points $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_N$ are to be selected from a set χ of r candidate points. In these design points measurements will be carried out, resulting in values $\underset{\sim}{u}(\underset{\sim}{x}_1), \ldots, \underset{\sim}{u}(\underset{\sim}{x}_N)$ for the responses $\underset{\approx}{u}(\underset{\sim}{x}_1), \ldots, \underset{\approx}{u}(\underset{\sim}{x}_n)$. Denoting the difference between $\underset{\approx}{u}(\underset{\sim}{x}_i)$ and $\underset{\sim}{u}(\underset{\sim}{x}_i)$ by $\underset{\approx}{e}_i$ it is seen that:

$$\underset{\approx}{u}(\underset{\sim}{x}_i) = \underset{\sim}{u}(\underset{\sim}{x}_i) + \underset{\approx}{e}_i = F^T(\underset{\sim}{x}_i)\underset{\sim}{\beta} + \underset{\approx}{e}_i \quad \text{for } i = 1, 2, \ldots, N \quad (3.4.11)$$

It is assumed that the elements of $\underset{\approx}{e}_i$ are uncorrelated random errors with expected value zero and known constant variance. Furthermore it is assumed that $\underset{\approx}{e}_i$ and $\underset{\approx}{e}_j$ are uncorrelated if $i \neq j$:

$$E(\underset{\approx}{e}_i) = \underset{\sim}{0} \; ; \; E(\underset{\approx}{e}_i \, \underset{\approx}{e}_j^T) = \delta_{ij} S^2 \quad \text{for } i, j = 1, 2, \ldots, N \quad (3.4.12)$$

Here, S is a known m*m diagonal matrix whose diagonal elements $\sigma_1, \ldots \sigma_m$ are the standard deviations of the response quantities $y_1, \ldots y_m$.

Because the error variances are unequal, $\underset{\sim}{\beta}$ should be estimated using the methode of weighted least squares. By premultiplying (3.4.11) with S^{-1} the new sample model becomes:

$$\underset{\approx}{w}(\underset{\sim}{x}_i) = W^T(\underset{\sim}{x}_i) \, \underset{\sim}{\beta} + \underset{\approx}{\eta}_i \quad \text{for} \quad i = 1, 2, \ldots, N \quad (3.4.13)$$

where $\underset{\approx}{w}(\underset{\sim}{x}_i) = S^{-1}\underset{\approx}{u}(\underset{\sim}{x}_i)$, $W^T(\underset{\sim}{x}_i) = S^{-1}F^T(\underset{\sim}{x}_i)$ and $\underset{\approx}{\eta}_i = S^{-1}\underset{\approx}{e}_i$

The errors $\underset{\approx}{\eta}_i$ still have zero expectation:

$$E(\underset{\approx}{\eta}_i) = S^{-1}E(\underset{\approx}{e}_i) = \underset{\sim}{0}$$

However, the error variance matrix of $\underset{\approx}{\eta}_i$, $V(\underset{\approx}{\eta})$, now becomes unity:

$$E(\underset{\approx}{\eta}_i \, \underset{\approx}{\eta}_i^T) = E(S^{-1}\underset{\approx}{e}_i\underset{\approx}{e}_i^T \, S^{-1}) = S^{-1}E(\underset{\approx}{e}_i\underset{\approx}{e}_i^T) \, S^{-1} = S^{-1}SSS^{-1} = I$$

Therefore model (3.4.13) satisfies the standard least squares assumptions and $\underset{\sim}{\beta}$ can be estimated by an ordinary least squares fit of the new model.

Estimates $\hat{\underset{\sim}{\beta}}$ for the model parameters $\underset{\sim}{\beta}$ can now be determined from the requirement that the weighted residual sum of squares,

$$KS_r = \sum_{i=1}^{N} [(\underset{\sim}{w}(\underset{\sim}{x}_i) - W^T(\underset{\sim}{x}_i)\hat{\underset{\sim}{\beta}})^T(\underset{\sim}{w}(\underset{\sim}{x}_i) - W^T(\underset{\sim}{x}_i)\hat{\underset{\sim}{\beta}})], \quad (3.4.14)$$

is minimal. As in section 3.1, this results in a set of k equations for $\hat{\underset{\sim}{\beta}}$ with the solution

$$\hat{\underset{\sim}{\beta}} = [\sum_{i=1}^{N} W(\underset{\sim}{x}_i) \, W^T(\underset{\sim}{x}_i)]^{-1} \sum_{i=1}^{N} [W(\underset{\sim}{x}_i) \, \underset{\sim}{w}(\underset{\sim}{x}_i)] \quad (3.4.15)$$

According to (3.4.15) a weighted least squares estimator $\hat{\underset{\sim}{\beta}}$ can be defined as

$$\hat{\underset{\sim}{\beta}} = [\sum_{i=1}^{N} W(\underset{\sim}{x}_i)\ W^T(\underset{\sim}{x}_i)]^{-1}\ \sum_{i=1}^{N} [W(\underset{\sim}{x}_i)\ \underset{\sim}{w}(\underset{\sim}{x}_i)] \tag{3.4.16}$$

While

$$E(\hat{\underset{\sim}{\beta}}) = [\sum_{i=1}^{N} W(\underset{\sim}{x}_i)\ W^T(\underset{\sim}{x}_i)]^{-1}\ \sum_{i=1}^{N} W(\underset{\sim}{x}_i)\ E\underset{\sim}{w}(\underset{\sim}{x}_i)$$

$$= [\sum_{i=1}^{N} W(\underset{\sim}{x}_i)\ W^T(\underset{\sim}{x}_i)]^{-1}\ \sum_{i=1}^{N} W(\underset{\sim}{x}_i)\ W(\underset{\sim}{x}_i)\underset{\sim}{\beta} = \underset{\sim}{\beta} \tag{3.4.17}$$

(3.4.16) is an unbiased estimator for the parameters $\underset{\sim}{\beta}$. Furthermore it can be shown that the variance-covariance matrix of the parameter estimators, i.e. $E(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})^T$ is given by

$$V(\hat{\underset{\sim}{\beta}}) = N\ M^{-1} \sum_{i=1}^{N} [W(\underset{\sim}{x}_i)\ V(\underset{\sim}{0})\ W^T(\underset{\sim}{x}_i)]\ N\ M^{-1}$$

$$= [\sum_{i=1}^{N} W(\underset{\sim}{x}_i)\ W^T(\underset{\sim}{x}_i)]^{-1} \quad = \quad N\ M^{-1} \tag{3.4.18}$$

where $M = \frac{1}{N} \sum_{i=1}^{N} [W(\underset{\sim}{x}_i)\ W^T(\underset{\sim}{x}_i)]$ is the information matrix.

If the parameters are estimated by an ordinary least squares fit of the original model (3.4.11) then

$$\hat{\underset{\sim}{\beta}}_0 = [\sum_{i=1}^{N} F(\underset{\sim}{x}_i)\ F^T(\underset{\sim}{x}_i)]^{-1}\ \sum_{i=1}^{N} [F(\underset{\sim}{x}_i)\ \underset{\sim}{y}(\underset{\sim}{x}_i)] \tag{3.4.19}$$

is obtained. $\hat{\underset{\sim}{\beta}}_0$ is unbiased, that is, $E(\hat{\underset{\sim}{\beta}}_0) = \underset{\sim}{\beta}_0$ however,

$$V(\hat{\underset{\sim}{\beta}}_0) = [\sum_{i=1}^{N} F(\underset{\sim}{x}_i)\ F^T(\underset{\sim}{x}_i)]^{-1}\ \sum_{i=1}^{N} [F(\underset{\sim}{x}_i)\ S^2\ F^T(\underset{\sim}{x}_i)] \cdot$$

$$[\sum_{i=1}^{N} F\ (\underset{\sim}{x}_i)\ F^T(\underset{\sim}{x}_i)]^{-1} \tag{3.4.20}$$

If $S^2 = I$ then $V(\hat{\underset{\sim}{\beta}}_0) = V(\hat{\underset{\sim}{\beta}})$. If $S^2 \neq I$ then $V(\hat{\underset{\sim}{\beta}}_0) > V(\hat{\underset{\sim}{\beta}})$, i.e., $V(\hat{\underset{\sim}{\beta}}_0) > V(\hat{\underset{\sim}{\beta}})$ is a positive definitive matrix and an ordinary least squares fit of (3.4.11) results in greater variances than a weighted least squares fit.

In order to estimate $\underset{\sim}{\beta}$, the error variance matrix S^2 should be known beforehand. When S^2 is unknown, S^2 and $\underset{\sim}{\beta}$ can be estimated using an iterative reweighted least squares procedure:

step 1. Compute the ordinary least squares estimations for the original model (3.4.11)

$$\hat{\underset{\sim}{\beta}}_0 = [\sum_{i=1}^{N} F(\underset{\sim}{x}_i)\ F^T(\underset{\sim}{x}_i)]^{-1}\ \sum_{i=1}^{N} [F(\underset{\sim}{x}_i)\ \underset{\sim}{y}(\underset{\sim}{x}_i)]$$

step 2. Compute initial estimations of the error variances according to

$$\hat{\sigma}^2_{j_0} = \frac{1}{N-k} \sum_{i=1}^{N} [y_j(\underset{\sim}{x}_i) - \hat{y}_{j_0}(\underset{\sim}{x}_i)]^2 \quad \text{for } j = 1 \ldots, m$$

where $\hat{y}_{j_0}(\underset{\sim}{x}_i) = \underset{\sim}{f}_j^T(\underset{\sim}{x}_i) = \hat{\underset{\sim}{\beta}}_0 \quad , \quad i = 1, \ldots, N$

and construct $\hat{S}_0 = \begin{bmatrix} \hat{\sigma}_{1_0} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{m_0} \end{bmatrix}$

step 3. Premultiply model (3.4.11) with $\hat{S}_0^{-1}$. This results in a transformed model:

$$\underset{\sim}{w}_1(\underset{\sim}{x}_i) = \underset{\sim}{W}_1^T(\underset{\sim}{x}_i)\ \underset{\sim}{\beta} + \underset{\sim}{n}_1(\underset{\sim}{x}_i) \quad , \quad i = 1, \ldots, N \quad (3.4.21)$$

step 4. Apply the ordinary least squares method to this transformed model (3.4.21). This results in new estimations $\hat{\underset{\sim}{\beta}}_1$ for which holds $V(\hat{\underset{\sim}{\beta}}_1) \leq V(\hat{\underset{\sim}{\beta}}_0)$

step 5. Compute new estimations of the error variances of the original model (3.4.11) according to

$$\hat{\sigma}^2_{j_1} = \sum_{i=1}^{N} [y_j(\underset{\sim}{x}_i) - \hat{y}_{j_1}(\underset{\sim}{x}_i)]^2/(N-k), \text{ for } j = 1, \ldots, m$$

where $\hat{y}_{j_1}(\underset{\sim}{x}_i) = \underset{\sim}{f}_j^T(\underset{\sim}{x}_i)\ \hat{\underset{\sim}{\beta}}_1 \quad , \quad i = 1, \ldots, N$

and construct $\hat{S}_1 = \begin{bmatrix} \hat{\sigma}_{i_1} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{m_1} \end{bmatrix}$

Step 3 - 5 are repeated with $\hat{S}_1$ instead of $\hat{S}_0$, then with $\hat{S}_2$, and so on, until

$$\| \hat{\underset{\sim}{\beta}}_i - \hat{\underset{\sim}{\beta}}_{i-1} \| \leq \text{eps} \ \| \hat{\underset{\sim}{\beta}}_i \| ,$$

where eps is the relative machine precision.

Using these results, unbiased estimators $\hat{\underset{\sim}{u}}(\underset{\sim}{x})$ for the response quantities can be derived from:

$$\hat{\underset{\sim}{u}}(\underset{\sim}{x}) = \underset{\sim}{F}^T(\underset{\sim}{x})\hat{\underset{\sim}{\beta}} \quad (3.4.22)$$

The variance-covariance matrix $V(\hat{\underset{\sim}{u}}(\underset{\sim}{x}))$ is given by:

$$V(\hat{\underset{\sim}{u}}(\underset{\sim}{x})) = F^T(\underset{\sim}{x})\ V(\hat{\underset{\sim}{\beta}})\ F(\underset{\sim}{x}) \tag{3.4.23}$$

From (3.4.18) it is seen that $V(\hat{\underset{\sim}{\beta}})$ is determined completely by the statistical properties, S, of the measurement errors, the model functions $F = F(\underset{\sim}{x})$ and the design points $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_N$. Assuming that the error variance matrix S is known, then, for a given number N of design points the matrix $V(\hat{\underset{\sim}{\beta}})$ can only be influenced by the choice of the design points $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_N$ from the set χ of candidate points $\underset{\sim}{\chi}_1, \ldots, \underset{\sim}{\chi}_r$. As in section 3.3, the objective of optimal design of experiments for simultaneous responses is to determine the number $N_i (N_i \geq 0)$ of observations at candidate point $\underset{\sim}{\chi}_i (i=1, 2, \ldots, r)$, such that $N_1 + N_2 + \ldots + N_r$ equals the number N of design points. Using (3.4.18) it is readily shown that the variance-covariance matrix $V(\hat{\underset{\sim}{\beta}})$ can be considered a function of the column $\underset{\sim}{N} = [N_1\ N_2\ \ldots\ N_r]^T$,

$$V(\hat{\underset{\sim}{\beta}}, \underset{\sim}{N}) = N\ M^{-1}(\underset{\sim}{N}) \tag{3.4.24}$$

where $M(\underset{\sim}{N})$ follows from (3.4.18)

$$M(\underset{\sim}{N}) = \frac{1}{N} \sum_{j=1}^{N} [W(\underset{\sim}{x}_j) W^T(\underset{\sim}{x}_j)] = \frac{1}{N} \sum_{i=1}^{r} [N_i W(\underset{\sim}{\chi}_i) W^T(\underset{\sim}{\chi}_i)] \tag{3.4.25}$$

3.4.3 Optimality criteria

In the previous subsection it has been shown that the accuracies of the parameter estimators $\hat{\underset{\sim}{\beta}}$ and of the response estimators $\hat{\underset{\sim}{u}}(\underset{\sim}{x})$ are determined by the variance-covariance matrix $V(\hat{\underset{\sim}{\beta}}, \underset{\sim}{N})$ and therefore by the elements of $\underset{\sim}{N}$. These elements may be subject to constraints. As in subsection 3.3.4 this leads to the following definition of an optimal experimental design:

> the design of an experiment, characterized by $\underset{\sim}{N}^*$, is optimal if the elements of $\underset{\sim}{N}^*$ satisfy all specified constraints and a specified criterion $C = C(\underset{\sim}{N})$ is minimal for $\underset{\sim}{N} = \underset{\sim}{N}^*$.

The D-, A- and E- criteria remain the same:

$$C(\underset{\sim}{N}) = \det(\hat{V}(\hat{\underset{\sim}{\beta}}, \underset{\sim}{N})) \text{ for D-optimality} \tag{3.4.26}$$

$$C(\underset{\sim}{N}) = \mathrm{tr}(\hat{V}(\hat{\underset{\sim}{\beta}}, \underset{\sim}{N})) \text{ for A-optimality} \tag{3.4.27}$$

$$C(\underset{\sim}{N}) = \lambda_{max}(\underset{\sim}{N}) \text{ for E-optimality} \tag{3.4.28}$$

$$\text{where } \lambda_{max}(\underset{\sim}{N}) \text{ is the maximal eigenvalue of } \hat{V}(\hat{\underset{\sim}{\beta}})(\underset{\sim}{N}) \tag{3.4.29}$$

Nagtegaal (1987) proposes the following G- and V- criteria

$$C(\underset{\sim}{N}) = \max_{i=1}^{r} [\mathrm{tr}(\hat{V}(\hat{\underset{\sim}{u}}(\underset{\sim}{\chi}_i)))] \text{ for G-optimality} \tag{3.4.30}$$

$$C(\underset{\sim}{N}) = \frac{1}{r} \sum_{i=1}^{r} [\mathrm{tr}(\hat{V}(\hat{\underset{\sim}{u}}(\underset{\sim}{x}_i)))] \text{ for V-optimality} \qquad (3.4.31)$$

Note: when the error variance matrix S^2 is unknown, successive estimations of S^2 can be obtained by applying the iterative model building procedure of Fig. 4.1.

3.5 Model fitting and testing

3.5.1 Introduction

The construction of a reliable model for a given system or structure is an iterative process. At the start of each iteration step the number k of model parameters $\beta_1, \ldots, \beta_k$ and the presumed model functions $f_i(\underset{\sim}{x})$, i = 1, 2, ..., k must be available. The iteration step then involves the design of an experiment to collect data, estimation of the parameters from the collected data and evaluation of the model. Evaluation implies answering questions like:

- Is the model valid?
- Are the estimated parameters accurate enough?
- Is the model accurate enough for all relevant values of $\underset{\sim}{x}$?
- Which parameters are relevant and which parameters can be dropped without affecting the predictive ability of the model too much?

In the next subsection procedures for selecting regressor variables will be discussed briefly. Testing procedures will be discussed in Subsection 3.5.4.

3.5.2 Selection of regressor variables

Experimental design as described in the previous sections, is based on a postulated regression model of the type (3.4.2). The problem now is to decide which terms or "regressor variables" should be contained in a regression model for a particular response quantity. For the selection of the regressor variables several algorithms have been developed.

The backward elimination algorithm starts with the full model and subsequently removes non-significant variables in successive steps. The forward selection algorithm starts with the most significant regressor variable and adds significant variables successively.

A popular and effective algorithm is the stepwise regression algorithm, developed by Efroymson (1960). Stepwise regression can be regarded as a combination of forward selection and backward elimination. A concise description is given below.

In order to decide whether or not a certain parameter β_i and its corresponding regressor variable $f_i(\underset{\sim}{x})$ should be preserved in the model or removed from the model a criterion is needed. For this purpose an F-test quantity $\hat{F}_i$ is used, defined by:

$$\hat{F}_i = \frac{\hat{\beta}_i^2}{\hat{V}(\hat{\underset{\sim}{\beta}})_{ii}} \quad \text{for } i = 1, 2, \ldots, k, \qquad (3.5.1)$$

where $\hat{\beta}_i$ is an estimate for the parameter β_i, and $\hat{V}(\hat{\underset{\sim}{\beta}})_{ii}$ is the estimated variance of that parameter. The term $\hat{\beta}_i f_i(\underset{\sim}{x})$ is said to be more significant than $\hat{\beta}_j f_j(\underset{\sim}{x})$ if $\hat{F}_i$ is larger than $\hat{F}_j$. Using these notions the stepwise regression algorithm adds the most significant term $\hat{\beta}_j f_j(\underset{\sim}{x})$ to the model for which $\hat{F}_j$ is larger than a predetermined F-value to enter and subsequently removes term $\hat{\beta}_i f_i(\underset{\sim}{x})$ for which $\hat{F}_i$ is smaller then a predetermined F-value to remove. Details of the algorithm can be found in a paper of Efroymson (1960).

3.5.3 Parameter estimation

Methods for estimation of the unknown parameters $\underset{\sim}{\beta}$ and their variances and for estimation of the response quantities and their variancies have already been described in the Sections 3.1 and 3.4. Section 3.1 treats a single response quantity, whereas Section 3.4 treats the general case of several response quantities, some of which possibly may be response quantities and their partial derivatives with respect to the control variables $\underset{\sim}{x}$.

3.5.4 Tests for model adequacy

Several tests and criteria for a judgement of the model adequacy are available. It is emphasized that the model should be tested using newly collected data. Tests based on data which were used to derive the model can be deceptive and dangerous. From careful inspection of the residuals, i.e. the discrepancies between observed and predicted values, one can learn much about the validity of the model. Assumptions about distributions of errors and the significance of the selected regressor variables can be tested. Draper et al. (1981) give a comprehensive treatment on the subject. Plots of residuals are very illustrative; common plots include normality plots and plots of residuals against predicted values or against control variables.

The variance estimator (3.4.24) gives a measure of the accuracy of the parameter estimators $\hat{\underset{\sim}{\beta}}$. If the observations follow a normal distribution, then the parameters $\hat{\underset{\sim}{\beta}}$ are also normally distributed and confidence intervals for the parameters can be calculated using:

$$\frac{\hat{\beta}_i - \beta_i}{V(\beta_i)} \tag{3.5.2}$$

which follows an N(0, 1) distribution if σ^2 and hence $V(\beta_i)$ is known. Based on estimated parameter variances confidence intervals for the parameters can be derived from:

$$\frac{\hat{\beta}_i - \beta_i}{(\hat{V}(\hat{\underset{\sim}{\beta}})_{ii})^{1/2}} = t_i \tag{3.5.3}$$

where t_i is a random variable following a Student's t-distribution with n-k degrees of freedom.

If the normal equations are ill conditioned, the parameters will become unstable. A slight modification of the observations will result in considerable changes in parameter values. Hence, the condition number of the matrix M provides another criterion for parameter judgement.

For a judgement of the predictive ability of the model the residuals of the responses should be examined. The mean of the squared residuals provides an "overall" measure for the predictive quality. It is an estimate of the response variance. For example, the estimated response variance of $\underline{y}$ can be computed from:

$$\hat{V}(\underline{y}) = \frac{1}{t} \sum_{i=1}^{t} (\hat{y}_i - y_i)^2 \qquad (3.5.4)$$

As recommended before, here newly collected data should be used in t adequately chosen points. The new set of data points may be generated using a block (or fraction) of a 2^n-design, which has not yet been used. Experimental design methods render efficient facilities to generate a set of "optimal" testing points, by augmenting the experimental design (Nagtegaal (1987)). When added to the old data, they may result in improved parameter estimates.

If the outcomes of the model testing ask for further model improvement, it is necessary to initiate another model building cycle consisting of design, data collection, model fitting and again testing. The model building process will be considered in Chapter 4.

3.6 Experimental design of deterministic processes

The experimental design theory has been developed for the planning and evaluation of physical experiments with random errors. The consequences of applying this theory for the planning and evaluation of numerical experiments on very complicated deterministic mathematical models, for example, finite element analyses, are not trivial and will be discussed below.

Experimental design

In general the postulated linear model in an experimental design for some deterministic process is biased and, by definition, does not show a random error. In order to find an optimal design for this situation the J-criterion for robust designs could be used. However, the following approach may yield better results.

In practical situations often a realistic linear model can be postulated using experience and/or preliminary investigations. Then experimental designs can be obtained in the same way as for stochastic processes. Optimization using the common algorithms and criteria may yield designs for which the condition of the set of normal equations will be improved.

Parameter estimation

In deterministic processes estimation of the parameters $\underset{\sim}{\beta}$ can be carried out in the same way as in stochastic processes. The measures of accuracy of the estimated parameters and response quantities can be computed as in the stochastic case. However, there is a difference because the variance of a deterministic response quantity is zero. We loosely keep speaking of "variances", although such quantities are not defined in deterministic processes.

The variance $\hat{V}(\underset{\sim}{\beta})$ of the parameters $\underset{\sim}{\beta}$ is defined analogously to (3.4.18) as:

$$\hat{V}(\underset{\sim}{\hat{\beta}}) = N\, M^{-1} \tag{3.6.1}$$

The variance-covariance matrix of the deterministic response quantities, $V(\underset{\sim}{u}(\underset{\sim}{x}))$, can be computed analogously to (3.4.23) by:

$$\hat{V}(\underset{\sim}{\hat{u}}(\underset{\sim}{x})) = F^T(\underset{\sim}{x})\hat{V}(\underset{\sim}{\hat{\beta}})\, F(\underset{\sim}{x}) \tag{3.6.2}$$

where $\hat{V}(\underset{\sim}{\hat{\beta}})$ is computed from (3.6.1).

Selection of regressor variables in deterministic processes

For the selection of regressor variables F-tests are used which are analogous to (3.5.1):

$$\hat{F}_i = \frac{\hat{\beta}_i^2}{\hat{V}(\underset{\sim}{\hat{\beta}})_{ii}} \quad , \quad i = 0,\ 1,\ \ldots\ k \tag{3.6.3}$$

In physical experiments a certain parameter β_i is regarded significant ($\beta_i \neq 0$) or not ($\beta_i = 0$), depending on whether or not $\hat{F}_i$ exceeds a critical value taken from an F-table. The test is used to include the corresponding regressor variable in the model if β_i is significant, or to remove that variable if β_i is not significant.

In deterministic processes we can use this selection procedure as well; however, it cannot be stated that β_i is significant or not. In general the correct variables will be selected in the regression model, but this cannot be proved. Tests for model adequacy should be used as a final approval of the regression model.

Model testing

The test for model adequacy can be performed using the same procedures as for stochastic processes. However, confidence intervals for estimated parameters cannot be defined.

3.7 CADE, an interactive program for computer aided design and analysis of experiments

Nagtegaal (1987) developed a computer program called CADE, which stands for "Computer Aided Design of Experiments". Also facilities for the analysis of experiments have been implemented. For the experimental design part, the core of the program ACED (Welch (1985)) has been used. In CADE the optimality criteria and algorithms of ACED have been generalized to the case of simultaneous observation of several response quantities. Facilities for the selection of regressor variables have been implemented. The selection procedures are capable of handling partial derivatives by means of an iterative reweighted least squares procedure. The interactive program is menu-driven using the command interpreting program COIN (Banens (1981)). CADE contains approximately 21000 lines, has been coded in Fortran 77, and runs on an Apollo D3000 workstation.

