A Response Surface Approach to Noise Optimization of Engine Structures

R.A. Hall

Submitted for the degree of Doctor of Philosophy in the Department of Transport Technology, Loughborough University, 1993.

Synopsis

The work presented within this thesis concerns the optimization of finite element models of engine structures to reduce radiated noise. For many engineering problems, current methods of structural optimization provide an efficient means by which to identify an optimum design, subject to a set of imposed bounds and constraints. They do not, however, have the flexibility to carry out efficient investigation of a range of different constraint criteria, and this is often a requirement of a noise optimization study.

In order to address this restriction, an alternative method of noise optimization is developed, which is based on the techniques of experimental design theory and response surface methodology. The main feature of this approach is that values of the response functions of interest are calculated at a number of selected points within the design variable space, from which an approximating mathematical model is generated. It is this analytical model of the original responses which is used as the basis of the optimization procedure.

Experimental design theory is employed in order to ensure that a sufficiently accurate model can be generated with the minimum number of function evaluations. A number of competing experimental designs and mathematical models are considered, and numerical trials are carried out to evaluate their performance in representing the noise function. A quadratic model is found to perform well throughout the design region, and can be estimated efficiently using a particular class of economic second-order designs.

A number of detailed noise optimization studies are presented, involving up to seven design variables, which illustrate the ways in which the requirements of the noise optimization problem can be met using the response surface approach.

Contents

Acknowledgements

The author owes an enormous debt of gratitude to his two supervisors, without whose advice and encouragement this work could never have been carried out. Firstly, to Dr. M.G. Milsted for his initial inspiration in seeing the potential of response surface methods, and for motivating the early investigations into their application. Secondly, but only chronologically, to Dr. T.J. Gordon, for ensuring that the statistical content of the work stands on a sound theoretical basis, and for suggesting many modifications and enhancements which have since become fundamental to the procedure.

Further thanks are due to Dr. T. Zhang and Mr. J.E.W. Ogendo for their work on the noise analysis procedure, and to Dr. P. Hanks, Mr. I.T. Martin and Mr. A. Haddock of Perkins Technology Ltd. for their technical support.

The work was funded by the Science and Engineering Research Council, and carried out in collaboration with Perkins Technology Ltd. The support of both is acknowledged with thanks.

Notation

The following represents the principal notation employed throughout the work. Additional notation is introduced and explained as required.

Symbol	Introduced	Meaning
c	Section 2.12	scaling factor
$\mathbf{D}$	Section 2.11	design matrix
E	Section 2.15	absolute lack of fit
E_R	Section 2.15	lack of fit expressed as a percentage of function range
F	Section 6.1	number of tests in factorial portion of CCD
F_h	Section 2.15	highest value of function within design region
F_l	Section 2.15	lowest value of function within design region
$\mathbf{I}$	Section 2.12	identity matrix
I_e	Section 3.5.2	identity effect
$\mathbf{J}$	Section 2.12	matrix of ones
$\mathbf{M}$	Section 2.11	moment matrix
M_i	Section 2.3	mean of the high and low bounds of variable i in original units
m_i	Section 2.3	dimensionless mean of the high and low bounds of variable i
N	Section 2.3	number of tests
n	Section 2.1.3	number of design variables
n_0	Section 6.1	number of centre point tests in CCD
$\mathbf{P}$	Section 5.2	transformation matrix
$\mathbf{X}$	Section 2.11	regressor matrix
X_i	Section 2.3	value of variable i in original units
x_i	Section 2.3	dimensionless value of variable i
$\mathbf{Y}$	Appendix 2A	vector of actual function value
Y	Appendix 2A	actual function value
y	Appendix 2A	predicted function value
α	Section 6.1	axial parameter of CCD
$\boldsymbol{\beta}$	Appendix 2A	vector of model coefficients
β	Section 2.5	model coefficient
δ	Section 2.12	order of design moment
Δ_i	Section 2.3	half-range of variable i
ε	Section 2.5	error term of statistical model
ε	Section 7.2	axial parameter of additional star portion of modified CCD
κ	Section 2.12	scaling factor
λ_2,λ_4	Section 2.12	scale factors in moment matrix
ρ	Section 5.3	radius from design centre
σ^2	Appendix 2A	variance