CADE consists of three main modules, being model input, design of experiments and model fitting.

In the model input module all kinds of linear models can be entered, stored in a file or read from a previously prepared file without the need for user supplied subroutines.

The module for the design of experiments offers the following facilities:

- optimal design for a single response and for several responses is possible.
- the D-, V-, G- and J-criteria are implemented.
- the available optimization algorithms include the "Detmax-excursion-algorithm", a "Branch & Bound-algorithm", both for exact designs, and an "Approximate-algorithm" for discrete designs.
- determination of the characteristics of user-supplied experimental designs.
- augmentation of experimental designs.
- generation of (fractional) 2^n-designs.

Finally, the main characteristics of the model fitting module are:

- regressor variables can be selected by means of Efroymson's step-wise regression procedure, extended with procedures for backward elimination and forward selection. Regressor variables can also be selected "by hand".
- regressor variables can be protected against removing from the model.
- parameters are estimated accurately by means of QR-decomposition, followed by an iterative refinement procedure.

CHAPTER 4: BUILDING AND APPLICATILITY OF REGRESSION MODELS BASED ON NUMERICAL COMPUTATIONS

4.1 Building regression models

Draper et al. (1981) give guidelines for building regression models based on physical experimentation. In this section we apply these rules to regression models based on numerical computations. First we will make some remarks about linear and nonlinear regression models. Next, aspects are discussed of response quantities, control variables and feasibility, followed by aspects of the preliminary choice of a regression model and the iterative development of that model. Finally, operational aspects of regression models are discussed.

4.1.1 Linear and nonlinear regression

Regression models can be catagorized in so-called functional, control and predictive models. We will give a short explanation of the differences.

Functional models, in literature also indicated as mechanistic models (Box et al. (1978)), can be used if there is a fairly detailed knowledge, often a set of differential equations, of the system under consideration. The unknown parameters generally appear in a nonlinear way in the model, and a least squares fitting of the model requires the solution of a nonlinear set of normal equations. Such a solution may be obtained by means of mathematical programming methods. In the present research we will concentrate on the combination of structural optimization and experimental design using less complicated linear regression models.

In structural optimization we are concerned with control of the structural behaviour by means of adjusting the values of the design variables. What we need is a so-called control model, in which the design or "control" variables are under the control of the investigator. With the term "control variables" we mean "design variables" or "factors" as far as structural optimization or experimental design is concerned respectively. For certain classes of structural analysis and optimization problems useful control models can effectively be derived by regression techniques using finite element analyses and experimental designs for linear models.

In many practical situations where experimentation has to be performed on running processes, the factors are not completely under the control of the experimentalist. In such situations experiments cannot be designed, but still one can often obtain a so-called linear predictive model, which at least can reproduce the main effects of the considered response quantity. Regression techniques are much used to construct predictive models. However, since we are concerned with designed (numerical) experiments, we will not discuss this type of model any further.

4.1.2 Selection of response quantities and control variables

In constructing a regression model the first important task is to define the problem as clearly as possible. This definition should include those response quantities and control variables which play a role in the problem. The pursued regression models should be able to estimate response quantities accurately, with values for the control variables as input data.

Response quantities

With a response quantity of a system we mean some behaviour variable of the system which, in our case, can only be evaluated by means of an analysis of the system, given a proper set of design variables. In general structural analyses produce a lot of response quantities. It is important to get clear which of these quantities are needed in the model building process. The relevant response quantities are closely related to the aimed goals of the investigation, so a formulation of these goals has to be available. To obtain afterwards data for an overlooked response quantity usually requires great extra effort or may even be impossible at all.
In many cases response quantity values emerge directly from the analysis of the system, but sometimes one or more transformations are necessary to derive such a value from analysis results. In the latter situations two approaches are possible. The first is to perform those transformations first and to use the resulting response value in the model building process. In the second approach regression models are first fitted on the separate analysis results and the transformation is performed on estimates computed using these regression models. The advantage of the first approach is that finally less data (regression parameters) is involved. In the second approach the regression models may be fitted more easily in case the individual regression models are of relatively low order.

Control variables

The investigator who tries to define the control variables often ends up in conflicting situations. On the one hand he wants to take "everything" into account to be sure that nothing important is overlooked. On the other hand the regression model should be as simple as possible, resulting in less model building effort, and in less variance in the model due to not including redundant variables. Nevertheless, the investigator should start with much freedom to consider a certain variable of possible importance. The selection process can be supported by adequate computer programs, but one should never solely rely on computer outcomes. It will be wise to consult experts both on model building and on the physical system under study.

In many problems the control variables can readily be identified as a set of discrete quantities which define the system and which, of course may be of different individual importance. However, sometimes control variables are defined more complex, for instance as certain classes of functions. The geometry in shape optimization problems may be regarded as such a control variable.

It is recommended to transform both the response quantities and the control variables to a set of dimensionless quantities. This can be done by means of a dimensional analysis (Langhaar (1960), Kline (1965)). Introduction of dimensionless quantities may have the following advantages:

- The total number of response quantities and control variables can be reduced by the number of the elementary dimensions in the problem.
- Insight into the behaviour of the system may be enlarged.
- Scaling the variables to obtain a well conditioned regression problem, will be more straightforward.

4.1.3 Aspects of collecting the data

Concerning the data collection in numerical experimentation, at an early stage the following aspects have to be considered.

The investigator must build and test an efficient and accurate numerical model, often a finite element model, of the system under study. Generally the control variables for the regression model must be transformed in input data for the finite element model (FEM), and the output of the finite element analysis usually also has to be transformed in order to obtain observations for the required response quantities. Sometimes the link between a certain response quantity and the FEM-output is simple: for instance the lowest eigenfrequency of the structure, or the Von Mises stress in a certain element node. User-friendly FEM-packages have features to gather such information in a manageable file. In more complex situations, one may be obliged to write an interface program between the FEM-output file and the regression analysis program.
At this stage of the project the investigator should check whether or not appropriate hardware and software is available to perform the transformations and computations.

First check on feasibility

At this stage a rough estimate of the required budget has to be made. Till now, rather little human and computational effort have been invested, but already a considerable knowledge is gathered about the problem. Using this knowledge, a first feasibility study can be carried out. The study should result in one of the following three outcomes:

- The project is infeasible and has to be stopped.
- The project should be revised and checked again.
- The project is allowed to proceed; in this case we proceed according to the next subsection.

4.1.4 A preliminary regression model

In deriving a preliminary regression model we have the following objectives:

- Gathering more detailed information about the behaviour of the system under study; we use this information to formulate a more established experimental design.

- Making a realistic time planning of the project, and producing fairly accurate estimates of the human and computational efforts and of the budget that is required.

The preliminary computations need not necessarily be formulated as an experimental design. If very little is known about the behaviour of the system one can start with computations in which the control variables are changed one at a time. A disadvantage of this approach is that interactions between the variables will not be revealed. Even a rather simple regression model, derived from a very first formulation of an experimental design, is a more structured way to achieve the objectives mentioned above. Whatever the approach may be, it is recommended to start tentatively and to spend not more than one quarter (25%-rule of Box et al. (1978)) of the budget as estimated at the first feasibility check.

The way in which the preliminary model is derived may be essentially the same as one iteration cycle of the further development of the regression model (Fig. 4.1) and will be discussed in the next subsection.

4.1.5 Development of the regression model

The development of the preliminary model into the ultimate regression model globally proceeds according to the scheme shown in Fig. 4.1.

The preliminary regression model can be used to perform a second screening of the control variables in two ways.

First, the X^TX-matrix of the underlying experimental design can be written in the so-called correlation form (Draper et al. (1981)). Marquardt (1970) gives the rule that the diagonal elements of the inverse of this correlation matrix should be larger than 1.0 but certainly smaller than 10.0. Violation of the latter limit indicates that the current data are not appropriate to produce valuable estimates for the regression coefficients. It may be an indication to modify the underlying linear model.

Next, for every response variable the correlation with the control variables is checked. Every response should show one or two strong correlations. If this is not the case, perhaps an important variable was overlooked or the range of variation of one or more of the variables is too small, resulting in poor predictive properties of the regression model.

Once the experimental design is more established, more reliable estimates can be made for the financial budget, human effort and needs for hardware and software. The schedule can be updated and the whole project has to be submitted for approval. The possible outcomes of this second checkpoint are similar as before:

- The project is stopped.
- The project is revised and submitted again.
- We proceed the project with the further development of the regression model.

Revision of the experimental design

Using the information obtained from the preliminary or the current regression model, the experimental design can now be revised with respect to the following aspects:

- The set of relevant control variables and their ranges and the regression equations must be adapted.
- Depending on the outcomes of the previous point, for every control variable a suitable number of discrete levels and values for the levels must be chosen. These choices depend on whether or not partial derivatives of the response quantities will be used to estimate the model parameters.
- If more than one response quantity is involved, one may decide to treat these quantities one at a time and formulate experimental designs for every individual response quantity. Generally such an approach is very inefficient and one of the following two approaches is used:
  1. By inspection we formulate a linear model incorporating the expected models of the individual response quantities. This model is used to formulate the experimental design. (Of course, after the observations having been made, for every response quantity its own set of parameters is estimated.)
  This approach is feasible if the linear models for the individual response quantities do not differ too much; in this case all methods described in Chapter 3 can be used.
  2. If quite different linear models are related to the response quantities, and/or if partial derivatives of the quantities will brought into account, it is then recommended to use those methods described in Chapter 3, through which an experimental design can be optimized based on several simultaneous linear models.
- Next, the set of design points which build a complete experimental design can be formulated. Generally a complete design is infeasible because it contains far too many design points and, in the case of 2^n-designs, a fractional 2^n-design is developed. In the case of optimal experimental design the points of a complete design build a possible set of candidate points. So many candidate points, however, may be infeasible as well and we should reduce their number. A reasonable choice may be defining the set of candidate points as a fractional 2^n-design. Sometimes there are practical reasons to eliminate certain points from the set of candidate points. As an example of such a situation the reader is referred to the derivation of the general bell model in Section 6.3.
- Finally we estimate the number of design points needed in the ultimate fractional design, and we decide how many points will be used in the next experimental design. Choices depend on the method used in the experimental design. In fractional 2^n-designs one may use blocks derived from related defining contrasts. In optimal experimental design methods one is freer to choose the number of design points. Generally the points already used in the preliminary design will be maintained and the next experimental design is formulated augmenting the current one.

We proceed with short explanations of the remaining steps of the scheme in Fig. 4.1.

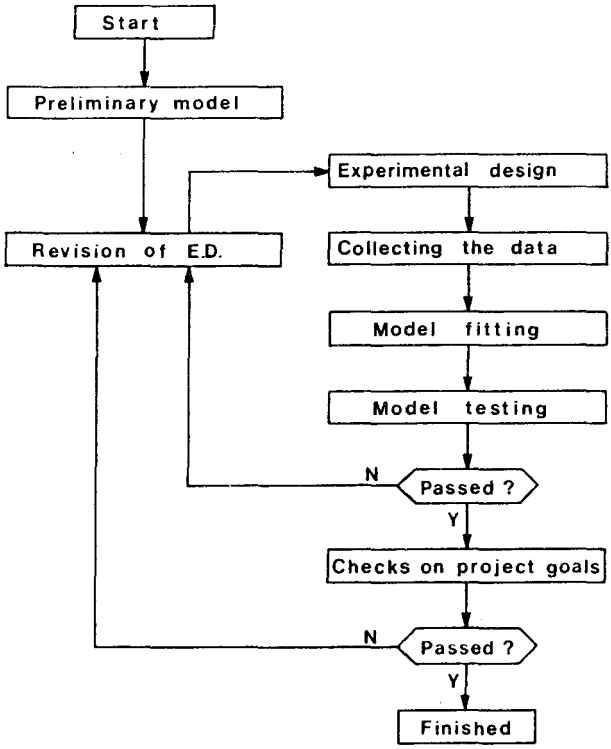


Fig. 4.1 Scheme for model building

Experimental_design

The major task in this step is to determine the set of design points. Possible choices are, for instance, a new block in a 2^n-design, or an augmentation of the current set with a rather arbitrary number of new points in the case that the optimal design theory is used.

Collecting_the_data

Usually the revision of the experimental design has no consequences for the collection of the data, and the same procedures as described in Subsection 4.1.3 will be used.

Model fitting and testing

After the data has been collected, for every response quantity a linear model must be fitted. These models must subsequently be tested on their validity and predictive ability. The tests which are relevant for our purpose are mentioned in Section 3.5. Once more we emphasize not to rely solely on computer tests; one should also check the models with common sense and consult experts on the problem under study for their appraisal.

Checks on project goals

Finally the goals and the resources of the project will be checked and we must find out whether the model building process can be successfully terminated or a next cycle will have to be initiated.

We will conclude this subsection with a typical use of the scheme in Fig. 4.1 in conjuction with optimal experimental design. If the experimental design is augmented with a certain set of design points, the optimality criterion (see Subsection 3.3.2) does not always have to remain the same. In subsequent model building cycles a reasonable alternative may be chosen, for instance

1. minimization of $\det((X^TX)^{-1})$, resulting in good parameter estimates (D-criterion).
2. minimization of the maximum response variance, resulting in good predictive ability in limited areas (G-criterion).
3. minimization of the mean response variance, resulting in good predictive ability in a wide area (V-criterion).

Neither the number of points added in the subsequent augmentations need not always be the same. If the model building process evolves, the model becomes of higher quality and the number of points in augmentations may be lowered.

4.1.6 Operating regression models

Regression models based on numerical experiments, can be used in several ways; in the next section we sum up the most important kinds of application. In this thesis we are most concerned with procedures for structured development of regression models. However, till now the use of the developed models happened merely according to "ad hoc" procedures. Such a use of regression models is not the most efficient one, and we will write down some ideas which probably will enhance future use.

After a regression model has been developed, the following data should be properly recorded:

- the coding and physical meaning of the response quantities.
- a measure for the predictive ability, for instance the variance of the response quantity.
- the coding and physical meaning of the control variables.
- the ranges of the control variables and perhaps some additional constraints. The user of the model should respect these limits as a necessary demand to avoid extrapolation using the model, probably resulting in poor behaviour of the model. Some additional comments on this subject will be made later.

- the number and types of the model functions in the regression equations.
- the estimated model parameters.

Finally it is recommended to record the set of design points on which the regression model is based. This information can be used to check whether a future design point perhaps may specify a so-called hidden extrapolation (Montgomery & Peck (1982)). Although a point lies within the ranges of the control variables, it may lie in a region of the design space which is not occupied by points used in the experimental design. Montgomery and Peck point out some procedures to check for hidden extrapolation. Such procedures should become tools in future software which supports the use of regression models.

The means by which the data of the experimental design are collected should also be recorded; for instance one can think of the used finite element package and the file containing the element model. This ensures the possibility to check future results of the regression model and to update it if this might be necessary. For this purpose it will also be useful to establish procedures to gather user's experiences.

4.2 Applicability of regression models

The main reason for the development of a regression model is the need for a fast analysis model of a certain system which can be described with not too many variables, whereas nevertheless the system is so complex and the wanted analysis accuracy so high, that this can only be achieved by elaborate numerical analysis. Formulated in this way, regression models may be developed for a very wide class of applications. However, we are concerned with structural analysis and optimization, and we will restrict the following considerations to this case and some related topics.

4.2.1 Fast analysis modules for design offices and education

Applicability

In mechanical engineering design offices a wide class of machine parts is used to build complex structures. In most cases these parts can be described by not too many (say less than 10) design variables. Textbooks on machine parts show many examples, usually together with rather simple analytical analysis models for the parts. Fig. 4.2 gives some typical examples, which are all pretty standard in mechanical engineering. A second class of interesting parts may be less general, but quite common in certain production organizations or branches; see Fig. 4.3 for some examples. A third typical class of problems is that of more or less isolated continuum mechanics problems, for instance stress concentration problems. The book of Peterson (1974) shows a lot of problems; Fig. 4.4 shows some examples. Most of the results in Peterson's book have been obtained in the past from comprehensive experimental investigations.

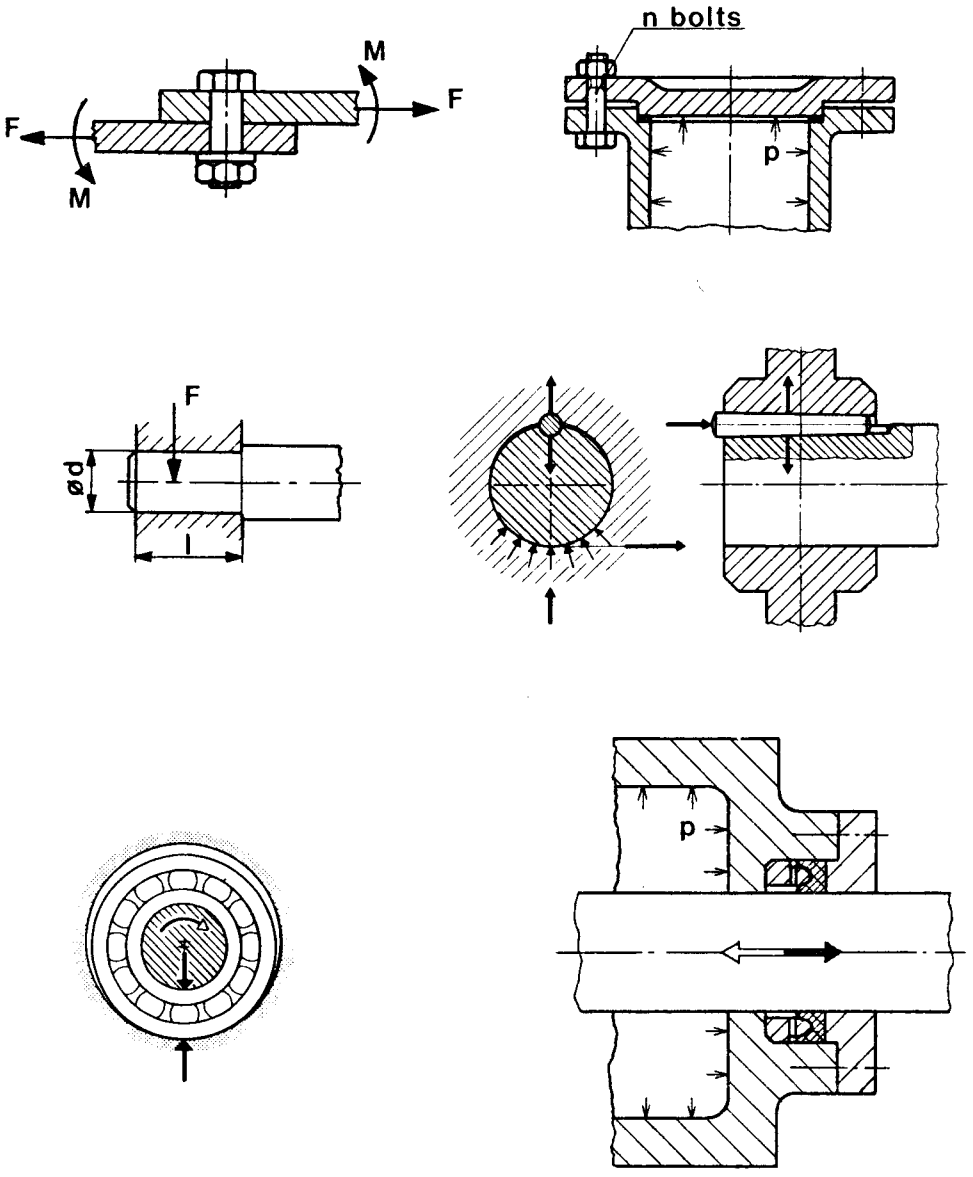
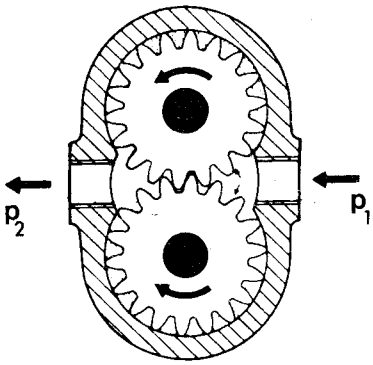
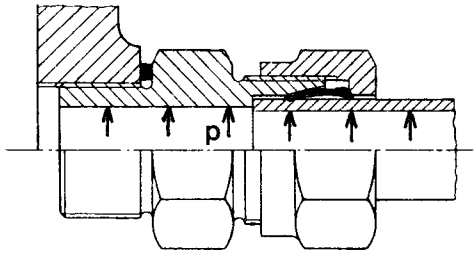
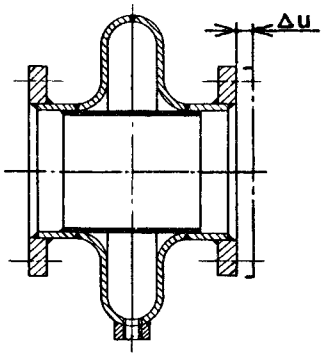
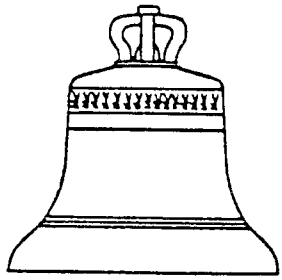


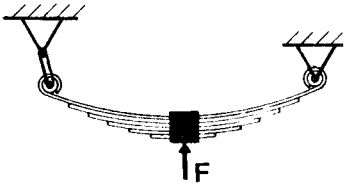
Fig. 4.2 Structural design elements

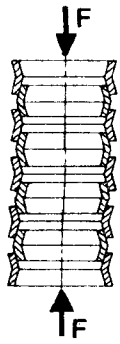




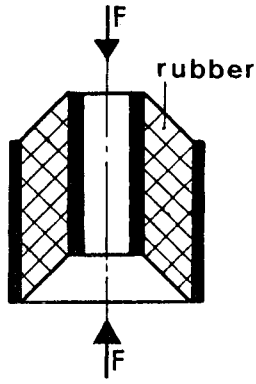






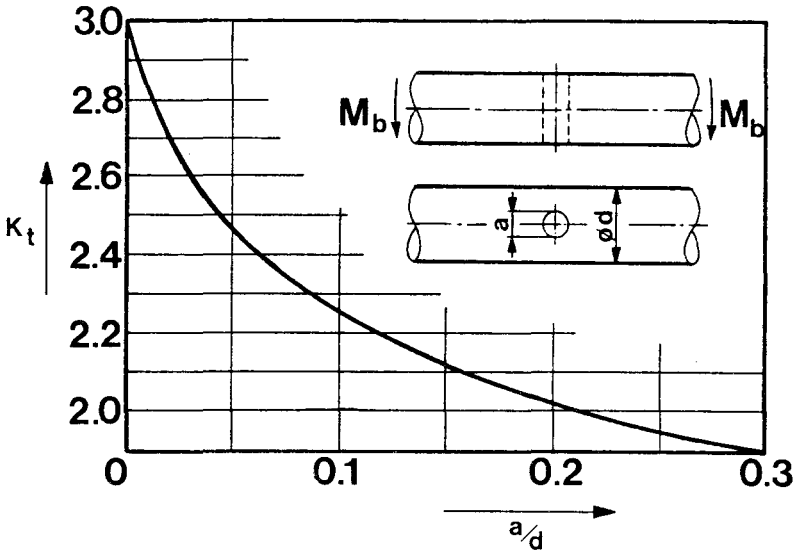


spring



vibration absorber

Fig. 4.3 Standardized special products



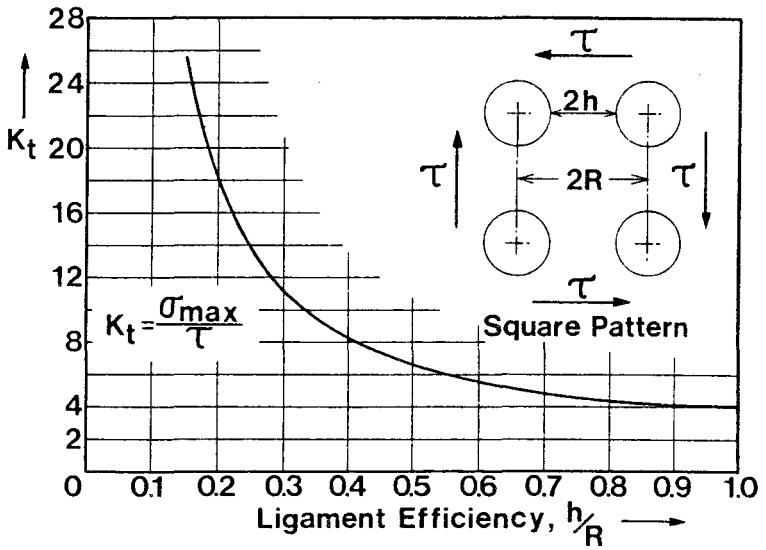


Fig. 4.4 Stress concentration problems

Nowadays the real mechanical behaviour of the machine parts mentioned before can be approximated much better using the finite element method. FEM-analysis of problems like stress concentrations demands also less effort than experimental investigations. However, even for such simple parts and problems FEM-analysis may soon become too comprehensive for ordinary use in a design office. Similar difficulties arise in a classroom where students attend computer assisted exercises to learn about the real behaviour of systems. Due to limitations on hardware and time only rather simple problems can be treated.

In situations as described above the execution of direct FEM-analyses can be troublesome and the development and use of a number of well chosen regression models may be useful. The regression models can be developed exploiting sophisticated hard- and software apart from the user's spot. For the use of regression models a personal computer or even a pocket calculator will suffice. Some other reasons whether to develop or to use regression models are:

- One has not available the hardware, software, knowledge or budget to perform comprehensive FEM-analyses.
- One tries to avoid repeated expensive FEM-analyses.
- Building a catalogue of standard regression models may be a vendible product for software houses. Collecting the FEM-data can be used to reduce idling of computer capacity.

Feasibility

Concerning the feasibility of regression models the following aspects are relevant:

1. The measure of similarity of specimens of the considered machine part or mechanical problem.
2. The demanded accuracy for the response quantities.
3. The effort for development of the regression model versus its profits.

Sub 1.

A necessary condition for the development of a regression model is a well defined and stable set of design variables and their ranges, by means of which all specimens of the structure can adequately be described. If during the development changes occur in the set of design variables, one has to decide which part of the already collected data remains valuable.

The demand for similarity cannot always be fulfilled easily, and some additional modelling may be necessary. It can also result in dependencies between design variables; we will give an example. If the elements of a certain class of similar machine parts vary much in size, very often some material properties vary together with the size. For this reason one may decide to incorporate, besides variables concerning the size, those material properties in the set of design variables. It will be clear such a set of design variables is dependent.

Sub 2.

Evaluation of a developed regression model in a certain design point will result in estimates of the response quantities, which in turn

are estimates of the real behaviour of the structure. Since the regression model is fitted on FEM-data, these estimates obviously are less accurate than those obtained directly from a FEM-analysis. However, depending on the effort that is spent on the development of the regression model, approximation of direct FEM-results may be sufficiently close.

Sub 3.

The effort for the development of a regression model is influenced by the following, certainly dependent factors:

- the number of control variables and their ranges.
- the number of parameters in the regression model and the types of the model functions.
- the number of levels for each control variable.
- the number of design points in the experimental design and the number of candidate points.
- the number of response quantities and the desired accuracy of response estimates.
- modelling and computer costs for collecting the FEM-data.
- computer costs for definition of the experimental design and model fitting and testing.

It is impossible to give general rules to derive the costs for the development of a regression model from the above factors. Our advice is to investigate a possible application carefully and to record the experiences for future learning. Following the procedures established in section 4.1 one is guided to estimate the costs of development at an early stage and this estimate is subsequently updated.

The possible profits of a developed regression model are determined by:

- the expected number of future FEM-analyses which possibly can be replaced by an evaluation of the regression model, and the savings in human effort and computer costs per analysis.
- the smaller schedule to produce relevant results for projects.
- the possible savings in hardware and software facilities.

It will also be clear that estimating these profits is not an easy task. A problem is that substantial investments for the development of a regression model precede possible profits veiled in future.

For the final decision whether or not to develop a particular regression model clearly no general rules can be given; good engineering judgment and some experience will be essential.

4.2.2 Tool for parameter study and model building

Parameter studies and model building are often based on physical experimentation and the experimental design theory is often used to structure such investigations. Nowadays physical experimentation in many cases is combined with numerical analysis (Van Heck (1984), Rousseau (1985)), thus resulting in better experiments and better analyses. With today's availability of very advanced numerical analysis packages, parameter studies in certain stages of a project may be solely based on numerical analyses. This situation is very similar to that in the previous subsection.

However, now the problem under study does not concern, for instance, a more or less standard machine part, but a unique physical problem. In such situations sometimes tremendous numerical investigations are performed, the results of which are analysed with comparatively simple methods. Planning of such parameter studies using the experimental design techniques as described in Chapter 3 will not raise the project's budget, but on the contrary, the project will only gain by doing so.

4.2.3 Use in structural optimization and simulation

At the end of Chapter 1 we already summed up some reasons for integration of experimental design into structural optimization. We repeat those reasons very shortly:

- The design variable concept appears in both disciplines.
- The need for minimizing the number of expensive elementary operations, whether it may be physical experiments or FEM-analyses.
- FEM-analyses can be regarded as numerical experiments.
- Structural optimization programs are suitable to collect data for experimental designs.
- A regression model can serve as a fast analysis model in a structural optimization program.
- Using such an optimization program an approximated global optimum can be found. It is possible to investigate several objective and constraint functions at low computing costs.
- The approximated global optimal design can be used as a starting point for a more accurate optimization, using direct FEM-analysis.

An additional, not yet mentioned reason is the use of gradients in both methods. In Chapter 2 we saw that the most effective optimization algorithms use gradients of the objective function and the constraints; in finite element formulations such gradients can be computed effectively.

In developing regression models such gradients can also be used with advantage, because in a structure with n design variables each FEM-analysis provides, per response quantity, 1 + n figures (response value + n gradient values) which can be used to estimate the parameters in the regression model. In other words: if gradients are used the fractional experimental design can be reduced globally by a factor of n.

Having discussed the potential of regression models in structural optimization, we must find out in which situations it can be used. Globally we may say that the same arguments apply as in the two preceding subsections, because use of experimental design in structural optimization starts with the development of one or more regression models. However concerning optimization some additional remarks can be made.

One remark refers to the extent of the experimental design in relation to the number of direct FEM-analyses used in iterative optimizations. In a problem with n design variables a lower bound to produce a local optimum, using an efficient optimization algorithm,

is about n FEM-analyses. If the problem is not too simple at least 5 to 10 different starting points must be used to allow a guess for the global optimum point. These countings apply for one optimization problem. If, for instance, several objective functions must be investigated, a multiple of the above mentioned 5 n to 10 n FEM-analyses will be necessary. Such figures may give an idea about a reasonable number of FEM-analyses planned to develop a regression model.

In the weighting process whether to use direct FEM-analyses or a regression model in a certain optimization problem, the following comparative evaluation of the features and drawbacks of both approaches may be helpful.

We consider the following two cases:

1. Iterative optimization based on direct FEM-analyses.
2. Development and use of regression models as an analysis module in iterative optimization processes.

We evaluate a number of aspects for these cases.

Number_of design_variables

Sub 1. Number may be large (50 up to several hundreds). Nevertheless one should try to reduce the number; design variable linking can be used for that purpose.

Sub 2. Number is rather small; maximum is about 10 if the number of levels per variable is larger than 1. If the used model is a polynomial of first order and if gradients of the response can be used, then 1 design variable level will be sufficient, and a larger number of design variables can be used (see the general bell model in Section 6.3).

Gain_of insight into_the_problem

Sub 1. Rather little insight is gained; this insight is restricted to a small area along the trajectory of the iteration steps.

Sub 2. Insight can rather easily be obtained over the full extent of the design variable space.

Accuracy_of solutions

Sub 1. Possible high accuracy which can be controlled by mesh refinement.

Sub 2. Lower accuracy, because regression models approximate a direct FEM-analyses; the element meshes are chosen in advance and usually are relatively coarse.

Flexibility of_problem_solution

Sub 1. With respect to solving the mechanical problem the flexibility is high; in the last extremity one can modify the FEM-model during every iteration step. Flexibility is low concerning the optimization problem, because modification of the objective and/or the constraint functions will require large additional computations.

Sub 2. With respect to solving the mechanical problem the flexibility is low, because the regression model is based on a certain

fixed set of design variables and a fixed design variable space. Flexibility is high concerning the optimization problem, because several objective and constraint functions can be investigated at low computer costs.

Value of the results of a FEM-analysis

Sub 1. The FEM-analyses usually are only intermediate results in the iteration process.

Sub 2. The FEM-analyses have enduring value due to their incorporation in the regression model.

Initial effort in problem solution

Sub 1. Little initial effort is required, because iterative optimization runs can be controlled well and the number of FEM-analyses can be dosed very well.

Sub 2. Great initial effort is required, because the solution procedure starts with the development of the regression model, usually requiring a fairly large number of FEM-analyses.

Again, to make the right choices, one will require good engineering judgement and experience.

Use of regression models in simulation programs

We have no experience with such applications of a regression model and we will confine ourselves to an example to indicate a possible type of use.

Consider the problem of how to steer a ship in a certain schematic flow field, see Fig. 4.5

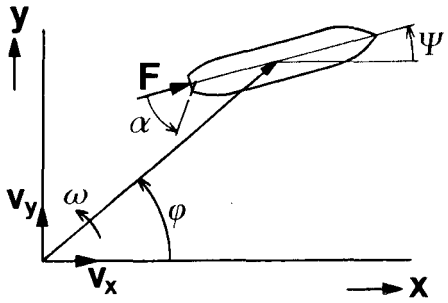


Fig. 4.5 Ship in schematic flow field

To develop a regression model which can be used in a numerical simulation of the steering problem, the following state variables can be regarded as "control variables":

v_x, v_y, ω : the translational and rotational velocities in the schematic flow field
r, φ, ψ : the position of the ship in the flow field
$\dot{r}$, $\dot{\varphi}$, $\dot{\psi}$: the velocity components of the ship
F : the thrust of the ship's propeller
α : the rudder angle
M, I_z : mass and moment of inertia about the z-axis of the ship.

The resulting forces and moments due to the water flow acting on the ship can be calculated using the finite element or the finite difference method. Such computations will be too time consuming for direct use in a real time simulation program which runs on a computer of modest capacity. A possible solution may be found in the development of a regression model in which the 13 state variables mentioned above each are varied on an appropriate number of levels. The resulting regression model may be fast enough for real time simulation of the steering process.

CHAPTER 5: SOME APPLICATIONS ON STRUCTURAL DESIGN ELEMENTS AND A PARAMETER STUDY

The procedures described in the preceding chapters have been applied to several practical problems. In this chapter four applications are presented. The regression models described here have been developed without using partial derivatives of the response quantities. Considered aspects of each application include problem description, experimental design techniques, data collection and results.

5.1 Two-dimensional pin and hole joint

The development of procedures as described in this thesis, have already been initiated in 1978. During his Master's-Degree study Aerts (1979) worked out test-cases concerning two types of pin and hole joints for hydraulic cylinders.

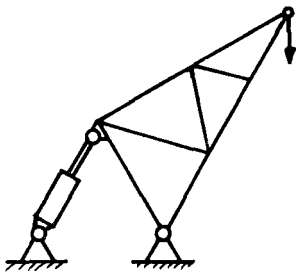


Fig. 5.1 Application of a hydraulic cylinder

In general, hydraulic cylinders are heavily loaded structures. Fig. 5.1 shows a typical application: the cylinder is linked to surrounding structures by means of pin and hole joints. Fig. 5.2 shows the two types of pin and hole joints investigated. The left one incorporates a ball bearing and is used in situations where the cylinder cannot properly be aligned. Here we will only describe the simpler joint shown in Fig. 5.2b. In this figure also geometrical parameters are indicated which possibly may serve as control variables. The objective of the investigations is to describe the deformations and stresses of the eye of the joint under statical loads. For this purpose the eye will be modelled as a two-dimensional plain stress problem. The cross-sections A-A and B-B in Fig. 5.2b are critical with respect to stresses. In Fig. 5.3 the points are indicated in which response quantities will be considered. These response quantities include the displacements of these points in x- and y-directions and furthermore the orthogonal stress components σ_x,

σ_y and τ_{xy} and the Von Mises stress σ_{id}. Fig. 5.4 shows the mechanical model and a sample of the element meshes.

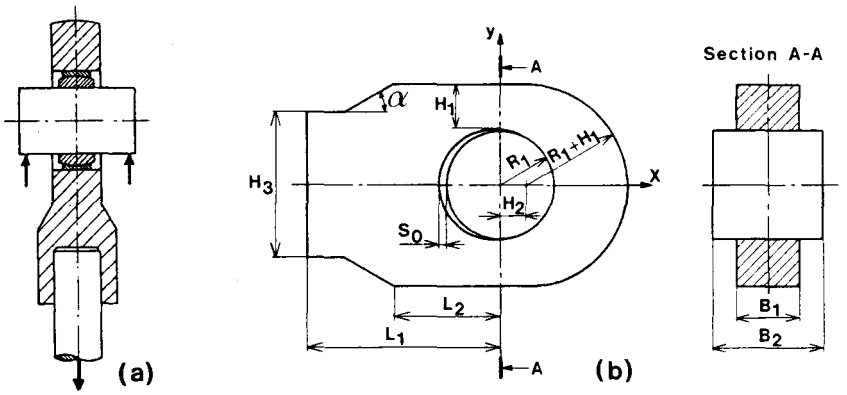


Fig. 5.2 Two types of pin and hole joints (Aerts, 1979)

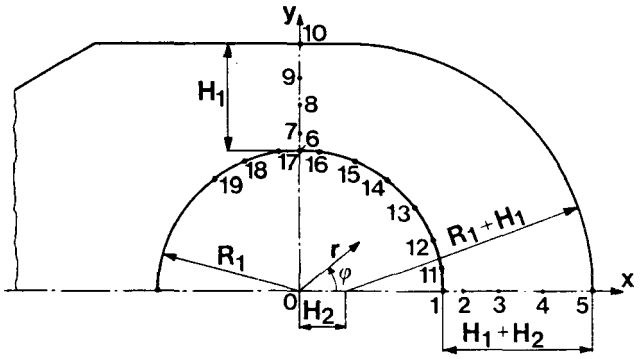


Fig. 5.3 Definition points for response quantities

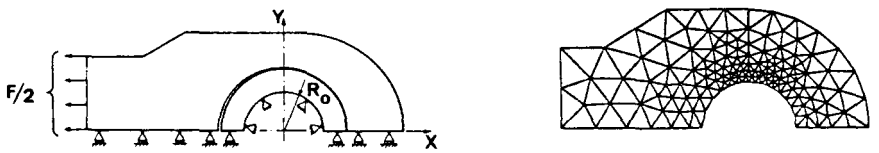


Fig. 5.4 Mechanical model and element mesh

The data have been collected using the mesh generating program TRIQUAMESH (Schoofs et al. (1978)) and the finite element program FEMSYS (Banens et al. (1976)).

Using FEMSYS, 2-dimensional contact problems with friction can be modelled easily by means of user-defined dependencies between nodal degrees of freedom (Van de Boom (1979)).

After some explorative calculations and using a dimensional analysis, the following five dimensionless control variables were formulated (see Fig. 5.2b):

$$\frac{H_1}{R_1} \quad , \quad \frac{H_2}{R_1} \quad , \quad \frac{S_0}{R_1} \quad , \quad f \quad , \quad \frac{P}{E_1 B_1 R_1} \qquad (5.1.1)$$

where P is the load of the joint, f is the coefficient of friction between pin and eye, and E_1 is Young's modulus of the eye.

As a first attempt, for all response quantities regression equations were assumed including first-order main effects and all first-order interaction terms. As experimental design a half fraction of a 2^5-design was used. After having estimated the parameters and tested the model, it was shown that the influence of the friction between pin and eye could be neglected. Furthermore it was shown that the first order models could not adequately describe the response quantities.

A second attempt was made based on the same control variables as in (5.1.1), except f. Second order main effects and all two-factor first-order interaction terms were added to the regression equations. Each of the control variables was varied on three levels and an orthogonal, one third fraction of a 3^4-design was used (Table 5.1).

		-1			0			1			X3
		-1	0	1	-1	0	1	-1	0	1	X4
-1	-1	1*	2	3	4	5	6*	7	8*	9	
	0	10	11*	12	13*	14	15	16	17	18*	
	1	19	20	21*	22	23*	24	25*	26	27	
0	-1	28	29*	30	31*	32	33	34	35	36*	
	0	37	38	39*	40	41*	42	43*	44	45	
	1	46*	47	48	49	50	51*	52	53*	54	
1	-1	55	56	57*	58	59*	60	61*	62	63	
	0	64*	65	66	67	68	69*	70	71*	72	
	1	73	74*	75	76*	77	78	79	80	81*	
X1	X2										

Table 5.1 A one third fraction of a 3^4-design indicated by *

Using the models, the displacements of the points indicated in Fig. 5.3 could be predicted within 1% from results obtained from direct finite element analyses. The Von Mises stresses could be predicted within 2.5%, whereas individual stress components showed deviations not exceeding 5%. The resulting regression models thus proved to be quite accurate.

5.2 Three-dimensional bearing problem

Dry running journal bearings are important connecting elements in mechanical engineering. Many practical bearing structures can be modelled according to Fig. 5.5. Here, the characteristic difference with the joint treated in Section 5.1, is the need for a three-dimensional model due to the inclination of the shaft.

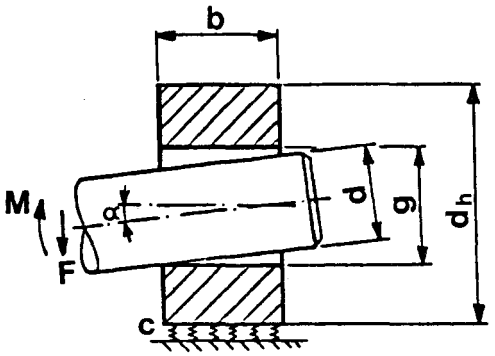


Fig. 5.5 Modelling of journal bearing (Wouters, 1986)

Wouters (1986) developed for this situation regression models describing the surface pressure in the contact zone between shaft and journal as a function of a number of control variables. Initially considered control variables are indicated in Fig. 5.5. After explorative calculations it was decided to develop regression models approximating the following dimensionless relations:

$$\frac{\sigma_v(i)}{E_1} = f\left(\frac{s}{d}\ ,\ \alpha\ ,\ \frac{F}{E_1 d^2}\right) \quad ,\quad i = 1,\ 2,\ \ldots \qquad (5.2.1)$$

where $\sigma_v(i)$ is the contact pressure in a discrete point, i, of the contact zone; $s = g - d$ is the clearance between shaft and journal.

From the explorative calculations it appeared that the regression equations should contain third-order main effects. All first-order interaction terms and a part of the second-order ones were inserted into the models.

The control variables were all varied on four levels (the minimum number for third order models without using partial derivatives of the response quantity). It was decided to carry out a complete 4^3-experiment, which was transformed to a complete 2^6- one. The required finite element analyses of 64 different contact problems were carried out using the I-DEAS-package of SDRC (1986). Fig. 5.6 shows a sample of the used three-dimensional element meshes.

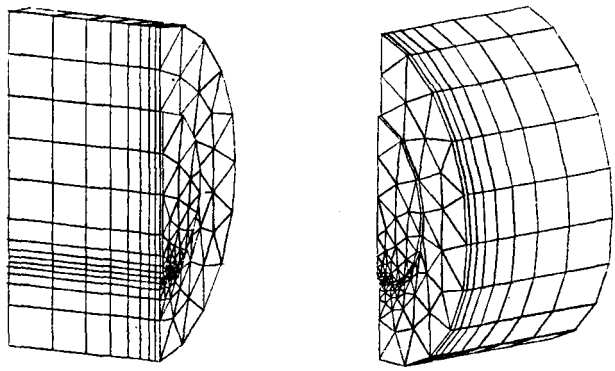


Fig. 5.6 3-D element mesh

Each analysis required about 2.5 hours of computing time on a VAX-11/750 computer.

Fig. 5.7 represents qualitative pictures of the contact pressure computed from the derived regression models for two different inclination angles of the shaft.

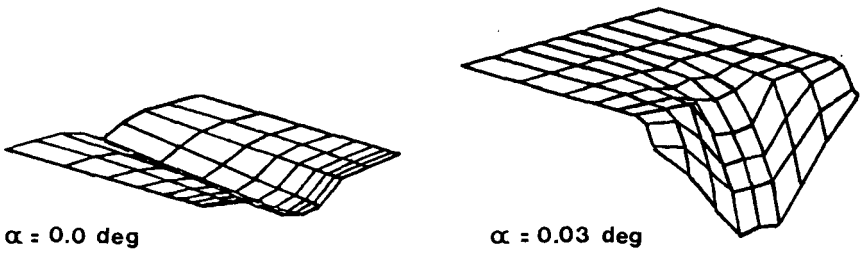


Fig. 5.7 Qualitative pictures of contact pressure

From Fig. 5.8 it can be seen that contact pressures computed from the regression models agree very well with results from direct finite element analyses.

Furthermore, Wouters (1986) made the following interesting comparison between the use of regression models and a direct finite element analysis.

From one finite element analysis using a mesh of 1000 nodal points and 1290 6-node elements the following results emerge:

number of displacements	: 1000 * 3	= 3000
number of nodal point forces	: 1000 * 3	= 3000
number of stress components	: 1290 * 6 * 6	= 46440
total number of response quantities		52440

The computing time for one finite element analysis of about 2.5 hours, results in an average computing time of 0.17 sec for one

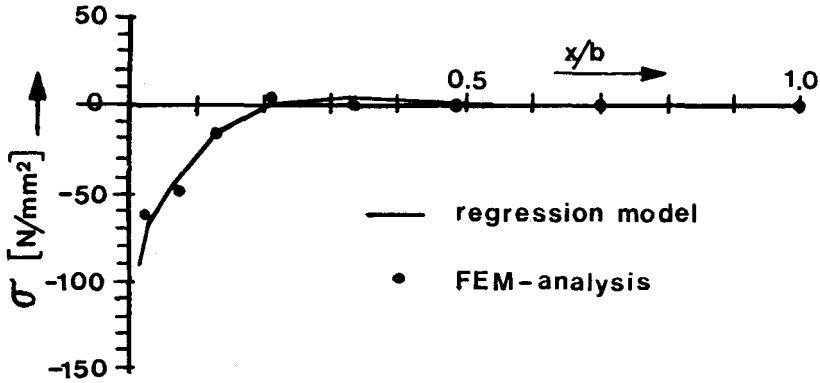


Fig. 5.8 Contact pressure from regression models versus direct finite element analysis

response quantity. The computing time required for one evaluation of a regression model is approximately 0.01 sec on the same computer. So, if regression models were available for all response quantities defined in a particular finite element model, evaluation of all those models would be a factor of 17 faster compared to a direct finite element analysis. This result may be of interest for the derivation of regression models for use in real time computer simulations, especially if a relatively small number of response quantities is relevant.

5.3 Cross-section design of aluminium beams

Extruded beams may have cross-sections with very complicated shapes; as a consequence, cross-sectional properties related to stiffness, strength and stability cannot be calculated analytically, nor can they be obtained from handbooks or from the manufacturer. Properties such as the cross-sectional area, the centroidal axes and the moments of inertia about those axes can be calculated straightforwardly using an appropriate mesh defined on the cross-section. For the calculation of some other cross-sectional properties, such as the co-ordinates of the shear centre, the warping constant and the torsion constant, the so-called torsion function, ψ, must be known (Menken et al. (1986)). This function is determined by the two-dimensional Laplace's equation $\Delta\psi = 0$. The numerical solution of this equation can be obtained using a finite element formulation (Davies (1980)). Once the torsion function is known, the cross-sectional properties can be determined straightforwardly.

For the solution of the equation $\Delta\psi = 0$ Van de Pasch (1985) wrote the finite element program GEOG8, incorporating an 8-node

isoparametric element. Using this program and some experimental design and optimization techniques, he developed a procedure for a more or less automatic optimization of extruded beams for typical applications. We will illustrate that procedure by means of the following case-study.

For the construction of greenhouses often extruded aluminium beams are used. The behaviour constraints of such beams comprise the maximum allowable deflection and the minimal wanted load carrying capacity. Usually, the beams fail due to lateral-torsional buckling. Fig. 5.9 shows a typical loading case and the cross-section of a particular beam. The cross-section is fixed, except the measures x_1 through x_4, which are used as design variables to minimize the weight of the beam. The applied element mesh is drawn in the cross-section.

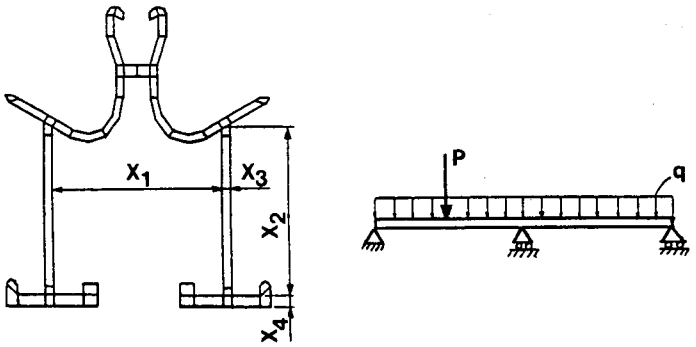


Fig. 5.9 Cross-section and loading case of aluminium beam (Van de Pasch, 1985)

The procedure starts with the generation of an experimental design. For the description of each of the response quantities, i.e. the cross-sectional properties and the behaviour constraints, standard a third order regression equation is assumed. The user has to indicate how many and which interaction terms should be used. For each design variable the range of variation and the number of equidistant levels are given. This information defines a complete experimental design, which will be used as a set of candidate points in the actual experimental design. The design is generated as follows. The user specifies the number of observations, N, (finite element analyses) he wants to spend in the investigation. This number should be larger than 1.5 to 2.0 times the number of parameters in the assumed regression equation. The N design points are randomly chosen from the candidate points; the only criterion used in this selection is preserving the matrix X^TX from becoming singular.

Using this N-point experimental design the relevant cross-sectional properties are calculated by means of the program GEOG8,

mentioned before. For that purpose the program first generates an element mesh defined by a particular set of design variables. Next, another program, the program KIP, is used to calculate the buckling load and the maximum deflection of the beam (behaviour constraints).