Preface

This work is arranged in 10 chapters. In Chapter 1, the aims and scope of the work are described, following which the general background to the noise optimization problem is introduced. A discussion of current approaches to structural optimization is presented, together with a brief survey of currently available software packages and an account of the use of a proprietary iterative optimization routine to perform noise optimization. The response surface approach to optimization is then outlined, and its advantages and disadvantages are discussed, together with restrictions on the use of the method.

A review of the development of response surface methods is given in Chapter 2, together with an introduction to a number of the concepts and definitions which are fundamental to the discussions of later chapters.

Chapters 3-8 discuss various experimental designs which may be employed in order to select the sample points at which the response functions are to be evaluated. Chapter 3 introduces first-order designs, which may be used to estimate the coefficients of linear main effects models and of models containing main effect and linear interaction terms. In Chapter 4, the results of experimental studies using first-order designs are presented, carried out using a concept level finite element model of a diesel engine cylinder block.

Chapter 5 provides a general introduction to second order models, in which quadratic terms are also represented, whilst Chapter 6 contains a detailed investigation into one particular second order design, the Central Composite Design (CCD). Numerical results using this family of designs are presented. In Chapter 7, modifications to the standard CCD are discussed, the aim of which is to address certain characteristics of the function evaluation process which are unique to the field of computer 'experimentation'. Chapter 8 contains a survey of economic second-order designs, which seek to fit a second-order model with the minimum number of test points. One of these strategies is chosen for more detailed investigation, and numerical trials conducted to assess its suitability to the present application.

Chapter 9 demonstrates the way in which the experimental designs introduced in previous chapters may be used within an optimization study. A number of examples are presented which illustrate the flexibility of the method in addressing different combinations of objective and constraint functions, and different levels of constraint. An example of the use of non-continuous variables is also presented.

Chapter 10 summarises the work presented in previous sections. Areas in which further work is required are highlighted.

1. Introduction

1.1 Research objectives

The principal aim of the work described in this thesis is to develop an approach to optimization of engine structures for minimum radiated noise which allows engine designers and analysts to make informed decisions as to the trade-off between competing design objectives. This will enable designers to select an engine specification which performs well when measured against a number of different criteria.

The motivation for seeking an alternative method of optimization is that currently available techniques, based on a direct iterative approach to optimization, do not provide the range of information required by engine designers, and make inefficient use of the computationally expensive finite element analyses which they perform. In particular, the following three characteristics of the standard direct iterative approach to optimization make it unsuitable for use in performing noise optimization of engine structures.

1. A separate optimization trial must be carried out for each combination of imposed constraint levels which are of interest. The consequent increase in computing requirements effectively prohibits the investigation of more than a few such alternative constraint scenarios.
2. The method is not robust with respect to the occurrence of local minima. A single trial cannot identify whether the optimum found is local or global.
3. The optimum design is generally sought with a higher level of precision than is appropriate to the characteristics of the noise analysis problem.

These issues are discussed further in Sections 1.4 to 1.8. An alternative approach is adopted, based on the theories of experimental design and response surface methodology, which has the potential to provide a far greater range of information.

1.2 Scope of work

The present work describes an investigation into the use of response surface methods to represent the variation in radiated noise of internal combustion engines with respect to changes in the structure of the engine. The aims of this investigation are to establish whether the variation of the noise function within an n-dimensional design variable space can adequately be

represented by a low order polynomial model, and to identify appropriate experimental designs which will facilitate this modelling process. A simplified mathematical model of the noise response surface is sought, which may be used as the basis of a structural optimization capability.

Optimization using the mathematical model of the original response surface is carried out using standard iterative techniques. Development of alternative methods of solving this final optimization problem is not addressed within the present work. Discussion of this issue is undertaken in Chapter 9.