Subsequently, for all response quantities the regressor variables are selected using a backward elimination procedure, and the parameters are estimated by means of the regression analysis program XPD, developed by Van Heck (1984).

Applying the derived regression models, a simple zero-order optimization could be carried out, resulting in a minimum weight beam satisfying the behaviour constraints. A conventionally, but carefully designed beam served as an initial design. Through the optimization the weight could be lowered by an amount of 6%, which profit may be considered important in view of the large scale on which these beams are being used.

5.4 Parameter study of a leaflet heart valve prosthesis

Presently used heart valve prostheses can be distinguished in mechanical and leaflet prostheses. In mechanical prostheses a ball, a disc or a rigid leaflet regulates the blood flow. These valves are mechanically strong and durable. However, their flow properties are bad and they cause damage to red blood cells. Leaflet valves resemble the human aortic valve. They have good flow properties and do not cause blood damages. Their main disadvantage is the relatively short life time due to tissue failure.

Leaflet valve prostheses can be distinguished in biological and artificial ones. In the biological valves the leaflets are prepared from porcine aortic valves. In the artificial valves the leaflets are made of synthetic materials. The artificial valves however are not yet clinically available.

In his Ph-D thesis Rousseau (1985) investigated artificial leaflet valves and formulated mechanical specifications for the design of such valves. As part of the investigations a numerical parameter study was carried out using experimental design techniques.

Fig. 5.10 gives a schematic view of the entire leaflet valve. Since it is assumed that the valve consists of three cyclically symmetric parts, the mechanical model can be restricted to one sixth of the valve comprising a half leaflet (Fig. 5.11). The applied element mesh is given in Fig. 5.12.

The valve is considered in the closed situation, where the leaflets are loaded by a blood pressure difference of 12 kPa, which is realized in 10 msec. The frame of the valve is made of linear, elastic material; the membranes and fibre reinforcements show visco-elastic material properties.

The analyses of the valve have been carried out using the nonlinear finite element package MARC (1984). After having

investigated an existing biological leaflet valve and a number of

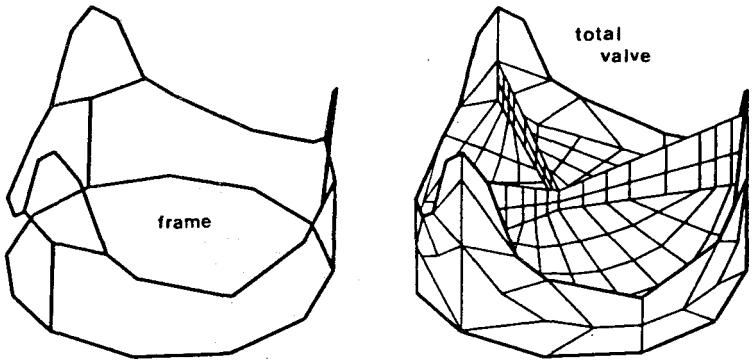


Fig. 5.10 Model of leaflet heart valve prosthesis (Rousseau, 1985)

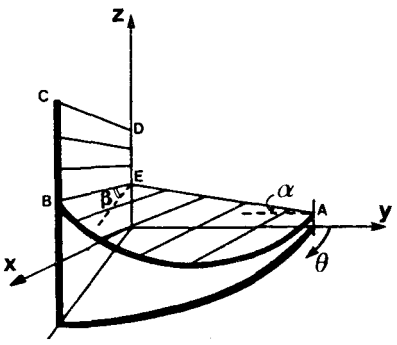


Fig. 5.11 One sixth of valve

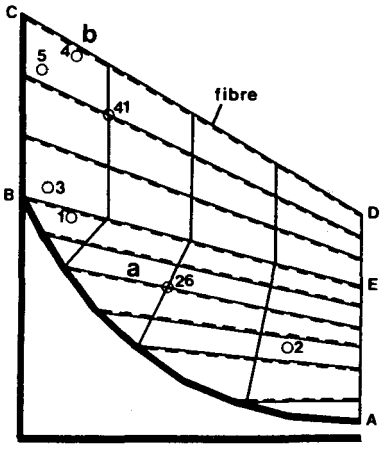


Fig. 5.12 Applied element mesh

explorative analyses of the synthetic valve, the following control variables were selected to be varied

- d_{FR} : the thickness of the frame
- α : the angle describing the free leaflet geometry
- d_f : the fibre thickness
- E_f : the elastic modulus of the fibre material

The following response quantities have been considered (see Fig. 5.12):

- the Von Mises stress in the points 1 through 4
- the stresses in the fibres a and b
- the minimum principal stress in the points 3 and 5

- the shear force per unit length in the points 26 and 41.

For all response quantities second order regression equations including all two-factor interaction terms have been assumed. All control variables have been varied on three levels. As experimental design a one third fraction of a 3^4-design has been used, requiring 27 nonlinear finite element analyses of the valve. The model parameters for the different response quantities have been estimated by exploiting the regression program XPD.

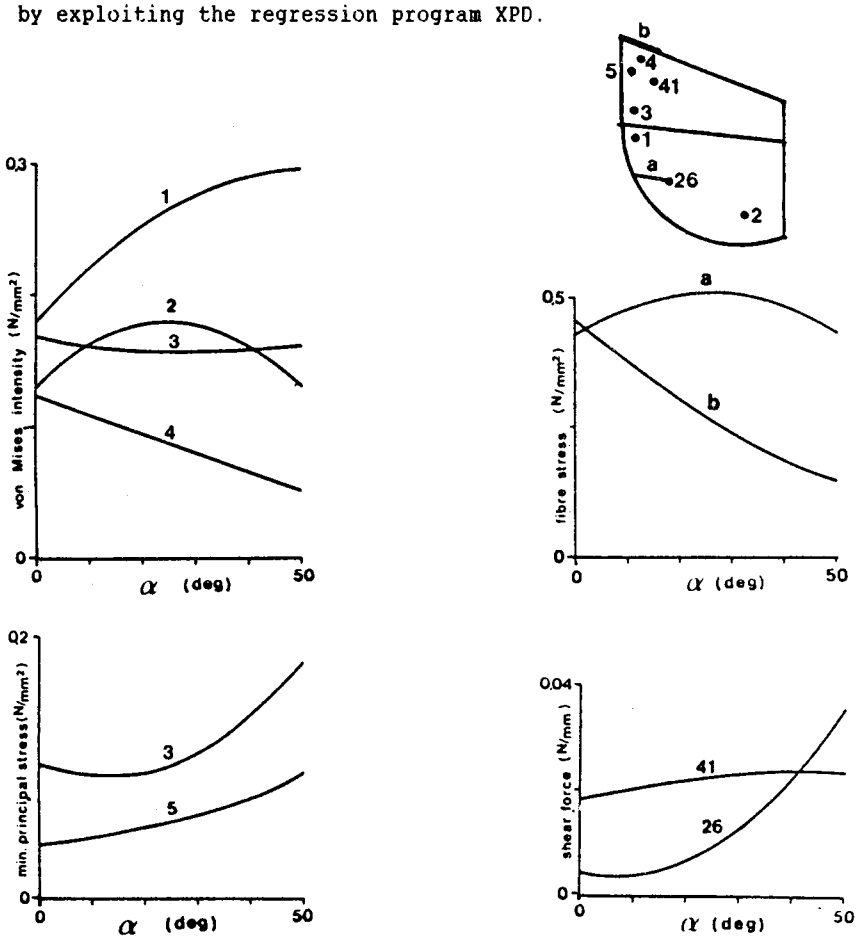


Fig. 5.13 Influence of the angle α on stress components

Applying the derived regression models the behaviour of the valve could be visualized by means of a number of graphs. Fig. 5.13 represents some results obtained from the regression models. Such results have been used to formulate design specifications for a new artificial leaflet valve.

CHAPTER 6: SHAPE OPTIMIZATION OF CHURCH- AND CARILLON BELLS

6.1 Introduction

6.1.1 History

During the summer of 1982 the research staff of the Division of Mechanical Engineering Fundamentals of the Eindhoven University of Technology visited the Royal Eijsbouts Bell-Foundry at Asten, The Netherlands. The visit was not the first contact between people of that division and the bell-foundry since already ten years earlier the master's-degree thesis of Banens (1972) was devoted to the analysis of church bells. During this visit problems concerning the design of a so-called major-third bell were discussed. The five lowest eigenfrequencies of a conventional minor-third bell have approximate ratios of 1 : 2 : 2.4 : 3 : 4 (or in tones based on c: c - c^1 - es^1 - g^1 - c^2), whereas a major-third bell should show approximately the ratios 1 : 2 : 2.5 : 3 : 4 (in tones: c - c^1 - e^1 - g^1 - c^2). The problem is to raise the ratio of the third eigenfrequency from 2.4 to 2.5. The slight change in ratio must be achieved by means of modification of the geometry of the bell. At first sight this does not seem to be a difficult problem. However, since the beginning of this century several experienced bell-founders have tried to find an appropriate bell geometry using a trial and error approach. They did not succeed, in spite of considerable experimental effort. Apparently the problem was far from simple.

We decided to adopt this challenging design problem, which presented to us a test-case for the structural optimization software, that was being developed in our division. On the one hand a bell has a simple symmetrical geometry in its horizontal cross-section, but on the other the geometry of the vertical cross-section is rather complicated and requires the finite element method for analysis of the bell. Furthermore, the dynamic problem is rather simple because the bell is a linear, elastic and almost undamped structure submitted to free vibrations, thus providing us the opportunity to concentrate ourselves on the optimization problem.

Bell-founding in all its aspects has been described extensively by Lehr in a large number of papers. Recently Lehr (1987a) presented a comprehensive paper especially devoted to bell design in the past and nowadays.

The Netherlands and Flanders have the oldest tradition in founding West European carillon bells with well defined pitches. Already in 1644 the brothers François and Pieter Hemony, advised by Jacob van Eyck, realized the first well-tuned carillon, being an important landmark in the history of music. In the 17th and 18th centuries beautiful carillons were founded, but later-on the necessary knowledge and principles of founding pure bells got lost. They were not rediscovered until the end of the 19th century in England, this being the start of the English carillon tradition.

Founding purely ringing bells has always remained a speciality. Even at this moment bell-founders who completely understand the art can be counted on the fingers of one hand.

In the past the contacts between scientists and bell-founders have always been somewhat poor. Bell-founders were not inclined to talk about their art and knowledge whereas most scientists did not listen to bells and studied the bell mostly from theoretical viewpoint. The bell behaviour proved to be so complicated that the scientists were not able to describe it adequately. This despite the famous names who paid attention to vibrating bells. Leonard Euler (1764) presented the first scientific paper on the vibration modes of rings. His discoveries are valuable, even to-day, for they led to the development of finite ring elements based on partial discretizations (Zienkiewicz (1977)). Other famous scientists who offered contributions to the bell study were Ernst Chladni (1756-1827) and Hermann Helmholtz (1821-1894) in Germany, and Lord Rayleigh (1842-1919) in England. Further important contributions came from Johannes Blessing (± 1890) in Germany, and from Abraham Vas Nunes in the Netherlands, who wrote a thesis on the subject in 1909.

Although these scientists could not provide an analysis method which was sufficiently accurate for bell design, their discoveries were important for a better understanding of the physical problem. We had to wait until the computer age before bells could be analysed adequately. Banens (1972) was the first who realized this. He wrote a finite element program in which ring elements were implemented using the circumferential vibration modes which were already described by Euler. Later on Perrin et al. (1983) also used the finite element method for frequency analysis of bells.

The introduction of structural optimization enabled us to reach another landmark in the history of bell design. Structural optimization offers a completely new and flexible way of bell designing. The design problem of the major-third bell could be solved (Van Asperen (1984), Maas (1985), Schoofs (1985)) and will be described in detail in Section 6.2. In Section 6.3 some other recently designed bells will be presented. We are convinced that in the near future application of the developed methods will lead to the discovery of other bells.

6.1.2 The overtone_structure_of the bell

Central in the campanology, the science of bells, is the investigation of the relationship between the profile of the bell, i.e. its half vertical cross-section, and the partials of the bell, i.e. its hum note and the overtones. Fig. 6.1 shows a bell profile; commonly used geometrical terminology in bell design has been included. The partials describe the behaviour of the bell from a musical point of view whereas vibration modes describe the mechanical behaviour. Every partial is associated with a unique vibration mode. The characteristics of the partials, including frequencies, sound volume and decay time, together determine the quality of the bell sound. The diagram of Fig. 6.2 illustrates the lowest and most important partials of a minor-third bell. It should be noted that the

bell structure is a continuum having an infinite number of partials. However, the lowest partials (lowest in frequency) are most important for the bell sound. Only a limited number of partials, say less than 10, is under the control of the skilled bell-founder.

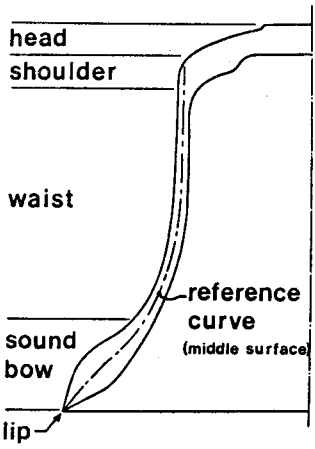


Fig. 6.1 Profile of a minor-third bell

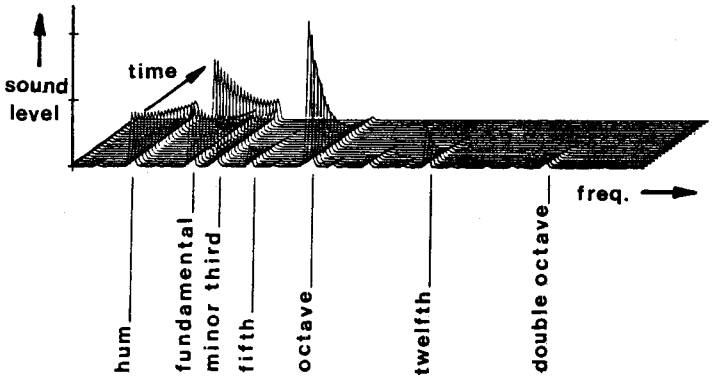


Fig. 6.2 Spectrum of partials of a minor-third bell

In this thesis we will restrict discussions to the frequencies and vibration modes of bells. Not because the volumes and decay times would be unimportant, but because the first condition for an acceptable bell sound is that the bell will ring the wanted chord constituted by the hum note and the important overtones. In addition, leaving loudness and decay time out of discussion will simplify the problem substantially, because acoustic phenomena need not be considered.

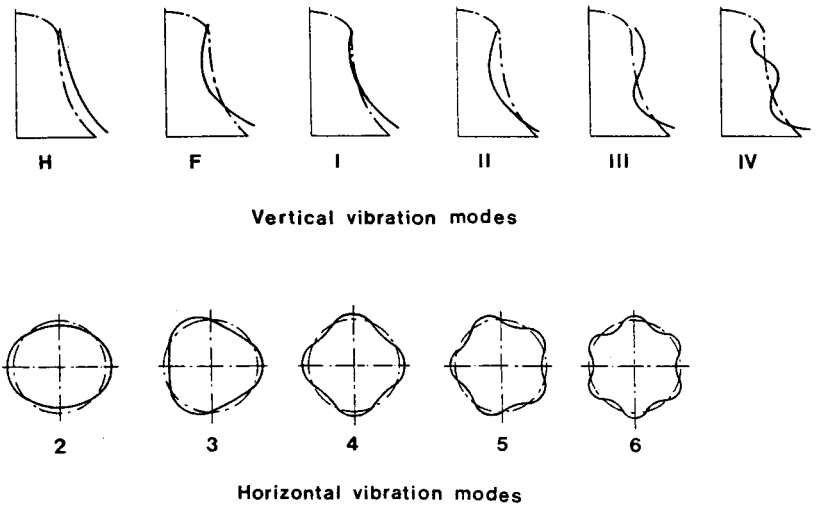


Fig. 6.3 (a) Vertical and (b) horizontal vibration modes

Fig. 6.3 shows qualitative pictures of the vibration modes of the bell in vertical and horizontal cross-sections. Each of the bell's vibration modes is a combination of one mode in the vertical and one mode in the horizontal cross-section, indicated by the partial code. Table 6.1 shows the most important frequencies of the minor-third bell. The column under "Frequency code" will be discussed later. In bell design the octave is commonly used as reference for the frequency ratios because this partial is the most important one in the perception of the bell's pitch.

The cent is a measure for music intervals where, by definition, one octave is equal to 1200 cents. The accuracy of the human hearing is about 5 cents. Cents values are computed from frequency ratios using the relation

$$c = \frac{1200 \log(4\ f/f_{oc})}{\log 2} \quad \text{[cents]} \qquad (6.1.1)$$

where f_{oc} is the frequency of the octave and f is the considered frequency. If the cents value is given, the frequency ratio $4f/f_{oc}$ can be calculated using the inverse relation

$$4\,\frac{f}{f_{oc}} = 2^{\frac{c}{1200}} \qquad (6.1.2)$$

Musical name	Partial code	Frequency code	Frequency ratios for swinging bells		Frequency ratios for carillon bells	
			$4f/f_{oc}$ [-]	cents	$4f/f_{oc}$ [-]	cents
Hum note	H - 2	2 - 1	1.0000	0	1.0000	0
Fundamental	F - 2	2 - 2	2.0000	1200	2.0000	1200
Minor third	I - 3	3 - 1	2.4000	1516	2.3784	1500
Fifth	II - 3	3 - 2	3.0000	1902	2.9966	1900
Octave	1 - 4	4 - 1	4.0000	2400	4.0000	2400
.						
Twelfth	I - 5	5 - 1	6.0000	3102	5.9932	3100
.						
Double octave	I - 6	6 - 1	8.0000	3600	8.0000	3600

Table 6.1 The most important frequencies of a minor-third bell.

As can be seen from Table 6.1 there is a difference in the tuning of swinging bells and carillon bells. The tuning according to the left column is the most harmonic and globally this tuning was used in the Renaissance; it is still in use for the tuning of solely ringing swinging bells. In the Western culture music instruments have been tuned according to the right column, which is common in music since the 18th century. Only in this tuning the musician can play in all keys. In the following we will only consider the frequency ratios as given for carillon bells.

After a bell has been founded, the five lowest partials must be tuned according to Table 6.1. The tuner uses a bell lathe (Fig. 6.4), and he consults tuning curves such as those shown in Fig. 6.5 to decide where metal should be removed from the inner surface.

The vertical and horizontal vibration modes have been used to categorize the partials. The partials showing the same vertical vibration mode belong to one group. The elements in a group are indicated by the number of meridian nodes, which is determined by the horizontal vibration mode. Lehr (1986) has measured a large number of partials of a tuned minor-third bell. Plotting the frequencies on a log-scale against the number of meridians, reveals the coupling between elements of a certain group (Fig. 6.6). From this plot it can be seen that the hum note (H-2) and the fundamental (F-2) can be regarded as degenerated elements of the groups I and II respectively.

The frequency ratios of partials belonging to the same group prove to be very strongly coupled. Ratios of a whole group with respect to another group can be changed much easier than mutual ratios within a group. In this respect the hum and the fundamental must be regarded as separate groups. Group I contains the important partials I-3, I-4, I-5 and I-6. Especially the strong coupling in this group makes realization of, for instance, the major-third bell so difficult.

Fig. 6.4 Tuning a large bell (Courtesy of Royal Bell-Foundry Eijsbouts, Asten, The Netherlands)

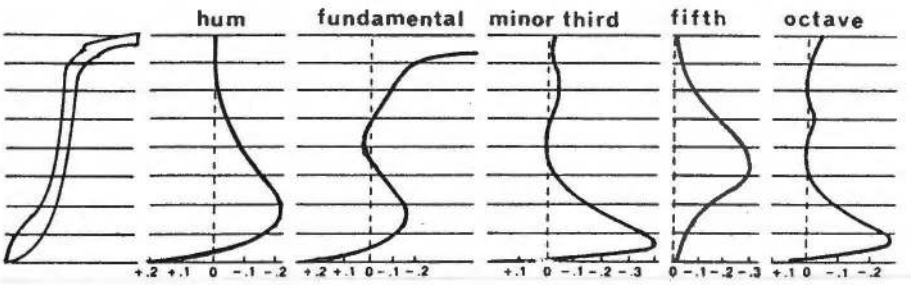


Fig. 6.5 Tuning curves for a minor-third bell

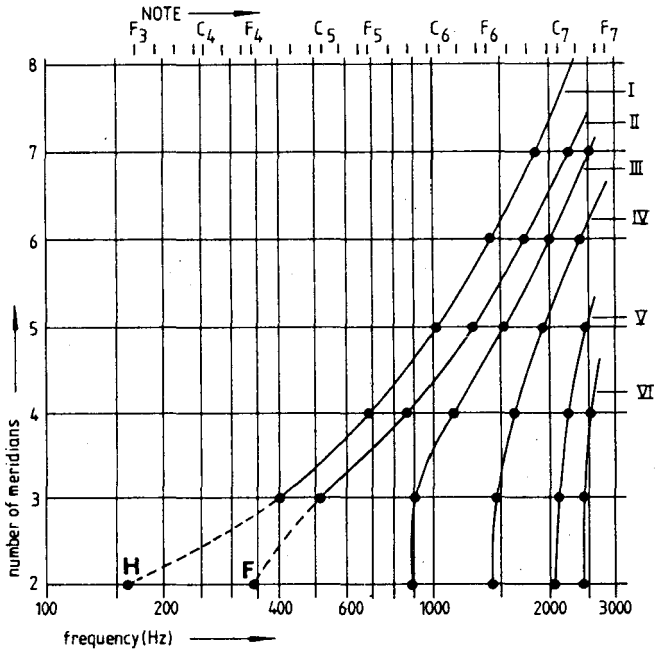


Fig. 6.6 Partial groups in bells (Lehr, 1986)

Perrin et al. (1983) have measured a very large number of partials of a tuned minor-third bell, too. In addition they have analysed the same bell using a finite element program and subsequently have matched the experimental and numerical results. Furthermore they proposed another classification scheme for partials. This scheme is based upon the number of nodal meridians, which is closely related to procedures in finite element programs. Here each specific number of meridians defines an eigenvalue problem and the eigenvalues of such a problem build a group. The interested reader is referred to Perrin's paper for more details.

The absolute value of the frequencies can be scaled easily to the desired value by means of proportional scaling of the dimensions of the bell, i.e. proportional scaling of the diameter, height, wall-thickness, etc. This can be illustrated by using the standard eigenvalue problem for free vibrations of a linear structure without damping.

$$\underset{\sim}{M}\ddot{u} + \underset{\sim}{K}u = \underset{\sim}{0} \qquad (6.1.3)$$

If we use the same scaling factor s for all dimensions, the mass matrix M is proportional to s^3 and the stiffness matrix K is proportional to s^2 (bending of the wall of the bell). Solving (6.1.3) it follows that the absolute frequencies are proportional to 1/s, leaving the frequency ratios unchanged! However, if this scaling is

used for all bells in a carillon, the treble bells will become too small and they cannot produce an appropriate sound volume. For this reason the treble bells are designed using a bell profile with increased wall thickness. In order to account for this phenomenum the so-called fD-parameter is used, which is the product of the frequency (in Hz) of the hum note and the diameter (in m) of the lip of the bell. In a large range of bell sizes the fD value should be from 190 to 220 m/s. The highest bell used in carillons has a pitch of c^6, giving a frequency of 4186 Hz for the hum note. Such a bell has a diameter of 0.18 m, resulting in fD ≈ 750 m/s.

When a large bell is struck by its clapper, first one hears the sharp clang of metal on metal. The sharp sound dies out quickly and the so-called strike note remains for a while. Most observers identify the pitch of the strike note at or near the second partial (the fundamental), but to others the pitch seems an octave higher. Apparently, the strike note cannot be measured physically, but must be determined by means of observers. As finally the sound of the bell ebbs, only the slowly decaying hum note, an octave below the fundamental, will remain to sound for a while.

6.2 Design of a major-third bell

6.2.1 Why a major-third bell?

Most musical instruments from the Western culture produce sounds with harmonic or nearly harmonic overtones, i.e. the frequency ratios of the partials build the row 1 : 2 : 3 : 4 : 5, etc., or in tones based on c: c - c^1 - g^1 - c^2 - e^2, etc. When playing a single note on such an instrument we actually hear an entire chord; however, we usually perceive only one sound with a single pitch and a certain timbre. On such an instrument music can be played in a major key as well as in a minor key. Of course, differences will be heard, but they are small because the major and the minor-third lie relatively high in the sound spectrum.

As already pointed out in Section 6.1, conventional bells show a different character. They have a strong minor-third in their lower partials. Special precautions must be taken when playing music in a major key on a carillon. For example, in general carillonneurs do not use a final major chord in order to avoid confusion with the lingering internal minor chord of the bells. Hence, the historical demand for major-third bells will be evident. Actually the first demand however emerged from the field of swinging bells. There the problem was met with a bell pealing that should give the opening of the Maria antiphon Salve Regina, which required the bells c1 - e1 - g1 - a1. In such a chord the c1 should be a bell with a major-third.

Major-third bells in a carillon may give problems as well; however, one may expect fewer problems due to the increased consonance of the bells. The Institute of Perception Research at Eindhoven investigated the appreciation for different bell types before the major-third bell had been realized. Houtsma et al. (1987)

modified the recorded sound of minor-third bells in that of major-third and "neutral-third" bells, using a computer-synthesizer. The neutral-third lies between the minor-third and the major-third. In the past those bells have been proposed as a general solution to the problem. In addition, neutral-third bells could be realized relatively easily. Using again the synthesizer, carillon melodies were composed in minor and in major key and listeners were asked which combination of melody and bell type they liked most. The results were surprising. It was found that (student) carillonneurs unanimously preferred minor-bells, whatever the key was in which they were played. In contrast, other musicians selected the right combinations, so minor melody on minor-bells and major-melody on major-bells. The average listener had an opposite opinion compared to the carillonneurs: in all cases he chose major-bells, whatever the key was in which they were played. The neutral-third bells were disliked by all listeners.

These results stimulated us in our quest for the major-third bell.

6.2.2 Problem formulation

The frequency ratios of a major-third bell should be equal to those of a minor-third bell (Table 6.1), except the third partial, which has to be changed into a major-third. The major-third must have the following frequency ratios:

- for swinging bells: 2.5000 or 1586 cents
- for carillon bells: 2.5198 or 1600 cents.

The bell-founder usually tunes only the frequency ratios of the five lowest partials. In order to create a tuning tolerance, the bell is cast with about 2 mm of extra material on the inner contour. The deviations after tuning should be less than 0.2% (≈ 3 cents). In practice such an accuracy is only attainable if the mutual deviations of the ratios of the founded bell are not larger than 3% (≈ 50 cents). After a common bell has been tuned, the untuned twelfth usually lies within 10 cents from the ideal 3100 cents (Table 6.1). The untuned double octave may lie up to 50 cents above 3600 cents. The twelfth and double octave of a major-third bell are allowed to shift somewhat. However, at least one of these partials should be close to the desired frequency ratio.

The design of a major-third bell is a shape optimization problem in which the design variables concern the geometry of the bell profile. The sum of squared residuals of frequency ratios of the five lowest partials must be minimized. This sum will be used as an objective function. The design constraints are trivial limitations on the geometry of the bell profile, such as positive wall thickness. In order not to exclude possible solutions, no constraints were placed on the bell profile. The restrictions with respect to the frequency ratios of the twelfth and the double octave, and the restriction with respect to the fD-parameter, can be formulated as behaviour constraints of the optimization problem.

6.2.3 A first attempt using iterative optimization

Considering the disappointing results of the bell-founders in searching a major-third bell, we were far from sure to find a solution. The optimization program should be based on finite element analysis of the structural behaviour, because of the very high accuracy of the frequency ratios which is needed.

We decided to develop a structural optimization program (DYNOPT, Van Asperen (1984)) suitable for axisymmetric, dynamically loaded structures. In this program the sequential linear programming method (SLP, Section 2.6.3) is used as optimization algorithm. The implemented finite elements include a 6-node and an 8-node iso-parametric ring element using partial discretization to describe the tangential displacements.

The program DYNOPT is suited to solve shape optimization problems, using the partial derivatives of the frequency ratios with respect to the design variables. The sum of squared residuals of absolute frequencies has been implemented as an objective function. The coupling between design variables and the finite element model can be defined by means of a user-supplied subroutine.

With the program DYNOPT the eigenfrequencies of bells can be calculated very accurately. The five lowest eigenfrequencies of a carefully measured and modelled minor-third bell could be computed within deviations of 1% from the measured frequencies. Fig. 6.7 shows the applied element mesh.

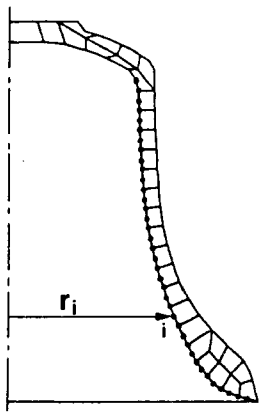


Fig. 6.7 Variable points for shape optimization

Van Asperen (1984) made a first attempt to find a major-third bell mainly by means of wall thickness variations. He defined as design variables the radial co-ordinates of the nodal points on the inner contour of the minor-third bell profile as indicated in Fig. 6.7. With the frequency ratios of a major-third bell as target values, DYNOPT changed the ratios as shown in Table 6.2. Although

most of the frequency ratios had improved a little, the ratio of the third partial still remained unacceptable. The range of variation of the bell geometry appeared to be too restricted due to the fixed outer contour of the bell profile. This confirmed the experiences of bell-founders that a major-third bell will not resemble a minor-third one. In order to find a solution a flexible geometrical model of the bell should be developed, enabling us to vary the bell geometry in a wide design space.

Name	Frequency ratios						
	ideal	initial values			after 10 iterations		
	$4f/f_{oc}$ [-]	$4f/f_{oc}$ [-]	deviations %	cents	$4f/f_{oc}$ [-]	deviations %	cents
Hum	1.0000	0.9841	-1.6	-27	0.9908	-0.9	-16
Fundamental	2.0000	2.0066	0.3	6	2.0231	1.2	20
Major-third	2.5198	2.3685	-6.0	-107	2.3781	-5.6	-100
Fifth	2.9966	3.0879	3.0	50	3.0300	1.1	20
Octave	4.0000	4.0000	0.0	0	4.0000	0.0	0

Table 6.2 Iterative optimization to search the major-third bell.

6.2.4 Solution_using_experimental_design

In a new effort to find a solution for the major-third bell we developed a flexible geometrical model (Maas (1985), Schoofs et al. (1987), Schoofs (1985)). This model was used in DYNOPT for explorative calculation in order to narrow the design space in which a solution had to be searched for. Defined on this design space and using experimental design techniques, regression models were developed relating frequency ratios to design variables. Finally, these regression models were used to find a solution.

6.2.4.1 Flexible_geometrical_model_of_the_bell

The geometry of the bell profile can be defined in several ways. We used the following description.

First we defined a set of basic points in a cylindrical co-ordinate system. Through these points a sufficiently smooth spline curve was fitted, which served as reference curve of the bell profile (Fig. 6.8a). At all basic points values for the wall thickness were defined. These wall thicknesses were measured perpendicularly and symmetrically with respect to the reference curve, which resulted in two additional sets of points. These inner and outer contour points of the bell were fitted by spline curves, resulting in a well determined bell profile (Fig. 6.8b).

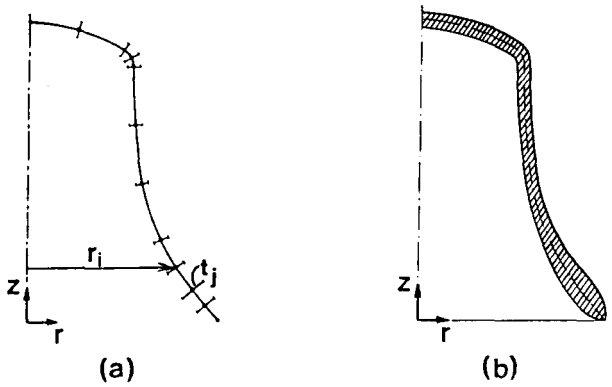


Fig. 6.8 (a) Reference curve and (b) bell profile

The r-co-ordinates of the basic points and the wall thicknesses at these points could be chosen either as fixed or as variable. Only those r-co-ordinates and wall thicknesses which were chosen to be variable formed part of the set of design variables. Fig. 6.9a shows the set of design variables and a sample of the element meshes used in the explorative calculations. The set of design variables included five radii and the wall thicknesses in the same points. In addition, the wall thickness at the important sound bow was also defined as a design variable. In this geometrical model the dimensions of the head of the bell were taken proportionally to the design variable r_5.

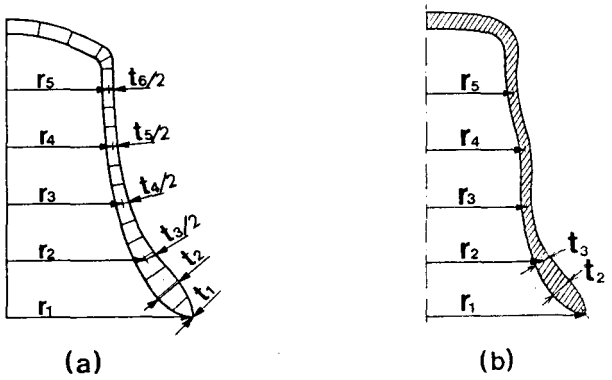


Fig. 6.9 (a) Design variables for explorative computations
(b) Prototype major-third bell with control variables for the experimental design

6.2.4.2 Explorative calculations; design of a prototype bell

Using this geometrical model we analysed bells, varying in shape from more or less common geometry to a very extreme geometry. The results pointed out that the frequency ratio of the third partial with respect to the octave could be influenced considerably. For the analysed bells the value of $4*f_{third}/f_{octave}$ ranged from about 2.14 up to 2.64, whereas a value of 2.5 is needed for a major third bell. Although in these variations some other important partials were usually out of range, the possibility to shift the third gave hope for finding a solution.

Next, a bell resembling the minor-third bell was considered. We tried to change the frequency ratio of the third from 2.4 to 2.5 by means of modifying the design variable which showed the (absolutely) largest partial derivative of the frequency ratio of the third partial. This resulted in a significant growth of design variable r_4. Here, the typical bulge of the major-third bell found later, was introduced. Using similar procedures for other design variables and other partials additional modifications were carried out, resulting in an intermediate design of which the frequency ratios are shown in Table 6.3. Of this design the ratios were pretty good, except for the fundamental, which was far too high. The bell founder advised us to stretch the height of the bell. From his long experience he knew that this would lower the fundamental without affecting the other partials. This resulted in the geometry

Name partial	Frequency ratios for					
	ideal major-third		intermediate design		prototype bell	
	$4f/f_{oc}$ [-]	cents	$4f/f_{oc}$ [-]	cents	$4f/f_{oc}$ [-]	cents
Hum	1.0000	0	1.0070	12	1.0163	28
Fundamental	2.0000	1200	2.1572	1331	1.9988	1199
Major-third	2.5198	1600	2.4909	1580	2.4852	1576
Fifth	2.9966	1900	2.9932	1898	2.9366	1865
Octave	4.0000	2400	4.0000	2400	4.0000	2400

Table 6.3 Intermediate design and prototype bell.

of a prototype major-third bell shown in Fig. 6.9b. The frequency ratios of this bell were very promising. In search of better solutions we explored the design space around this prototype using regression models of the frequency ratios.

6.2.4.3 Development of regression models for frequency ratios

From the explorative calculations mentioned in the previous subsection, the following limitations for the experimental design could be derived:

- All the design variables given by the radii appeared to have a rather strong influence on the frequency ratios. Of the wall thicknesses, only t_2 and t_3 had a strong influence. Therefore the radii r_1 through r_5 and the thicknesses t_2 and t_3 were taken as control variables in the experimental design. These control variables are indicated in Fig. 6.9b.
- The frequency ratios are nonlinear functions of these design variables. However, the nonlinearities with respect to radii appeared to be considerably stronger than those with respect to the wall thicknesses. It was therefore assumed that frequency ratios could be described sufficiently accurate by third-order models and a limited number of first-order interactions.
- The design space was limited to control variables varying $\pm$ 5% around their corresponding value in the prototype.
- Since partial derivatives of the frequency ratios were available too, the third-order models could be described by only two levels of the control variables. As experimental design a complete 2^7-design was used. (At that time no adequate software for an analysis of fractional designs using partial derivatives was available.)

Besides developing regression models for the four lowest frequency ratios, we decided to develop models for the ratios of the twelfth and the double octave too. Since the octave is the reference, it was not necessary to develop a regression model for this partial. The explorative calculations showed that these ratios are also considerably affected by the geometrical variations.

For all frequency ratios the same model functions were chosen in the regression equations. The complete 2^7-design existed of 128 finite element analyses, while one analysis required about 1100 sec. CPU on a PRIME 750 computer. The resulting regression models proved to be of such a high accuracy that they could replace direct finite element analyses.

6.2.4.4 Optimization_and_results

In order to find the geometry of a major-third bell, the following zero-order optimization was carried out using the derived regression models.

For each frequency ratio a sufficiently narrow band around the desired value was defined by

$$\left(\frac{f_i}{f_{oc}}\right)_{lower} \leq \left(\frac{f_i}{f_{oc}}\right)_{ideal} \leq \left(\frac{f_i}{f_{oc}}\right)_{upper}, \quad i = 1, \ldots, 7, \; i \neq 5 \qquad (6.2.1)$$

The variation range of each of the seven design variables was divided into a sufficiently large number (about 10) of equal intervals. Thus a narrow spaced 7-dimensional grid was defined on the design space. Subsequently we applied the following search procedure.

Consider a grid point and evaluate the first frequency ratio (i=1) according to (6.2.1). If the lower and upper limits are not

violated, then proceed with the next frequency ratio; otherwise pass

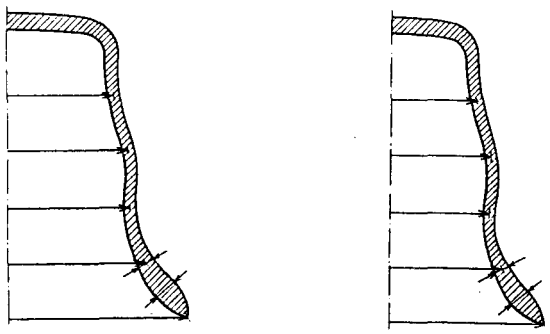


Fig. 6.10 Two solutions for major-third bells

to the next grid point. If all frequency ratios for a particular grid point satisfy (6.2.1), then retain this point as a possible solution for the geometry of a major-third bell.

Executing this procedure, eight different designs of major-third bells were found (Maas (1985)). Examined by means of direct finite element analyses, they were found to be very close to the ideal. Fig. 6.10 shows two of the resulting designs. From these designs, the most promising bell was selected and founded without adding extra material as tuning tolerance. Table 6.4 shows the frequency ratios of this bell.

Name	Frequency ratio			
	ideal cents	regr. models cents	DYNOPT cents	founded untuned bell cents
Hum	0	5	14	19
Fundamental	1200	1208	1210	1215
Major-third	1600	1589	1607	1616
Fifth	1900	1916	1907	1895
Octave	2400	2400	2400	2400
Twelfth	3100	3059	3057	?
Double octave	3600	3602	3596	?

Table 6.4 Results for the ultimate design of a major-third bell (? indicates that these frequency ratios were not measured).

Afterwards a second bell was founded and tuned exactly according to the frequency ratios of a major-third bell. The fD-value (see Section 6.1) of this new bell type was about 220 m/s, which is

acceptable, although this results in a lower sonority of the bell, compared to the conventional minor-bell. For carillon bells this lower sonority can be considered an advantage: the listener will be less confused by the lingering of the larger bells in the carillon (Lehr (1987b)). This bell, moulded by the Royal Bell-Foundry Eijsbouts, can be regarded as the first real major-third bell. Fig. 6.11 shows the final geometry of this bell, together with the typical shape of the conventional minor-third bell. A remarkable difference is it's pronounced bulge in the centre of the waist of the bell; furthermore, it is considerably taller than the minor-third bell.

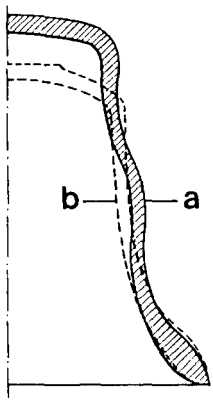


Fig. 6.11 Shape of (a) major-third bell
versus (b) minor-third bell

Applying major-bells, a complete carillon of four octaves was built (Fig. 6.12, Lehr (1987b)). The lowest bell had a diameter of 945 mm and a mass of 585 kg. For the smallest bell these values were 151 mm and 6 kg respectively. This first major-carillon was made mobile, so presentations at different locations could be realized easily. Although there was no systematic investigation of opinions of listeners during the various presentations of the major-third carillon, the conclusions of the perceptual investigation by Houtsma et al. (1987) mentioned before, were confirmed. The average carillonneur prefered the old trusted minor-carillon, the musician had a preference related to the key in which the music was played, while the layman had a strong preference for the major-carillon. It should be emphasized that these are averages. The spread in opinion was much greater for carillonneurs than for laymen; some carillonneurs rejected the major-carillon totally, whereas others reacted very enthousiastically. The extremes for layman were much more together and were completely within the range of acceptation. Using the new insights Lehr (1987a) designed an alternative major-bell for which the fD-value could be lowered to the conventional 200

m/s. The sonority of this bell lies between the sonority of the first major-bell and the sonority of the minor-bell.

Fig. 6.12 First carillon with major-third bells
(Courtesy of Royal Bell-Foundry Eijsbouts)

6.3 General bell design

6.3.1 Objective

In the conventional design of church and carillon bells the bell-founder reasons from his experience using a number of empirical design rules. These rules are linearized relationships between the bell geometry and characteristics of the sound of the bell, in particular the frequency spectrum. The geometrical modifications remain relatively small and there is little chance of finding new bells. Bell-founders also carried out many experiments, exploiting large variations of the bell profile, but compared to the effort and costs the profits remained small.

In the previous section a combined use was made of optimization procedures and experimental design techniques, resulting in the successful design of a major-third bell. The new bell profile differs considerably from that of conventional minor-third bells. The derived regression models for the frequency ratios can be used as design tools for bell profiles in the design space around the prototype bell shown in Fig. 6.9b.

In this section the shape optimization problem of bells will be extended to much larger geometrical variations. The design space will be defined as a broad area around the profile of the minor-third bell. For the eigenfrequencies of the seven most important partials (Table 6.1) we will derive regression models using optimal experimental design techniques as described in Chapter 3. Such regression models can be used for shape optimization problems. After having obtained an optimum the resulting geometry can be used as a starting point in a final optimization run based on direct finite element analyses.

6.3.2 Derivation of regression models for bell frequencies

In the following subsections the different aspects of deriving regression models for bell frequencies will be elucidated and discussed.

6.3.2.1 Response quantities

Table 6.5 shows a set of partials, ordered in groups of increasing number of meridians. Here a striking difference between groups of partials appears.

In conventional bell design partials with the same vertical vibration mode (Fig. 6.3a) constitute musically one group. For example, the group containing the partials I-4, I-5 and I-6 is very important for the strike note of the bell.

Mathematically the horizontal vibration mode (Fig. 6.3b), characterized by the number of meridian nodes, defines a unique eigenvalue problem. The lower eigenfrequencies of a particular eigenvalue problem constitute mathematically one group. Members in such a group will be ordered according to increasing eigenfrequency. We will indicate these members by the number of meridian nodes, followed by their eigenfrequency number. We will call this the

frequency code. Although for common bells the partial code and the frequency code coincide, it is emphasized that for relatively large profile variations discrepancies will occur.

Operating with large profile variations the musical names as given in Table 6.5 in general are no longer valid. For this reason in the sequel the frequency code will be used. Only if the frequency ratios satisfy musical intervals, we will apply their musical names.

We will restrict ourselves to the seven most important partials as mentioned in Table 6.1. The regression models for the limited set of seven partials may be used to search for musically interesting frequency spectra of bells.

Because the eigenfrequencies proved to be almost linear functions of wall thickness variables, we will develop regression models for absolute frequencies and not for frequency ratios.

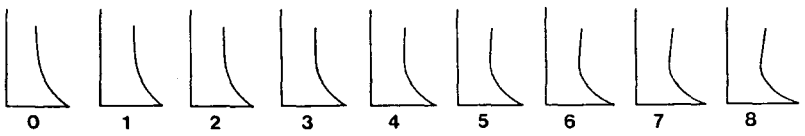
Musical name	Partial code	Frequency code	Frequency $4\ f/f_{oc}$	Ratio cents
Hum note	H - 2	2 - 1	0.9988	-2
Fundamental	F - 2	2 - 2	2.0000	1200
1st eleventh	III - 2	2 - 3	5.2174	2860
-	IV - 2	2 - 4	8.3542	3675
-	V - 2	2 - 5	11.8900	4286
Minor-third	I - 3	3 - 1	2.3977	1514
Fifth	II - 3	3 - 2	3.0053	1905
2nd eleventh	III - 3	3 - 3	5.3240	2895
-	IV - 3	3 - 4	8.5051	3706
-	V - 3	3 - 5	12.4883	4371
Octave	I - 4	4 - 1	4.0000	2400
Major tenth	II - 4	4 - 2	5.1396	2834
Thirteenth	III - 4	4 - 3	6.6384	3277
	IV - 4	4 - 4	9.5192	3901
Twelfth	I - 5	5 - 1	6.0175	3107
Fourteenth	II - 5	5 - 2	7.5597	3502
-	III - 5	5 - 3	8.9590	3796
-	IV - 5	5 - 4	11.4386	4219
Double octave	I - 6	6 - 1	8.3445	3673
-	II - 6	6 - 2	10.3627	4048

Table 6.5 Measured frequency ratios of a tuned minor-third bell (Lehr (1986)).

The parameters in the assumed linear relationships of the thickness variables can be computed using two levels for each wall thickness variable. However, when using FEM-analysis these parameters can be computed in a much more efficient way by adjusting the wall thickness variable to one level and computing at this level the eigenfrequency and it's partial derivative with respect to the wall thickness. This is a great advantage in designing the FEM-experiments, because design

variables adjusted to one level do not enlarge the number of required FEM-analyses.

The vertical vibration modes of common bells are given in Fig. 6.3a. If the geometry of the bell profile is varied considerably, then in general the vertical modes will change too. Even switching of the modes is possible. Fig. 6.13 shows a typical example. In Fig. 6.13a reference curves of nine bell profiles are given, beginning



(a) Profile reference curves

(b) Vibration modes of frequency 2-1

(c) Vibration modes of frequency 2-2

Fig. 6.13 Variation of a bell in eight gradual geometrical steps

with the profile of a minor-third bell, in eight gradual steps changing to a considerably modified profile. In Figs. 6.13b and c these reference curves have been stretched to straight lines. On the straight lines the corresponding computed vertical vibration modes have been plotted for the frequencies 2-1 and 2-2 respectively. Two very remarkable things occur. First, the vibration mode of frequency 2-1 changes continuously to the mode of frequency 2-2 of a minor-third bell. For frequency 2-2 the opposite is the case. Secondly, concerning the frequencies it is seen that initially their values approach each other and then move away from each other again (Fig.

6.14a). Regarding Fig. 6.14a one may ask whether or not the graphs of the frequencies 2-1 and 2-2 have actually crossed each other. Three additional computations in the circled area pointed out that this was not the case. In Fig. 6.14b the result is plotted together with similar results for other bell profiles. The numbers along the graphs indicate the vibration modes given in Figs. 6.13b and c.

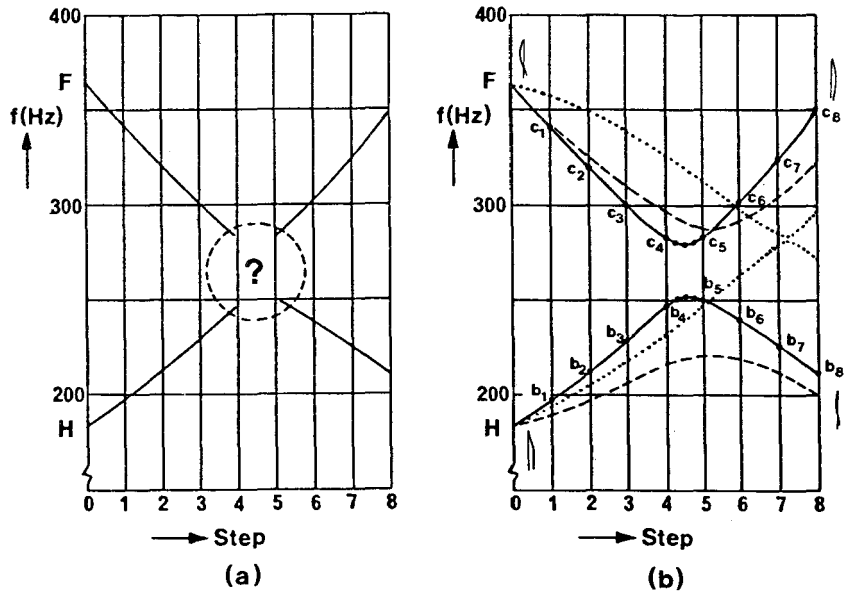
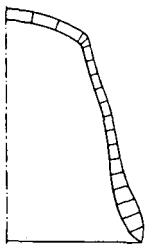


Fig. 6.14 (a) Frequency variations due to gradual changes in the bell profile
(b) Similar situations for different ultimate bell profiles

From these results we deduce that the graphs of the frequencies 2-1 and 2-2 do not cross if the bell profile is varied. Further computations pointed out that for higher frequencies shifts of vibration modes also occur. These shifts have not yet been investigated thoroughly.

In finite element analyses of bells different numbers of meridian nodes lead to different eigenvalue problems. In each eigenvalue problem a certain number of frequencies will be calculated. Graphs of frequencies belonging to different groups of meridian nodes may cross each other due to geometrical variations. Fig. 6.15 gives an example of a bell where this occurs.



Frequency code	Frequency ratio (cents)
2 - 1	-394
2 - 2	2307
3 - 1	1345
3 - 2	2407
4 - 1	2400
5 - 1	3149
6 - 1	3725

Fig. 6.15 Bell with an extravagant frequency spectrum

6.3.2.2 Geometrical model, design variables, ranges, levels and candidate points

For the definition of a geometrical model of the bell profile several methods can be used (Maas (1985)). Here we will use a model similar to the model used for the design of the major-third bell, described in section 6.2. However, the following modifications will be made.

We start with a nominal reference curve defined by a spline curve fitted through a number of np basic points between the lip and the shoulder of the bell (Fig. 6.16). The basic points can be treated as fixed or as variable; let nr be the number of basic points that are chosen to be variable. These variable points are related to the design variables in the following way. They are allowed to shift along a line perpendicular to the nominal reference curve. The magnitude of the shift, Δn_i, i = 1, 2, ..., nr, is proportional to it's corresponding design variable, x_i, i = 1, 2, ..., nr, and to the nominal radius, r_j, j $\in$ (1, 2, ..., np) of the basic point (Fig. 6.16); wr_i is a weighting factor:

$$\Delta n_i = wr_i * x_i * r_j \quad , \quad \begin{matrix} i = 1, 2, \ldots, nr \\ j \in (1, 2, \ldots, np) \end{matrix} \qquad (6.3.1)$$

For all weighting factors wr_i the value 0.05 will be used, resulting in a Δn_i equal to 5% of r_j if the variable x_i is shifted to 1.0. Through the fixed and the shifted basic point another spline curve, the actual reference curve, is fitted.

In the basic points nominal values, tn_j, j $\in$ (1, 2, ... np), for the wall thicknesses are defined, which too may be chosen fixed or related to nt design variables t_i, i = 1, 2, ..., nt. The latter nominal wall thicknesses result in the actual ones, ta_i, defined by (see Fig. 6.16):

$$ta_i = tn_j * (1 + wt_i * t_i) \quad , \quad \begin{matrix} i = 1, 2, \ldots, nt \\ j \in (1, 2, \ldots, np) \end{matrix} \qquad (6.3.2)$$

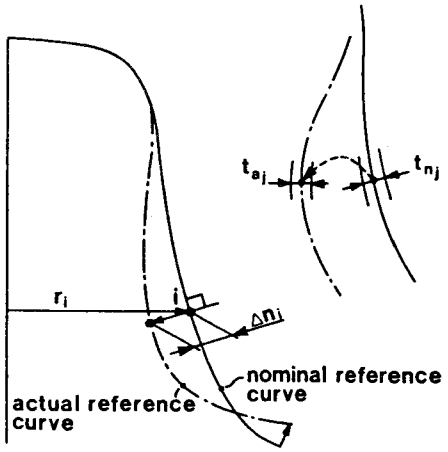


Fig. 6.16 Nominal and actual reference curve and definition of Δn_i, tn_j and ta_j

For all weighting factors wt_i the value 0.1 will be used, resulting in 10% change of the wall thickness if the variable t_i is shifted to 1.0. Through the fixed and variable discrete values of the wall thickness another spline curve is fitted. This results in a continuous function for the thickness. Thickness values emerging from this function are taken perpendicular to the actual reference curve, resulting in the actual bell profile from the lip to the shoulder.

The profile of the head of the bell is taken proportional to the radius of the bell at the shoulder. In the resulting profile an appropriate finite element mesh is generated; Fig. 6.15 shows an example of such a mesh.

For deriving generally applicable regression models of bell frequencies we start with the profile of the minor-third bell as a nominal profile. Geometrical variations are allowed in a limited area around this old and thoroughly developed bell profile.

The design variables will be related to seven curvilinearly equidistant basic points from lip to shoulder. We believe seven points are a good compromise resulting in sufficient geometrical flexibility of the model and in reasonable computing cost. The design variables will be related to the wall thicknesses in the basic points and in intermediate points (Fig. 6.17). The number of design variables related to wall thickness parameters is not critical from the viewpoint of experiment design, because these design variables will not be varied.

In order to choose the model functions a number of explorative computations were carried out. Fig. 6.18 shows five types of

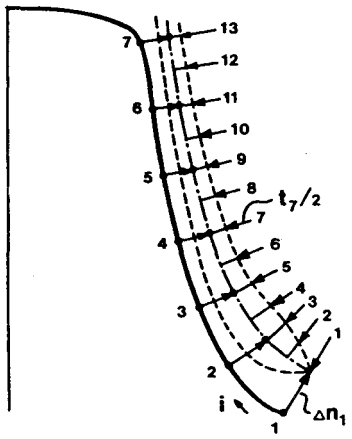


Fig. 6.17 Design variables in the experimental design

variation of the design variables x_i, i = 1, 2, ..., 7. The types of variation are defined by the following functions of the parameter α depicted in the figures:

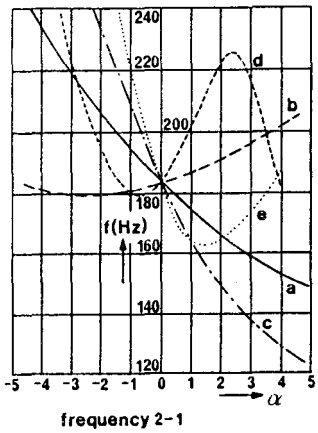
a. $x_i = \alpha$, i = 1, 2, ..., 7
b. $x_i = \alpha\ (4-i)/3$, i = 1, 2, ..., 7
c. $x_i = \alpha\ (1 - |4-i|/3)$, i = 1, 2, ..., 7
d. $x_2 = \alpha$, $vr_i - 0$ for i = 1, 3, 4, 5, 6, 7
e $x_3 = \alpha$, $vr_i = 0$ for i = 1, 2, 4, 5, 6, 7

Fig. 6.18 gives also the computed eigenfrequencies of the five lowest frequencies, grouped according to frequency. They show strong nonlinearities. From the results of the variation types a, b and c it can be seen that the nonlinearities become larger for increasing values of α, so the range of the geometrical variations should be limited, allowing us to use relatively simple (low order) regression equations. For this reason the design space will be constrained to:

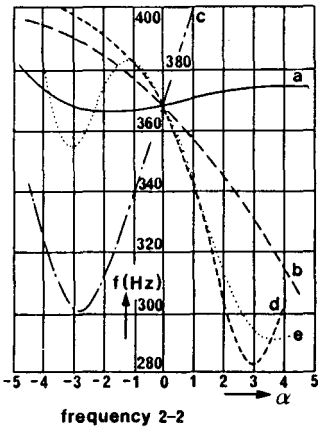
$$-3.0 \leq x_i \leq 3.0 \quad , \quad i = 1, 2, \ldots, 7 \qquad (6.3.3)$$

The local geometrical variations given by the cases d and e in Fig. 6.18 result in considerably stronger nonlinearities than the more global variation types a, b and c. In order to limit these nonlinearities, extra constraints will be applied:

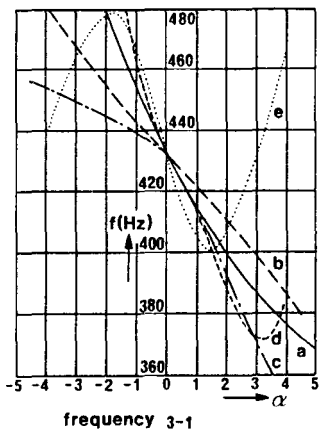
$$|x_i - x_{i-1}| \leq 2.0 \quad , \quad i = 2, 3, \ldots, 7 \qquad (6.3.4)$$



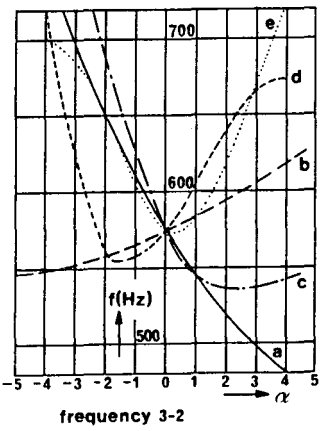
frequency 2-1



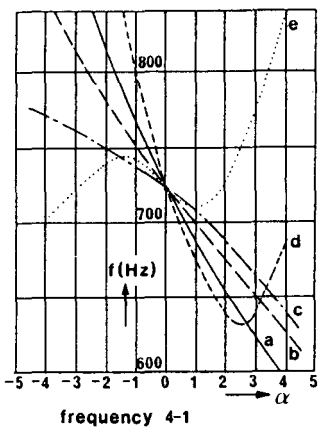
frequency 2-2



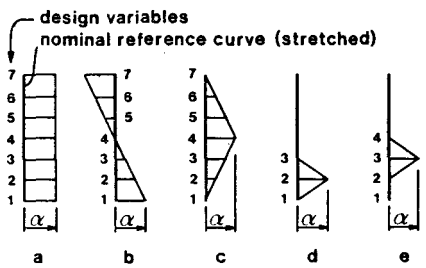
frequency 3-1



frequency 3-2



frequency 4-1



Types of variation of the design variables

Fig. 6.18 Frequency variations due to geometrical variations

Computations, in which the wall thicknesses were varied in ranges from 0.7 to 1.4 times the nominal thicknesses (i.e. concerning design variables ranging from -3.0 to +4.0), pointed out that the resulting frequencies in good approximation (within 2%) can be described by linear functions. As discussed before, those linear functions can adequately be estimated by exploiting partial derivatives of the frequencies with respect to the wall thicknesses. Hence, these design variables do not have to be varied in the experimental design. They will be held zero, that is the nominal wall thicknesses will be used (see (6.3.2)).

The ranges (6.3.3) and the constraints (6.3.4) can be realized by means of the following four equidistant levels for each of the design variables x_i, i = 1, 2, ..., 7:

$$x_i \in (-3.0, -1.0, +1.0, +3.0) \tag{6.3.5}$$

From the explorative results given in Fig. 6.18 we conclude that regression equations should be at least of order three or four. Such regression equations can adequately be described using four levels, since partial derivatives of the frequencies will be used.

The set of candidate points for the experimental design can now be generated. First, a complete 4^7-design is chosen, i.e. consisting of all possible combinations of the design variables x_i, i = 1, 2, ..., 7, on four levels. From this design all those points are skipped that show a difference of more than one level for contiguous variables (constraints (6.3.4)). This results in a set of 1220 candidate points.

6.3.2.3 Building the regression models

The design variables will be stored in two columns $\underset{\sim}{r}$ and $\underset{\sim}{t}$ defined by:

$$\underset{\sim}{r} = [x_1\ x_2\ \ldots\ldots\ldots\ x_{nr}]^T \tag{6.3.6}$$

and $$\underset{\sim}{t} = [t_1\ t_2\ \ldots\ldots\ldots\ t_{nt}]^T \tag{6.3.7}$$

We will start the development of the experimental design based on a regression equation which is chosen to be of third order in the variables $\underset{\sim}{r}$ and linear in the variables $\underset{\sim}{t}$. It is convenient to write the regression equation for a frequency $f_i = f_i(\underset{\sim}{r}, \underset{\sim}{t})$ as:

$$f_i(\underset{\sim}{r}, \underset{\sim}{t}) = fr_i(\underset{\sim}{r}) + ft_i(\underset{\sim}{r}, \underset{\sim}{t}) \tag{6.3.8}$$

For $fr_i(\underset{\sim}{r})$ a relatively simple polynomial will be chosen of the form:

$$
\begin{aligned}
fr_i = \beta_0 + & \\
& \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3 + \\
& \cdots \\
& \beta_{19} x_7 + \beta_{20} x_7^2 + \beta_{21} x_7^3 + \\
& \beta_{22} x_1 x_2 + \beta_{23} x_2 x_3 + \beta_{24} x_3 x_4 + \beta_{25} x_4 x_5 + \\
& \beta_{26} x_5 x_6 + \beta_{27} x_6 x_7 + \\
& \beta_{28} x_1 x_3 + \beta_{29} x_2 x_4 + \beta_{30} x_3 x_5 + \beta_{31} x_4 x_6 + \beta_{32} x_5 x_7 + \\
& \beta_{33} x_1 x_4 + \beta_{34} x_2 x_5 + \beta_{35} x_3 x_6 + \beta_{36} x_4 x_7
\end{aligned}
\qquad (6.3.9)
$$

For $ft_i(\underset{\sim}{r}, \underset{\sim}{t})$ we chose a polynomial which is linear in the variables t_j, $j = 1, 2, \ldots$ nt of the form:

$$
\begin{aligned}
ft_i(\underset{\sim}{r}, \underset{\sim}{t}) = & \gamma_1 t_1 + \gamma_2 t_2 + \ldots + \gamma_{nt}\, t_{nt} + \\
& \gamma_{nt+1} g_1(\underset{\sim}{r})\, t_1 + \ldots \gamma_{nt+nt} g_{nt}(\underset{\sim}{r})\, t_{nt}
\end{aligned}
\qquad (6.3.10)
$$

where $g_j(\underset{\sim}{r})$, $j = 1, 2, \ldots,$ nt are common polynomials in the components of $\underset{\sim}{r}$.

The components of $\underset{\sim}{t}$ will not be varied, but fixed on the level "zero"; the linear relationship can be estimated using partial derivatives of the frequencies with respect to the components of $\underset{\sim}{t}$. This is a rather unusual situation, which is not supported by standard experimental design and regression software. But this problem can still be tackled using standard software if the two terms on the right-hand side of (6.3.8) are treated separately. The parameters of the model (6.3.9) can then be estimated without worrying about the variables $\underset{\sim}{t}$.

In order to estimate the parameters of the model (6.3.10) we differentiate this model with respect to the components of $\underset{\sim}{t}$:

$$
\frac{\partial ft_i}{\partial t_j} = \gamma_j + \gamma_{nt+j} g_j(\underset{\sim}{r}) \quad , \quad j = 1, 2, \ldots, nt \qquad (6.3.11)
$$

Explorative computations pointed out that in the area $-3.0 \leq t_j \leq 4.0$, $j = 1, 2, \ldots,$ nt, a particular frequency f_i in good approximation can be estimated by:

$$
\hat{f}_i = fr_i + \sum_{j=1}^{nt} \left(\frac{\partial ft_i}{\partial t_j}\right) t_j \qquad (6.3.12)
$$

We will begin the model building with frequency 3-1, because in bell design this frequency is very crucial. Maas (1985) could develop

adequate regression models in seven variables using a complete 2^7-design, requiring 128 design points.
Now we consider a wider design space and we expect that at least 300 to 400 design points will be required in the final experimental design. Remembering the 25%-rule (Box et al (1978)), we chose 100 design points in the first experimental design. Using the program CADE (Nagtegaal (1987)) the design is optimized applying the excursion algorithm and the D-optimality criterion.

INPUT SUMMARY

ALGORITHM	EX	CANDIDATES	1220	MODELTYPE	P
CRITERION	DO	OBSERVATIONS	100	PARAMETERS	37
		PROTECTED	0	VARIABLES	7
		REPLICATIONS	NO	DEPENDENT	1
		UPPER LIMIT	150	DERIVATIVES	7
		LOWER LIMIT	100		

DESIGN PROPERTIES

DESIGN REF	ITERATIONS	DET**(1/37) COVARIANCE MATRIX	MAX.TRACE VARIANCE MATRIX	AVE.TRACE VARIANCE MATRIX	MAX E-VAL COVAR. MATRIX	TRACE COVAR. MATRIX	|CORRELATION| MAX	AVER AGE	%NON ZERO
1	135	0.549E-01	0.413E+02	0.308E+02	0.777E+01	0.168E+02	0.83	0.25	20

ALL PROPERTIES NORMALIZED FOR NUMBER OF OBSERVATIONS AND ERROR VARIANCE

DESIGN POINTS

DESIGN REF														
1	3	10	18	27	33	45	47	69	76	87	89	90	92	97
	100	102	108	109	122	123	130	179	181	183	219	221	223	227
	228	229	235	259	260	269	273	280	282	283	304	320	322	336
	468	477	571	574	577	579	595	596	611	623	626	628	644	651
	704	742	752	866	883	903	912	919	941	946	952	956	961	974
	977	989	990	991	993	1000	1023	1036	1038	1040	1063	1093	1099	1105
	1106	1114	1119	1125	1126	1133	1134	1145	1174	1181	1186	1187	1201	1205
	1218	1220												

CPU-TIME FOR THIS JOB IN SECONDS : 0.693E+05

Table 6.6 Optimal design for 100 observations

The optimization takes into account that function values and partial derivatives of the frequency will be used in estimating the model parameters. Table 6.6 gives the CADE-run on an Apollo D-3000 workstation. The properties of the design are good. The maximum variance is 41.3 which is not much bigger than the theoretical minimum of 37 being the number of parameters. However, it is emphasized that the value 41.3 might be somewhat too optimistic, due to the fact that we are operating with a fixed set of candidate points. Quite remarkable is the huge amount of required computing time (19.2 hrs) due to the large dimensions of the problem: we have 1220 candidate points, eight response quantities (1 function + 7 partial derivatives), and 37 model parameters.

Using collected data in the generated design points the model parameters have been estimated by means of the iterative reweighted

least squares procedure. Table 6.7 gives a report of the CADE-run. In order to test the predictive ability of the regression models we generated a set of 100 testing points uniformly distributed over the design area and distinct from the candidate points. Tested against this data the estimated regression model shows a standard deviation

WEIGHTED ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGR OF FREEDOM	MEAN SQUARE	F-LEVEL	MULTIPL CORRELA	STAND ERROR OF ESTIMA
REGRESSION	0.38054E+04	24	0.15856E+03	225.529	0.93528	0.83848E+00
RESIDUAL	0.54486E+03	775	0.70305E+00			
TOTAL	0.43502E+04	799				

ESTIMATIONS OF PARAMETERS

PARAMETERS			REGRESSION COEFFICIENT	STANDARD DEVIATION	F-LEVEL
INTERCEPT			0.42610E+03		
MAIN EFFECTS:					
VARIABLE	POWER	REGRESSOR			
1	1	1	0.13405E+02	0.40562E-01	1109223.694
1	2	2	0.12722E+01	0.19690E-01	4174.374
2	1	4	-0.33263E+02	0.49348E-01	454349.901
2	2	5	0.53678E+01	0.26288E-01	41695.180
3	1	7	-0.12975E+02	0.48852E-01	70543.014
3	2	8	0.23369E+01	0.25199E-01	8600.466
4	1	10	0.84588E+01	0.76701E-01	12162.436
4	2	11	0.17747E+01	0.24871E-01	5091.774
4	3	12	0.14194E+00	0.61256E-02	536.915
5	2	14	0.68085E+00	0.20836E-01	1067.748
5	3	15	0.10345E+00	0.37351E-02	767.141
7	1	19	-0.13896E+01	0.33642E-01	1706.062
7	2	20	0.19579E+00	0.15863E-01	152.341
INTERACTION		REGRESSOR			
1		22	-0.53464E+01	0.34119E-01	24554.218
2		23	-0.22676E+01	0.37338E-01	3688.306
3		24	-0.62830E+01	0.36597E-01	29474.710
4		25	-0.67649E+00	0.34497E-01	384.565
6		27	-0.49801E+00	0.21958E-01	514.400
7		28	0.10805E+01	0.24720E-01	1910.534
8		29	0.85366E+00	0.29515E-01	836.536
9		30	0.55921E+00	0.28497E-01	385.081
12		33	0.24105E+01	0.21133E-01	13009.939
13		34	0.61330E+00	0.22819E-01	722.373
14		35	0.56560E+00	0.18766E-01	908.415

MODEL	WEIGHTS
1	0.0892831
2	0.0880495
3	0.0555363
4	0.0455186
5	0.1192538
6	0.1506882
7	0.2865387
8	0.6046083

CPU-TIME FOR THIS JOB IN SECONDS : 169.314

Table 6.7 Model fitting using 100 observations and partial derivatives of frequency 3-1

for frequency 3-1 of 13.2 Hz with a mean of 436 Hz. This results in a 95% confidence interval of ± 6%, or ± 102 if expressed in cents. Such a small predictive ability is insufficient for bell design and the experimental design should be augmented. Augmentation of the design with 100 design points lowered the standard deviation for frequency 3-1 to 11.2 Hz, which is still insufficient.

Using the 200-points design to fit regression models of type (6.3.9) for the other frequencies showed somewhat better results for the frequencies 4-1, 5-1 and 6-1. However, the results for the frequencies 2-1, 2-2 and 3-2 were so disappointing, that satisfying regression models of type (6.3.9) could hardly be expected by raising the number of design points to 400.

In order to find upper bounds on the predictive ability, we permitted ourselves to collect data in all 1220 candidate points. At the same time we augmented the main effects in the regression equation to fourth order and we changed the set of interaction terms, resulting in the following regression equation:

$$
\begin{aligned}
fr_i = \beta_0 \; + & \qquad (6.3.13)\\
& \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3 + \beta_4 x_1^4 + \\
& \cdots \\
& \beta_{25} x_7 + \beta_{26} x_7^2 + \beta_{27} x_7^3 + \beta_{28} x_7^4 + \\
& \beta_{29} x_1 x_2 + \beta_{30} x_2 x_3 + \beta_{31} x_3 x_4 + \\
& \beta_{32} x_4 x_5 + \beta_{33} x_5 x_6 + \beta_{34} x_6 x_7 + \\
& \beta_{35} x_1 x_2 x_3 + \beta_{36} x_2 x_3 x_4 + \beta_{37} x_3 x_4 x_5 + \\
& \beta_{38} x_4 x_5 x_6 + \beta_{39} x_5 x_6 x_7 + \\
& \beta_{40} x_1 x_3 + \beta_{41} x_2 x_4 + \beta_{42} x_3 x_5 + \\
& \beta_{43} x_4 x_6 + \beta_{44} x_5 x_7 + \\
& \beta_{45} x_1 x_2^2 + \beta_{46} x_1^2 x_2 + \beta_{47} x_2 x_3^2 + \beta_{48} x_2^2 x_3 + \\
& \beta_{49} x_3 x_4^2 + \beta_{50} x_3^2 x_4 + \beta_{51} x_4 x_5^2 + \beta_{52} x_4^2 x_5 + \\
& \beta_{53} x_5 x_6^2 + \beta_{54} x_5^2 x_6 + \beta_{55} x_6 x_7^2 + \beta_{56} x_6^2 x_7
\end{aligned}
$$

The reweighted least squares parameter estimation required 1.5 to 2 hours of computing time per regression model. Table 6.8 shows the results of testing the models against the set of 100 testing points.

Note that the deviations in the model for frequency 4-1, which is the reference, reduce to zero when frequency ratios are considered. These deviations are added to the deviations of the other

models, thereby decreasing the predictive ability of the other models.

Frequency code	2-1	2-2	3-1	3-2	4-1	5-1	6-1
Mean frequency (Hz)	192	360	436	605	735	1095	1499
Standard deviation(Hz)	8.4	15.4	8.3	12.3	8.9	9.5	9.0
Standard deviation(cents)	78	67	19	43	0	8	13

Table 6.8 Test of regression models with 57 parameters and based on all candidate points, against 100 testing points.

The predictive ability of the models for the frequencies 3-1, 4-1, 5-1 and 6-1 is satisfactory. We pose that the 95% confidence interval (2 x standard deviation) should be less than 50 cents, which is the case for the four frequencies mentioned above. However, the frequencies 2-1, 2-2 and 3-2 have 95% confidence intervals of $\pm$ 156, $\pm$ 134 and $\pm$ 86 cents respectively, which are much too large.

Table 6.9 (page 6.32) shows the CADE-report of fitting the model for frequency 2-2. It can be seen that 26 of the 28 interaction terms have been selected in the model, and we may ask ourselves whether the augmentation of the number of regression terms would help. The F-levels in Table 6.9 point in that direction. The F-levels of the main effects gives no arguments to augment the order of the main effects. We strongly augmented the number of interaction terms in the regression equation, namely from 28 to 80 and we fitted the models again. Table 6.10 shows the results.

Frequency code	2-1	2-2	3-1	3-2	4-1	5-1	6-1
Mean frequency (Hz)	192	360	436	605	735	1095	1499
Standard deviation(Hz)	6.1	12.3	7.4	7.5	8.0	8.5	8.0
Standard deviation(cents)	55	58	17	34	0	7	11

Table 6.10 Test of regression models with 109 parameters and based on all candidate points, against 100 testing points.

A significant improvement of the models for the frequencies 2-1, 2-2 and 3-2 can be seen, but these models are still not satisfactory. Nevertheless we believe that the difficulties in fitting models for these three frequencies can be solved by means of a "lucky strike" in choosing suitable potential interaction terms in the model. This idea is based on the common vibration modes of the frequencies. The modes for the frequencies 3-1, 4-1, 5-1 and 6-1 are often the same as for the third, the octave, the twelfth and the double octave of common

WEIGHTED ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGR OF FREEDOM	MEAN SQUARE	F-LEVEL	MULTIPL CORRELA	STAND ERROR OF ESTIMA
REGRESSION	0.20686E+05	45	0.45969E+03	477.680	0.82991	0.98099E+00
RESIDUAL	0.93481E+04	9714	0.96233E+00			
TOTAL	0.30034E+05	9759				

ESTIMATIONS OF PARAMETERS

PARAMETERS			REGRESSION COEFFICIENT	STANDARD DEVIATION	F-LEVEL
INTERCEPT			0.35016E+03		
MAIN EFFECTS:					
VARIABLE	POWER	REGRESSOR			
1	2	2	-0.19318E+01	0.70119E-02	75901.150
2	1	5	-0.16065E+02	0.24367E-01	434636.759
2	2	6	-0.19365E+01	0.93291E-02	43086.410
3	1	9	-0.10966E+02	0.33805E-01	105226.003
3	2	10	-0.25493E+01	0.91399E-02	77797.832
3	3	11	0.75007E+00	0.50936E-02	21684.768
4	1	13	0.22552E+02	0.34970E-01	415866.103
4	2	14	-0.15130E+01	0.17109E-01	7820.031
4	3	15	-0.24868E+01	0.51075E-02	237067.924
4	4	16	0.24177E-01	0.10817E-02	499.503
5	1	17	0.28956E+02	0.35236E-01	675310.964
5	2	18	-0.28932E+01	0.17192E-01	28322.811
5	3	19	-0.22888E+01	0.52939E-02	186925.704
5	4	20	0.26719E-01	0.10816E-02	610.279
6	1	21	-0.94593E+01	0.32181E-01	86401.250
6	2	22	-0.29636E+01	0.17089E-01	30075.515
6	3	23	0.55174E+00	0.40315E-02	18729.717
6	4	24	0.16096E-01	0.10763E-02	223.653
7	1	25	-0.89897E+01	0.18418E-01	238225.620
INTERACTION		REGRESSOR			
1		29	0.40129E+01	0.12396E-01	104798.300
2		30	0.45950E+01	0.13283E-01	119663.785
3		31	0.23066E+01	0.13259E-01	30263.855
4		32	0.10087E+02	0.13333E-01	572390.933
5		33	0.30034E+01	0.13431E-01	50003.553
6		34	0.21514E+01	0.97949E-02	48243.891
7		35	-0.30953E+00	0.57146E-02	2933.833
8		36	0.16298E+01	0.62049E-02	68989.517
10		38	0.18399E+00	0.62065E-02	878.854
11		39	0.46832E+00	0.56360E-02	6904.630
12		40	-0.78145E+00	0.87327E-02	8007.643
13		41	-0.62094E+01	0.90987E-02	465724.994
14		42	-0.11253E+01	0.91556E-02	15106.908
15		43	-0.92034E+00	0.91011E-02	10226.059
16		44	-0.26374E+01	0.86557E-02	92840.471
17		45	0.48814E+00	0.59727E-02	6679.385
18		46	-0.13797E+00	0.46633E-02	875.283
19		47	0.64452E+00	0.87851E-02	5382.498
20		48	-0.10134E+01	0.74052E-02	18727.841
21		49	0.38232E+01	0.97233E-02	154607.111
22		50	-0.40682E+01	0.10063E-01	163436.917
23		51	0.90478E+00	0.10069E-01	8073.831
24		52	0.70311E+00	0.97290E-02	5222.791
25		53	-0.26615E+01	0.10082E-01	69689.661
26		54	0.31968E+01	0.10074E-01	100695.615
27		55	-0.13033E+00	0.35319E-02	1361.637

CPU-TIME FOR THIS JOB IN SECONDS : 6235.118

Table 6.9 Model fitting for frequency 2-2 using data in all candidate points

bells. These vibration modes show very small movements in the upper half of the bell. Hence, the interactions between the design variables are mainly restricted to the lower half of the bell. For the frequencies 2-1, 2-2 and 3-2, whose vibration modes often resemble those of the hum, the fundamental and the fifth in common bells, the interactions between the design variables are distributed over the full height of the bell.

Possibly better solutions might be found when a suitable linear model is used in which terms are chosen guided by, for example, the analytical theories for a circular and/or a conical shell. We will leave this for future research.

The comprehensive set of collected data in all candidate points can be studied directly in order to find starting points for optimization runs using direct FE-computations. Although the candidate points constitute a very widely spaced grid over the entire design area, we were able to design three new bells in this way. We will describe these bells in the next subsection.

6.3.3 Shape optimization

Besides the frequency spectra of minor-third and the major-third bells there exist several other musically interesting frequency spectra; see Table 6.11, taken from Lehr (1987a). The bell types have been ordered according to decreasing consonance. This can be seen from the row of the most simple integer ratios. The lower and the more regular the figures in this row are, the more consonant the bell is. To build carillons the bell types no. 1 through 4 are suited and, to a less extent, also the types 5 and 6; in accordance with the tuning of carillon bells, the cents values of these bells are given in hundreds (see also Table 6.1). The other bell types in the list can only be used as swinging bells.

In the next subsection we will report on our attempt to find the most consonant bell: a bell with harmonic partials.

6.3.3.1 Attempt to find a bell with harmonic frequency ratios

In an attempt to find a bell with harmonic frequency ratios we applied a systematic search on the data collected in the 1220 candidate points. Besides the five ratios mentioned in Table 6.11, the frequencies 5-1 and 6-1 were also considered. Their ideal ratios should be 3100 and 3600 cents respectively. Hence, the ideal frequency ratios of a harmonic bell are:

frequency code	2-1	2-2	3-1	3-2	4-1	5-1	6-1
ratio in cents	-400	800	1500	2000	2400	3100	3600

Bell type	Musical tones and cents values					Ratio $4f/f_{octave}$ and most simple integer ratios				
No. and name	2-1	2-2	3-1	3-2	4-1	2-1	2-2	3-1	3-2	4-1
1. bell with harmonic partials	c1	c2	g2	c3	e3	4/5	8/5	12/5	16/5	4
	-400	800	1500	2000	2400	1	2	3	4	5
2. fifth-bell	c1	g1	c2	g2	c3	1	3/2	2	3	4
	0	700	1200	1900	2400	2	3	4	6	8
3. major-third bell	c1	c2	e2	g2	c3	1	2	5/2	3	4
	0	1200	1600	1900	2400	2	4	5	6	8
4. minor-third bell	c1	c2	es2	g2	c3	1	2	12/5	3	4
	0	1200	1500	1900	2400	5	10	12	15	20
5. minor-sixth bell	e1	c2	e2	g2	c3	5/4	2	5/2	3	4
	400	1200	1600	1900	2400	5	8	10	12	16
6. major-sixth bell	es1	c2	es2	g2	c3	6/5	2	12/5	3	4
	300	1200	1500	1900	2400	6	10	12	15	20
7. major-sixth bell with diminished fundamental	es1	bes1	es2	g2	c3	6/5	9/5	12/5	3	4
	316	1018	1516	1902	2400	6	9	12	15	20
8. minor-sixth bell with diminished fundamental	e1	a1	e2	a2	c3	5/4	5/3	5/2	10/3	4
	386	884	1586	2084	2400	15	20	30	40	48
9. twelfth bell	g0	c2	e2	g2	c3	2/3	2	5/2	3	4
	-702	1200	1586	1902	2400	4	12	15	18	24
10. octave bell with diminished fundamental	c1	bes1	d2	f2	c3	1	9/5	9/4	27/10	4
	0	1018	1404	1720	2400	20	36	45	54	80

Table 6.11 Lowest frequencies of musically interesting bells.

Furthermore we assumed an ideal fD-value of 200 m/s. The quality of the design points $\underset{\sim}{x}_j$ with respect to the harmonic bell could then be judged by the standard deviation $S(\underset{\sim}{x}_j)$, defined by:

$$S^2(\underset{\sim}{x}_j) = \Big[\sum_{i=1}^{7} w_i^2\{c_i(\underset{\sim}{x}_j) - c_i^*\}^2 + w_8^2\{fD(\underset{\sim}{x}_j) - 200\}^2\Big] \Big/ \sum_{i=1}^{8} w_i^2$$

$$j = 1, 2, \ldots, r \qquad (6.3.14)$$

where w_i, $i = 1, 2, \ldots, 8$ are weighting factors, $c_i(\underset{\sim}{x}_j)$ are frequency ratios in cents, computed in the design points, c_i^* are the corresponding desired frequency ratios and $fD(\underset{\sim}{x}_j)$ is the fD-value computed in point $\underset{\sim}{x}_j$. Table 6.12 contains the best 25 points, if in (6.3.14) all weighting factors are chosen equal to one. It appears that for these 1220 design points:

1. The frequency ratio for 2-1 is at least 100 cents too high.
2. The ratio for 3-1 is at least 100 cents too low.
3. The ratios for 2-2, 3-2, 5-1 and 6-1 vary around their desired values.
4. Lower values for ratio 2-1 usually coincide with the lower values for ratio 3-1 and the higher values for 5-1 and 6-1.
5. Most of the fD-values lie above the ideal of 200 m/s.

The ratios for 2-1 and 3-1 clearly will make it difficult for us to find a solution. If we concentrate our search on these ratios by choosing their weighting factors equal to one, and all the other weighting factors equal to zero, then the set of 25 points in Table 6.13 is obtaind. Although the range in which the ratios for 2-1 now lie has moved towards the desired value of -400, no ratio is lower than -385 cents, whereas the range in which the ratios for 3-2 lie remained approximately the same. We have plotted in Fig. 6.19 four bells, selected from the top of Table 6.12.

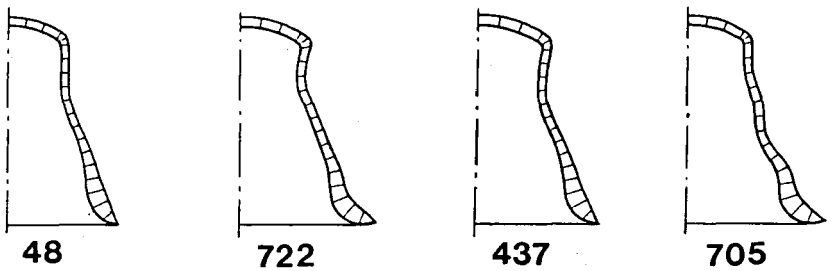


Fig. 6.19 Ideas for the geometry of a harmonic bell

Our conclusion is that a harmonic bell will not be found in this design area. To our opinion designing a harmonic bell will be extremely difficult.

CAND. POINT NO.	S	FREQUENCY RATIOS IN CENTS 2-1	2-2	3-1	3-2	4-1	5-1	6-1	FD	DESIGN VARIABLES X1	X2	X3	X4	X5	X6	X7
70	63	-286	860	1400	1939	2400	3096	3625	239	-3	-3	-1	1	-1	-3	-1
48	69	-313	783	1373	1914	2400	3127	3677	232	-3	-3	-1	-1	-3	-3	-3
722	78	-235	794	1413	1899	2400	3081	3582	259	1	-1	1	1	-1	-3	-1
437	78	-266	842	1372	2013	2400	3139	3705	217	-1	-1	1	1	-1	-1	1
705	86	-213	878	1413	1912	2400	3094	3610	255	1	-1	1	-1	-1	-3	-3
438	86	-234	796	1380	2085	2400	3134	3695	219	-1	-1	1	1	-1	1	-1
49	87	-274	654	1374	1977	2400	3127	3677	232	-3	-3	-1	-1	-3	-3	-1
706	89	-178	759	1412	1948	2400	3094	3610	255	1	-1	1	-1	-1	-3	-1
1098	91	-211	929	1397	2011	2400	3110	3641	240	3	1	3	1	-1	-3	-3
40	93	-203	775	1390	2101	2400	3124	3675	238	-3	-3	-1	-3	-1	-3	-3
187	94	-244	846	1377	2119	2400	3143	3719	201	-3	-1	1	-1	-1	-1	-1
1121	95	-235	963	1410	1905	2400	3092	3610	245	3	1	3	3	1	-1	1
204	97	-307	911	1353	2093	2400	3160	3748	192	-3	-1	1	1	-1	-1	1
1125	98	-170	744	1427	1885	2400	3082	3589	249	3	1	3	3	1	3	1
461	98	-278	891	1403	1795	2400	3103	3640	227	-1	-1	1	3	1	3	3
707	99	-152	730	1420	1939	2400	3093	3608	255	1	-1	1	-1	-1	-1	-3
448	99	-263	910	1400	1835	2400	3132	3696	218	-1	-1	1	1	1	3	3
420	99	-192	834	1385	2092	2400	3138	3705	220	-1	-1	1	-1	-1	-1	-1
215	99	-312	966	1375	1923	2400	3154	3740	193	-3	-1	1	1	1	3	3
186	99	-272	918	1375	2133	2400	3143	3718	201	-3	-1	1	-1	-1	-1	-3
58	99	-286	733	1390	1794	2400	3123	3673	233	-3	-3	-1	-1	-1	1	-1
436	100	-321	1004	1369	1942	2400	3139	3705	217	-1	-1	1	1	-1	-1	-1
419	100	-222	907	1383	2107	2400	3138	3704	220	-1	-1	1	-1	-1	-1	-3
50	100	-239	621	1381	2053	2400	3121	3667	234	-3	-3	-1	-1	-3	-1	-3
35	100	-216	657	1380	2055	2400	3129	3682	235	-3	-3	-1	-3	-3	-3	-3

Table 6.12 Selection of design points with respect to a harmonic bell; all weighting factors are equal to one

CAND. POINT NO.	S	FREQUENCY RATIOS IN CENTS 2-1	2-2	3-1	3-2	4-1	5-1	6-1	FD	DESIGN VARIABLES X1	X2	X3	X4	X5	X6	X7
452	69	-333	1325	1401	1855	2400	3126	3687	220	-1	-1	1	1	3	3	1
453	71	-321	1273	1405	1819	2400	3125	3686	220	-1	-1	1	1	3	3	3
449	72	-356	1498	1382	1965	2400	3130	3692	220	-1	-1	1	1	3	1	-1
450	73	-346	1424	1385	1927	2400	3130	3693	220	-1	-1	1	1	3	1	1
462	74	-327	1932	1393	2117	2400	3091	3624	220	-1	-1	1	3	3	1	-1
85	74	-307	1564	1411	1891	2400	3060	3566	235	-3	-3	-1	1	3	1	-1
61	74	-342	1206	1384	1856	2400	3123	3673	234	-3	-3	-1	-1	1	-1	-3
898	75	-353	1540	1378	1890	2400	3141	3717	204	1	1	3	3	3	3	3
86	75	-299	1487	1416	1841	2400	3061	3575	234	-3	-3	-1	1	3	1	1
64	75	-323	1040	1395	1767	2400	3121	3672	234	-3	-3	-1	-1	1	1	-1
463	76	-322	1747	1392	1975	2400	3098	3635	220	-1	-1	1	3	3	1	1
220	76	-372	1312	1371	1910	2400	3150	3733	195	-3	-1	1	1	3	3	3
62	76	-334	1133	1385	1829	2400	3123	3675	234	-3	-3	-1	-1	1	-1	-1
897	77	-373	1643	1368	1946	2400	3143	3719	204	1	1	3	3	3	3	1
457	77	-331	1403	1384	1797	2400	3112	3655	224	-1	-1	1	3	1	1	-1
454	77	-338	1622	1381	1948	2400	3113	3658	224	-1	-1	1	3	1	-1	-3
451	77	-322	1305	1390	1902	2400	3129	3691	220	-1	-1	1	1	3	1	3
219	77	-385	1365	1367	1944	2400	3150	3733	195	-3	-1	1	1	3	3	1
87	77	-297	1331	1413	1771	2400	3075	3590	235	-3	-3	-1	1	3	1	3
77	77	-304	1651	1406	1952	2400	3082	3616	233	-3	-3	-1	1	1	-1	-3
65	77	-314	987	1397	1736	2400	3121	3673	234	-3	-3	-1	-1	1	1	1
458	78	-322	1223	1388	1758	2400	3110	3653	225	-1	-1	1	3	1	1	1
455	78	-331	1365	1383	1890	2400	3112	3655	224	-1	-1	1	3	1	-1	-1
445	78	-355	1280	1371	1801	2400	3139	3709	216	-1	-1	1	1	1	1	1
78	78	-304	1464	1404	1844	2400	3089	3623	233	-3	-3	-1	1	1	-1	-1

Table 6.13 Selection of design points with respect to a harmonic bell; the weighting factors are equal to one for the frequencies 2-1, 3-1 and 4-1 and are equal to zero for the other frequencies

6.3.3.2 Design of three new bells

Frequency 3-1 is crucial in bell design. The range of variation of its frequency ratio for the whole set of candidate points appeared to be:

$$1322 \text{ cents} \leq \text{ratio } 3\text{-}1 \leq 1753 \text{ cents} \qquad (6.3.15)$$

This means that in this design area a fifth-bell (see Table 6.11) is not likely to be found. Furthermore, only a few musically interesting frequency ratios lie within this range: 1500 cents, 1600 cents, and a musically less interesting, 1700 cents.

We therefore ordered the candidate points according to increasing values of ratio 3-1 and selected three typical design points; Table 6.14 gives the frequency ratios for these points; the profiles have been plotted in Fig. 6.20.

Cand. point No.	Frequency ratio in cents						
	2-1	2-2	3-1	3-2	4-1	5-1	6-1
754	12 (0)	1652 (1600)	1609 (1600)	1892 (1900)	2400 (2400)	3050 (3100)	3598 (3600)
943	406 (400)	1171 (1200)	1606 (1600)	2389 (2400)	2400 (2400)	3055 (3100)	3599 (3600)
1165	674 (700)	1241 (1200)	1628 (1600)	2376 (2400)	2400 (2400)	3049 (3100)	3582 (3600)

Table 6.14 Frequency ratios and target values of three peculiar bell profiles.

The bells are peculiar because in each of them two of the lowest five frequencies coincide. In common bell design this is very unusual. Futhermore, for two of the three bells the frequency 2-1 is raised considerably compared to the conventional hum note.

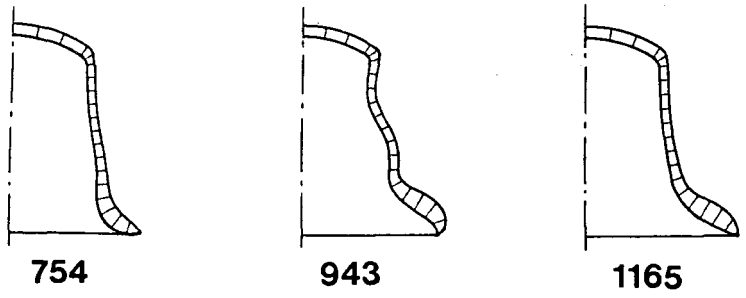


Fig. 6.20 Three new bell types

For the optimization the program DYNOPT, based on the SLP algorithm (see Section 2.5.3), was applied. For each bell about 15 iteration steps were needed. Table 6.15 gives the results after optimization and after the bells were founded by the Royal Bell-Foundry Eijsbouts. The largest diameter of the bell "1165" is 460 mm and the bell has been tuned by means of a set of special tuning curves, which were also computed using the program DYNOPT. Table 6.15 gives the results of the tuned bell. The largest diameters of the bells "754" and "943" are respectively 387 mm and 437 mm. These bells have been founded without applying a tuning tolerance and hence they have not been tuned.

Bell type No.	Frequency ratios in cents						
	2-1	2-2	3-1	3-2	4-1	5-1	6-1
"754" DYNOPT	4	1612	1600	1880	2400	3057	3608
Founded	-31	1635	1582	1883	2400	3057	3605
"943" DYNOPT	405	1197	1598	2399	2400	3064	3611
Founded	407	1221	1595	2405	2400	3047	3620
"1165" DYNOPT	700	1199	1613	2401	2400	3056	3593
Founded + tuned	713	1199	1600	2406	2400	?	?

Table 6.15 Results of optimized and founded bells (? indicates that the frequency ratio has not been measured)

We will call bell 1165 a major chord bell, because the lowest three partials give the notes g - c - e, which gives the experience of a major chord since the pitch of the bell lies on c. In the bell-foundry this bell is called the "Mexican hat".

At this moment only a preliminary musical opinion can be expressed about these bells. The bells are of increasing interest in the sequence 1165-943-754.

In the major chord bell (1165) the strike note is not very distinct. The listener hesitates between the strike note at 1200 cents and another one coinciding the strong major-third 3-1. This is caused by the hum which is raised with a fifth interval. Furthermore, the hum is much weaker then usually is the case. Due to these two phenomena the strike note is but little supported by the hum and hence becomes less distinct.

Bell 943 is a variant of the major-third bell. Although most people will think it ugly, the sound is quite reasonable and certainly better then that of the "Mexican hat". Here, due to the raised hum the strike note is also less distinct, compared to common bells, but it is better than with bell 1165.

Bell 754 is also a variant of a major-third bell. One might call this bell a TTNP-bell, i.e. "Two Thirds and No Prime", where

"prime" is a synonym of "fundamental". Although in this bell the fundamental has gone, the strike note is very distinct at the common location. Furthermore, it is a clear, transparently sounding bell and it is judged as being very interesting by the experts.

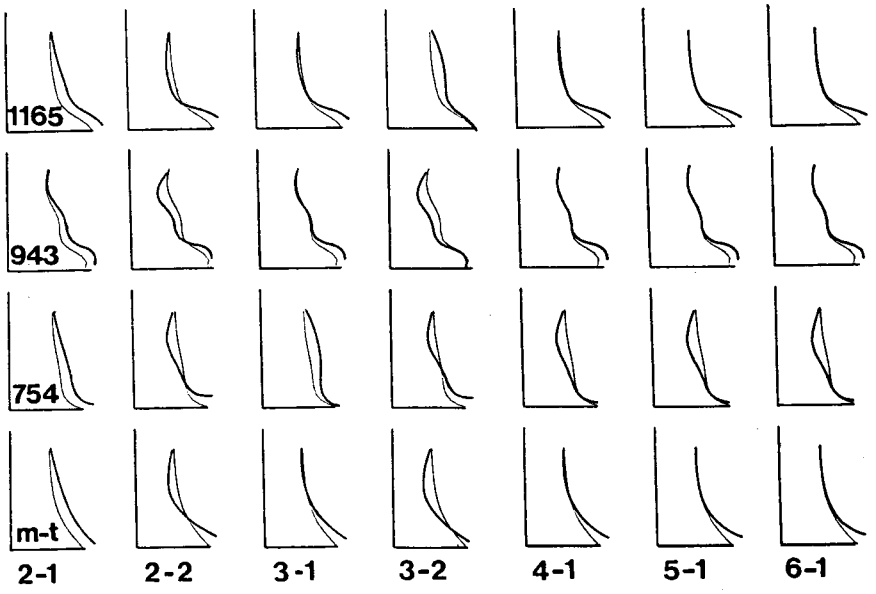


Fig. 6.21 Vibration modes of a minor-third bell and of three new bells

Fig. 6.21 shows the corresponding vibration modes of a minor-third bell and of the three new bells. In general the modes of the bells 943 and 1165 resemble those of the minor-third bell. The same thing can be said of the modes 2-1 and 2-2 of bell 754. However, the modes 3-1 through 6-1 are quite different; the modes 4-1, 5-1 and 6-1 have a node at the sound bow, whereas common bells there show a ventral segment. Considering these vibration modes, bell 754 should be preferably struck in the waist instead of at the sound bow. The future will learn the value of this bell. In any case, the shape of the bell is less open to question than that of the bells 943 or 1165.

6.4 Conclusions and suggestions

We may conclude that structural optimization using direct FEM-analyses and experimental design techniques offers efficient tools in bell design. They enabled us to design the challenging major-third bell, which was chased by the bell-founders for more than half a century. The major-third bell and carillons built of these bells, earn much success.

The development of a polynomial model of the bell by means of planning FEM-analyses using experimental design techniques and regression analysis, was not wholly successful. The developed geometrical model proved applicable in a wide design area. For four of the seven most important frequencies (3-1, 4-1, 5-1 and 6-1) satisfying regression models could be derived, although with the use of data collected in all candidate points. However, we expect that for these frequencies adequate models can be obtained with much fewer data. The regression models for the frequencies 2-1, 2-2 and 3-2 do not reach the predictive ability which is needed in bell design. We suspect that the regression models can be improved by choosing more adequate regressor variables (polynomial or otherwise) in the model.

Although the predictive ability of the models for the frequencies 2-1, 2-2 and 3-2 is too low for bell design, these apparently difficult relationships have been estimated with an accuracy which certainly is satisfactory in many branches of mechanical engineering. The regression models are extremely fast. Computing the seven frequencies of a bell using regression models with 109 parameters requires 7 * 0.02 = 0.14 sec. of computing time. A FEM-computation of these frequencies requires 1200 sec. CPU on the same computer, that is 8500 times 0.14 sec.

And what about all those finite element computations in the candidate points? In any case these results can be used to derive regression models in interesting subareas of the design area. For example, selection of design points with variables in ranges from -1.0 to +1.0 results in a 2^7 experimental design in a relatively small area. The results of Maas (1985) showed that regression models derived in such an area, are sufficiently accurate for bell design. However, such procedures have little to do with (optimal) experimental design, but more with mere regression analysis.

The collected data is used to search for new bell types. A bell with harmonic partials and a fifth bell (Table 6.11) could not be found in the design region occupied by the candidate points. The design of a harmonic bell is a challenge; whether a solution might be possible is hardly to say at this moment.

Three new bell types were found. Each of these bells is peculiar in the sense that two of the five lowest frequencies coincide. Such frequency spectra are new in bell design. Musical evaluation of these bells has yet to be carried out. At this moment the bells "1165", "943" and "754" may be characterized as musically of little interest, of moderate interest and of great interest, respectively.

Applying large profile variations, the frequency ratios and vibration modes can change very much, far beyond the changes that appear in common bell design. Therefore new classifications of partials should be developed, and new insights must be obtained as to how bells with completely new sound spectra will be experienced by the listeners. This requires investigation of all aspects of the sound spectrum of bells (frequencies, vibration modes and strength, internal and external damping and sound radiation).

We may conclude this chapter posing that, due to the application of shape optimization, a new landmark has been reached in bell design.

CHAPTER 7: SUMMARY AND CONCLUSIONS

In this thesis the integration of experimental design techniques in structural optimization and related topics has been discussed. The arguments for such an integration are:

- The need for fairly accurate and fast mathematical models of structures. Such models can be derived by the planning and analysis of a set of comprehensive numerical experiments (for example, carried out by means of a finite element package).
- The concepts of design variables in structural optimization and control variables in experimental design are closely related.
- Structural optimization programs are suitable to collect data for numerical experimental designs.
- Regression models can be used as fast analysis modules in optimization programs. A solution of the optimization can serve as an initial starting point in a final optimization run using direct FEM-computations.

Experimental designs can be formulated in several ways. In the more classical approach Latin squares and p^n-designs are used. Here, the objective is to minimize the confounding in the design, in order to obtain uncorrelated parameter estimates and a sufficient predictive ability of the regression models.

In optimal experimental design the variance of the parameter estimators and/or the response estimators is minimized. Methods and criteria are described in Chapter 3, and have been generalized to the case of simultaneous observations of several responses. This feature is applicable in the derivation of regression models in case observations of function values and partial derivatives of the response quantity with respect to the control variables are available. Computation of the required partial derivatives can be carried out efficiently in FEM-formulations.

The guidelines for model building based on physical experiments can also be used in building models based on numerical experiments.

Regression models based on numerical computations can be applied in different ways. They can serve as efficient stand-alone mathematical models of a structure or system. Such models can be evaluated on a microcomputer or even on a pocket calculator. One can think of a library of regression models for use in design offices or in education. Besides derivation of models for more or less standardized structures or machine parts, models can also be derived for unique physical problems, in order to carry out parameter studies.

The derived methods have been applied successfully on several mechanical engineering problems. We recall a pin and hole joint, a journal bearing, cross-section design of aluminium beams and a leaflet heart valve prosthesis.

The methods described in this thesis have been tested extensively and were used in shape optimization of church and carillon bells. The major-third bell, which was searched for by bell-founders for more than half a century, has been found and appeared to be a great success.

As a final application, especially of optimal experimental design, we tried to build regression models for bell design which had to be applicable in a relatively wide design area. Only for four of the seven relevant frequencies satisfying regression models could be derived. Additional research will be necessary to improve the three remaining models. The outcome of such investigations might be that our objectives for the general bell model were too ambitious. Nevertheless model building procedures could be tested thoroughly. The program CADE proved to be a very valuable tool in experimental design. Our first use of the program was in numerical experimentation, but application to physical experimentation is standard.

Although a satisfying general bell model was not achieved, the collected FEM-data could be used to find new bells. It follows that in the design area exploited by us, a bell with harmonic partials and a fifth bell probably will not be found. Solving these problems seems to be very difficult.

Initial designs were found for three typical bells. In the frequency spectra of each of these bells, two of the lowest five frequencies coincide. The three bells were optimized to target frequency spectra by means of the program DYNOPT, applying direct FEM-computation in the analysis step.

Application of numerical shape optimization of bell profiles has opened new ways in designing musically interesting bells.

References

Aerts, G.F.M. (1979).
"Parameter study of pin and hole joints based on analyses via the finite element method" (in Dutch). Report WFW 79.05, Univ. Press, Eindhoven.

Asperen, F.G.J. van (1984).
"Optimization of the eigenfrequencies of axi-symmetric constructions, applied to a swinging bell or a carillon bell" (in Dutch). Report WFW 84.012, Univ. Press, Eindhoven.

Avriel, M. (1976).
"Nonlinear Programming: Analysis and Methods". Prentice-Hall Englewood Cliffs, N.J.

Banens, J. (1972).
"The frequency spectrum of the bell. Determination of the stiffess matrix of the Triax6-element. Some methods for determining the largest eigenvalues of band matrices" (in Dutch). Report WFW, Univ. Press, Eindhoven.

Banens, J. and Peters, F. (1976).
"Femsys, a system for calculations based on the finite element method" (in Dutch). Univ. Press, Eindhoven.

Banens, J. (1981).
"COIN, A Command Interpreting program, User's Manual". Report Departm. of Mech. Eng., Eindhoven Univ. of Techn. (unpublished).

Blessing, J. (1894-1897).
"Ueber Glocken und ihre Musik". Gregorius-Blatt, Vol. 19, 20, 21 and 22.

Boom, J. v.d. (1979).
"The nonlinear behaviour of a parabolic spring support; annex 1: FEMLAN-procedure WRIJVING" (in Dutch). Report WFW 79.04, Univ. Press, Eindhoven.

Box, G.E.P., and Draper, N.R. (1959).
"A Basis for the Selection of a Response Surface Design". Journal of the American Statistical Association, 54, pp. 622-653.

Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978).
"Statistics for Experimenters". John Wiley, New York.

Carpenter, W.C., and Smith, E.A. (1977).
"Computational efficiency of nonlinear programming methods on a class of structural problems". Int. J. Num. Meth. Engng., 11, pp. 1203-1223.

Cochran, W.G., and Cox, G.M. (1950).
"Experimental Designs". John Wiley, New York.

Dantzig, G. (1963).
"Linear Programming and Extensions". Princeton University Press, Princeton, N.J.

Davidon, W.C. (1959).
"Variable Metric Method for Minimization". Argonne Natl. Lab. ANL-5990 Rev., Univ. of Chicago.

Davies, A.J. (1980).
"The Finite Element Method: a First Approach". Clarendon Press, Oxford.

Doornbos, R. (1984).
"Calculus of probability and statistics 2" (in Dutch). Syll. No. 2.319.O, Univ. Press, Eindhoven.

Draper, N.R., and Smith, H. (1981).
"Applied Regression Analysis". John Wiley, New York.

Efroymson, M.A. (1960).
"Multiple Regression Analysis". In Ralston, A., and Wilf, H.S.: "Mathematical Methods for Digital Computers". John Wiley, New York (1960).

Euler, L. (1764).
"Tentamen de sono companarum (Novi Commentarii academiae scientiorum imperalis Petropolitanae". Tom. X. pro anno MDCCLXIV, pp. 261-281.

Fedorov, V.V. (1972).
"Theory of Optimal Experiments". Academic Press, New York.

Fletcher, R., and Powell, M.J.D. (1963).
"A Rapidly Convergent Descent Method for Minimization". Computer J. 6(2), pp. 163-168.

Fletcher, R., and Reeves, C.M. (1964).
"Function Minimization by Conjugate Gradients". Br. Comp. J. vol. 7, no. 2, pp. 149-154.

Fox, R.L. (1971).
"Optimization methods for engineering design". Addison-Wesley, Reading Mass.

Gill, P.E., and Murray, W. (eds) (1974).
"Numerical Methods for Constrained Optimization". Academic Press.

Heck, J.G.A.M. van (1984).
"On the dynamic characteristics of slideways". Ph. D. Thesis, Eindhoven Univ. of Techn., Wibro Press, Helmond (Neth.).

Houtsma, A.J.M., and Tholen, J.G.M. (1987).
"A carillon of major-third bells II: A perceptual evaluation". Music Perception Vol. 4, No. 3, Spring 1987, Univ. Press, San Diego, California.

Imai, K. (1978).
"Configuration Optimization of Trusses by the Multiplier Method". Ph. D. Thesis, Univ. of California, Los Angelos.

Kiefer, J., and Wolfowitz, J. (1959).
"Optimum Designs in Regression Problems". Canadian J. Math., 12, pp. 363-366.

Kirsch, U. (1981).
"Optimum Structural Design". McGraw-Hill, New York.

Kline, S.J. (1965).
"Similitude and Approximation Theory". McGraw-Hill, New York.

Krishnamurthy, N., Huang, H-T., Jeffrey, P.K. (1979).
"Analytical M-O Curves for End-Plate Connections". Transportation Eng., J. of ASCE, Vol. 105, No. 1, pp. 133-145.

Langhaar, H.L. (1960).
"Dimensional analysis and theory of models". John Wiley, New York.

Lehr, A. (1986).
"Partial groups in the bell sound". The Journal of the Acoustical Society of America, Vol. 79, No. 6, pp. 2000-2011.

Lehr, A. (1987a).
"Designing swinging bells and carillon bells in the past and nowadays". Nationaal Beiaardmuseum, Asten, The Netherlands.

Lehr, A. (1987b).
"A carillon of major-third bells III: From theory to practice". Music Perception Vol. 4, No. 3, Spring 1987, Univ. Press, San Diego, California.

Maas, P.J.J. (1985).
"Investigation of the geometry of a major-third bell" (in Dutch). Report WFW 85.027, Univ. Press Eindhoven.

MARC (1984).
"MARC General Purpose Finite Element Program, Vol. A, B, C, D." MARC Analysis Research Corp. Palo Alto, California.

Marguardt, D.W. (1970).
"Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation". Technometrics, 12, pp. 591-612.

Menken, C.M., and Pasch, J.A.M. v.d. (1986)
"The calculation of geometrical constants for irregular cross-sections of rods and beams". Proceedings of the Second Int. Conf. on Microcomputers in Engineering, Swansea U.K. 7th-10th April, 1986, pp. 75-86.

Miura, H., and Schmit, L.A. (1979).
"NEWSUMT -a Fortran program for inequality constrained function minimization- users guide". NASA CR-159070.

Montgomery, D.C., and Peck, E.A. (1982).
"Introduction to linear Regression Analysis". John Wiley, New York.

Montgomery, D.C. (1984).
"Design and Analysis of Experiments". John Wiley, New York.

Morris, A.J., Editor (1982).
"Foundations of Structural Optimization: A Unified Approach". John Wiley, New York.

Murray, W. (1976).
"Methods for Constrained Optimization", in L.C.W. Dixon (ed.), "Optimization in Action", Academic Press.

Nagtegaal, R. (1987).
"Computer Aided Design of Experiments. A program for experimental design and model building" (in Dutch). Report WFW 87.005, Univ. Press Eindhoven.

N.B.S. (1957).
"Fractional Factorial designs at two levels". Applied Mathematics series, Report 48.

Pasch, J. v.d. (1985).
"Development of utilities for optimization of beams with complicated cross-sections" (in Dutch). Report WFW 85.002, Univ. Press, Eindhoven.

Patterson, H.D. (1976).
"Generation of Factorial Designs". Journal of the Royal Statistics Society, Ser. B, 38, pp. 175-179.

Perrin, R., and Charnley, T. (1983).
"Normal modes of the modern English church bell". Journal of Sound and Vibration, 90 (1), pp. 29-49.

Peterson, R.E. (1974).
"Stress Concentration Factors". John Wiley, New York.

Pichumani, R. (1975).
"Parametric studies of layered pavement systems using AFPAV code with fast equation solver". New Mexico University, Report No. AD-A 009 479.

Powell, M.J.D. (1964).
"An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives". Computer J., vol. 7 no. 4, pp. 303-307.

Ragsdell, K.M. (1984).
"The Utility of Nonlinear Programming Methods for Engineering Design". in Atrek, E et al. (eds.): "New Directions in Optimum Structural Desing". John Wiley, New York.

Rousseau, E.P.M. (1985).
"Mechanical specifications for a closed leaflet valve prosthesis". Ph. D. Thesis, Eindhoven Univ. of Techn., Wibro Press, Helmond (Neth.).

Sandgren, E. (1977).
"The utility of nonlinear programming algorithms". Ph.D. Dissertation, Perdue University, Univ. Microfiche, Ann Arbor, Michigan, Document No. 7813115.

Schittkowski, K. (1980).
"Nonlinear Programming Codes: Information, Tests, Performance". Lecture Notes in Economics and Mathematical Systems, Vol. 183, Springer Verlag, Berlin.

Schmit, L.A., and Miura, H. (1975).
"A New Structural Analysis/Synthesis Capability - ACCESS 1". AIAA/ASME/SAE Sixteenth Conference, Denver, Colorado.

Schmit, L.A., and Miura, H. (1976).
"An Advanced Structural Analysis/Synthesis Capability- ACCESS 2. AIAA/ASME/SAE Seventeenth Conference, King of Prussia, Penn.

Schmit, L.A., and Fleury, C. (1979).
"An Improved Analysis/Synthesis Capability Based on Dual Methods - ACCESS 3". AIAA/ASME/ASCE/AMS Twentieth Conference, St. Louis, Mo.

Schoofs, A.J.G. (1984).
"The combination of the finite element method and statistical analysis; a literature survey" (in Dutch). Internal report WFW, Eindhoven Univ. of Techn.

Schoofs, A.J.G. (1985).
"Eindhoven Univ. of Technology and Eijsbouts design first major-third bell" (in Dutch). I^2-werktuigbouwkunde, No. 10, pp. 39-47.

Schoofs A., Asperen, F. v., Maas, P., and Lehr, A. (1987).
"A carillon of major-third bells I: Computation of bell profiles using structural optimization". To be published in Music Perception Vol. 4, No. 3, Spring 1987, Univ. Press, San Diego, California.

SDRC (1986).
"Supertab Engineering Analysis, Model Solution and Optimization, User's Guide, I-DEAS, Level 3". CAE Intrnational, Milford, Ohio.

Smith, K. (1918).
"On the Standard Deviations of Adjusted and Interpolated values of an Observed Polynomial Function and its Constants and the Guidance They Give Towards a Proper Choice of the Distribution of Observations". Biometrica, 12, pp. 1-85.

Vanderplaats, G.N. (1973).
"CONMIN -A Fortran program for constrained function minimization: User's manual". NASA TM X-62282.

Vanderplaats, G.N., Sugimoto, H., and Sprague, C.M. (1983).
"ADS-1: a new general-purpose optimization program". AIAA Paper No. 83-0831, presented at the AIAA/ASME/ASCE/AHS 24th Structures, Structural Dynamics & Materials Conf., Lake Tahoe, Nevada.

Vanderplaats, G.N. (1983).
"A Robust Feasible Directions Algorithm for Design Synthesis". Proceedings, AIAA/ASME/ASCE/AHS, 24th Structures, Structural Dynamics & Materials Conf., Lake Tahoe, Nevada.

Vanderplaats, G.N. (1984).
"Numerical optimization techniques for engineering design". McGraw-Hill, New York.

Vas Nunes, A. (1909).
"Experimental investigation of bells of F. Hemony" (in Dutch). Doctor's Ph. D. thesis, Amsterdam.

Welch, W.J. (1982).
"Branch and Bound Search for Experimental Designs Based on D-Optimality and Other Criteria". Technometrics, 24, pp. 41-48.

Welch, W.J. (1983).
"A Mean Squared Error Criterion for the Design of Experiments". Biometrica, 70, pp. 205-223.

Welch, W.J. (1985).
"Algorithm for the Construction of Experimental Designs; User's Manual ACED, Version 1.6.1". Univ. of British Columbia, Vancouver, Can.

Widdershoven, T.C.H. (1981).
"Optimization of constructions" (in Dutch). Report WFW 81.10, Univ. press, Eindhoven.

Wouters, H. (1986).
"The stress distribution in dry running journal bearings" (in Dutch). Report WFW 86.012, Univ. Press, Eindhoven.

Zienkiewicz, O.C. (1977).
"The Finite Element Method". McGraw-Hill. U.K.

Samenvatting

Bij het numeriek optimaliseren van het mechanisch gedrag van konstrukties wordt meestal gebruik gemaakt van twee belangrijke methoden nl.:

1. de eindige elementenmethode (EEM) als flexibel en nauwkeurig modelerings- en analysegereedschap en
2. mathematische programmering als gestruktureerde werkwijze voor het zoeken van een gunstiger set van ontwerpvariabelen.

De boven aangeduide optimaliseringsproblemen zijn vrijwel steeds sterk niet-linear, hetgeen een iteratieve oplossingsmethode vereist: men analyseert de konstruktie en bepaalt vervolgens een schatting voor betere waarden van de ontwerpvariabelen. Dit proces wordt voortgezet totdat een, i.h.a. lokaal optimum is bereikt. Het eindresultaat van deze werkwijze is wat ons interesseert; de tussenliggende iteratiestappen zijn achteraf van weinig belang.

Indien men ervan uitgaat dat voor het oplossen van het optimaliseringsprobleem meerdere EEM-analyses nodig zijn, is ook een andere aanpak mogelijk. Bij die aanpak worden de uit te voeren EEM-analyses vooraf gepland, zowel wat betreft hun aantal als de waarden van de ontwerpvariabelen welke een rol spelen. De resultaten van de uitgevoerde analyses worden m.b.v. regressie-analyse "gecondenseerd" tot een rekenmodel in de vorm van één of meer lineaire modellen, bijvoorbeeld polynomen.

Voor het plannen en verwerken van de EEM-analyses kan met vrucht gebruik worden gemaakt van de statistische theorie van proefopzetten, welke is ontwikkeld ten behoeve van het plannen van omvangrijke fysische experimenten. EEM-analyses zijn te beschouwen als numerieke experimenten, welke kunnen dienen voor het formuleren van een numerieke proefopzet. Het resultaat van zo'n numerieke proefopzet is een efficient rekenmodel van de onderzochte konstruktie. Dergelijke rekenmodellen kunnen als zodanig gebruikt worden, bijvoorbeeld op een ontwerpafdeling. Voor ons is vooral de toepassing als analysemoduul in optimaliseringsprogrammatuur van belang.

In dit proefschrift worden werkwijzen behandeld zoals die gebruikelijk zijn bij de numerieke optimalisering en in de statistische theorie van proefopzetten. Verder worden modificaties van deze theorie behandeld welke nodig en nuttig zijn ten behoeve van numerieke proefopzetten en worden argumenten gegeven voor de integratie van deze theorie in optimalisering.

De mogelijkheden voor het gebruik van de ontwikkelde methoden worden afgebakend en er worden richtlijnen gegeven voor de toepassing.

De ontwikkelde werkwijzen zijn uitgebreid getest en gebruikt bij het optimaliseren van de geometrie van luid- en carillonklokken, met als resultaat enkele nieuwe, muzikaal zeer interessante kloktypen. Verder zijn de methoden met succes toegepast bij diverse werktuigkundige problemen en bij een biomechanica-probleem.

Nawoord

Aan het tot stand komen van dit proefschrift hebben veel mensen een bijdrage geleverd. Langs deze weg wil ik hen van harte bedanken.

Allereerst gaat mijn dank uit naar Anneke, Paul, Stan en Ivo voor hun begrip en geduld en de bereidwillige overname van allerlei klussen in en om ons huis. Bovendien controleerde Anneke nauwgezet het typewerk van dit proefschrift en Paul verzorgde de figuren.

Het onderzoek werd verricht in de vakgroep Fundamentele Werktuigkunde (WFW) van de Technische Universiteit Eindhoven (TUE). Alle vakgroepleden wil ik bedanken voor de prettige werksfeer in de groep; extra dank geldt hen die tijdelijk een deel van mijn vakgroepstaken voor hun rekening namen.

Jan Janssen nam het initiatief tot dit onderzoek. Ik ben hem erkentelijk voor de faciliteiten en de vriendelijke dwang om dit onderzoek in gang te kunnen zetten. Dick van Campen dank ik voor zijn stimulerende begeleiding bij het onderzoek en de inhoudelijke suggesties bij het schrijven van het proefschrift. Ook de vele kritische kanttekeningen en inhoudelijke bijdragen van Frans Veldpaus dwongen bewondering af. Een lange rij van enthousiaste afstudeerders heeft mij zeer veel geholpen. Ongeveer in chronologische volgorde zijn zij: Gé Aerst, Crétien Widdershoven, Frans van Asperen, Jan van de Pasch, Peter Maas, Huub Wouters, Arie Schepens, Rob Nagtegaal, Leon Levasier en Jan van Beukering. Zonder hen had dit proefschrift onmogelijk tot stand kunnen komen. Extra dank ben ik verschuldigd aan Rob Nagtegaal voor de assistentie ook nog na zijn afstuderen en voor de kritische opmerkingen en waardevolle suggesties bij meerder delen van dit proefschrift.

Lia Neervoort verdient alle lof voor de manier waarop zij het typewerk snel en nauwkeurig verzorgde en de eindeloze stroom correcties verwerkte.

Verder dank ik Prof. Doornbos en Emile van Berkum van de Vakgroep Stochastiek en Besliskunde van de Faculteit Wiskunde en Informatica van de TUE voor hun steun en adviezen bij het onderzoek.

Aad Houtsma van het Instituut voor Perceptie Onderzoek leverde waardevolle adviezen en was erg behulpzaam bij het realiseren van een publikatie.

Ook de bijdrage van oud-WFW-medewerker Jos van Heck, in de vorm van persoonlijke gesprekken over het onderwerp, mag niet onvermeld blijven.

Dhr. J.L.M. Dietvorst dank ik voor de snelle en accurate correctie van het gebezigde engels.

Tot slot, maar niet in de laatste plaats, wil ik André Lehr van de Koninklijke Eijsbouts Klokkengieterij B.V. danken voor het stimulerende enthousiasme waarmee hij aan het onderzoek meedeed, kritiek leverde en oplossingen suggereerde. De gesprekken die ik met hem mocht hebben, zijn voor mij erg waardevol geweest. Hij was bovendien bereid om, zo ongeveer op afroep, meerdere van mijn soms bizarre klokontwerpen in brons te verwerkelijken.

Levensbericht

15-10-1940	geboren te Bladel.
1953-1955	L.T.S. te Bladel, Afdeling Machine-Bankwerken.
1955-1959	Katholieke MTS te Eindhoven, Afdeling Werktuigbouwkunde.
1959-1962	Avondopleiding Nijverheids Opleiding te Eindhoven, Akte N IV.
1959-1967	Constructeur bij Hydraudyne b.v. te Boxtel.
1962-1965	H.B.S.-B aan de Avond HBS te Eindhoven.
1965-1971	Studie aan de Technische Universiteit Eindhoven, Faculteit Werktuigbouwkunde.
1971-	Universitair Docent aan de Technische Universiteit Eindhoven, Faculteit Werktuigbouwkunde.

STELLINGEN

Behorende bij het proefschrift

EXPERIMENTAL DESIGN AND STRUCTURAL OPTIMIZATION

1. Omvangrijke numerieke berekeningen waarin parametervariaties een rol spelen, kunnen zinvol en efficient gepland worden door gebruik te maken van werkwijzen uit de statistische theorie van proefopzetten.
   - Dit proefschrift.

2. Door middel van wijziging van een klokprofiel is het niet mogelijk om de laagste twee eigenfrekwenties van de trillingsmodes met twee meridiaan-knooplijnen te doen samenvallen.
   - Dit proefschrift, hoofdstuk 6.

3. De bewering van G.E.P. Box, W.G. Hunter en J.S. Hunter dat de theorie van optimale proefopzetten leidt tot "oversimplification" van de experimentele situatie, is in zijn algemeenheid niet juist.
   - G.E.P. Box, W.G. Hunter, J.S. Hunter, "Statistics for Experimenters", blz. 304, John Wiley, New York (1978).

4. Technische opleidingen vragen om twee categorieen van applicatieprogrammatuur:
   - in de basis: onderwijs gerichte programma's welke zeer snel en "student-proof" zijn en waarin "gestoeid" kan worden met principes, methoden en relatief eenvoudige problemen.
   - in de afsluitende fase: geavanceerde praktijkgerichte programma's voor het modelleren en oplossen van realistische problemen.

5. De in computerspelletjes gebruikte principes voor het stimuleren van de speler dienen ook toegepast te worden bij de ontwikkeling van educatieve programmatuur.

6. Bij handbellen bestaat het probleem dat de grotere bellen relatief te weinig geluid afstralen*). De geluidsafstraling kan verbeterd worden door het klokprofiel zodanig te ontwerpen dat de frekwentieverhoudingen van de niet gestemde partialen zoveel mogelijk elementen vormen uit de in muzikaal opzicht harmonische reeks: 1 : 2 : 3 : enz.
   *) T.D. Rossing and R. Perrin, "Vibration of Bells". Applied Acoustics. Vol. 20. No 1, pp. 41-70 (1987).
   - Dit proefschrift.

7. De publiciteit rond tamelijk specialistische onderzoekprojecten kan sterk bevorderd worden door de keuze van een tot de verbeelding sprekend probleem bij het testen van onderzoekmethoden.

8. Muziekinstrumenten van allerlei soort vormen een uitgebreide klasse van interessante konstrukties met betrekking tot het optimaliseren van hun dynamisch gedrag.

9. In de regelgeving met betrekking tot het verlenen van vergunningen voor het organiseren van lawaai-intensieve evenementen, wordt de modelvliegsport achtergesteld. Dit is des te opmerkelijker daar juist in de modelvliegerij relatief veel moeite gedaan wordt om geluidoverlast te beperken.
   - KB 13 november 1986, nr. 26
   - KB 9 oktober 1986, nr. 109
   - L.F. Doorduijn, "Jurisprudentie", Geluid en Omgeving, maart 1987, blz. 33-35.

10. Uit didactisch oogpunt verdient het aanbeveling op basisscholen en in het voortgezet onderwijs slechts die zakrekenmachines toe te staan, waarbij op het display de decimale punt niet zichtbaar is.