The measure of radiated engine noise which is used as the objective function of the optimization problem is A-weighted sound power, summed from one-third octave band contributions in the range to 5 kHz. Evaluation of this function is carried out using a purpose-written computer code based on finite element analysis techniques. Detailed description of this program is not undertaken within the present work, but a summary of the theory underlying the noise analysis procedure, together with some notes on its development, is given in Appendix 1A. Further details are given by Ogendo and Zhang (1989), Zhang (1992) and Milsted, Zhang and Hall (1993).

A computer implementation of the response surface and optimization procedure which are discussed in this thesis are incorporated in the computer program *optrsm*, which is introduced in Appendix 1B. Further details of the capabilities and operation of this program, together with a number of numerical examples, are given in Chapter 9. See also Hall (1992). The numerical examples which are presented within the work have been conducted using a finite element model of a representative engine structure. This model is introduced and described in Appendix 1C.

The remainder of the present chapter is arranged as follows. Firstly, the general background to the noise optimization problem is introduced, with discussion of previous approaches to the subject (Section 1.3). In Section 1.4, the direct iterative approach to optimization is described. A survey of currently available software packages for carrying out structural optimization is presented in Section 1.5, with particular emphasis on reduction of radiated engine noise. This is accompanied by an account of the use of a proprietary iterative optimization routine to perform structural optimization of the finite element model mentioned above (Section 1.6). It is this exercise which provides the motivation for seeking an alternative noise optimization strategy, which is outlined in Section 1.7. The practical advantages of the response surface technique are discussed in Section 1.8, together with restrictions on the use of the method.

1.3 Background to noise optimization

Over the past few decades, much pressure has been brought to bear on manufacturers of motor vehicles to reduce the levels of noise generated by their products. This pressure has come from a number of sources, many of which are closely related; a general trend towards a higher level of in-vehicle refinement has been coupled, in recent years, with a growing awareness of environmental issues. Indeed, it is largely this latter concern which has led to the introduction of increasingly strict legislation, both nationally and internationally, regarding the noise levels of all classes of vehicle. Environmental concerns, however, have also been instrumental in the rapid growth in the use of diesel engines in the passenger car sector, which has further increased the pressure on manufacturers to produce low-noise engine designs, since the combustion characteristics of a compression ignition engine make it intrinsically noisier than a comparative spark ignition engine.

Within industry, early attempts at engine noise reduction were largely confined to experimental studies of particular engines, with modifications often being made on the basis of past experience. A wide range of experimental approaches are reviewed by Priede (1980), many of which contributed greatly to the theoretical understanding of the mechanisms of noise generation, transmission and radiation. As this theoretical understanding developed, it was natural for workers to try to predict the noise of engines during the design phase, so that low-noise characteristics could be built into a design at an early stage. Early attempts at dynamic simulation were carried out using discrete element models of the engine system, which employed lumped mass approximations to different parts of the structure, connected by spring and damper elements. Examples of this approach are given by DeJong (1976) and DeJong and Manning (1979), who investigated the effect of main bearing stiffness and structural modifications on dynamic behaviour.

The frequency range which could accurately be modelled by such discrete element techniques was necessarily limited by the number of parameters which could feasibly be included in the model, and this in turn was constrained by the computing power available at that time. As theoretical techniques and computing facilities improved, however, discrete element modelling was gradually superseded by the finite element technique, with a number of commercial software packages becoming available to aid in the analysis. Using large mainframe computers, reasonably sophisticated finite element models could be used to calculate natural vibration characteristics of the main engine components, and comparison with experimental modal analysis data enabled good correlation to be achieved within a far wider frequency range than had previously been possible. Turner, Milsted and Hanks (1984), for example, described an analysis of a four-cylinder in-line engine structure, in which good correlation of mode shapes and frequencies was obtained up to 2000 Hz, using a half-engine model containing fewer than 1000 degrees of freedom.

With the increasing use of the finite element method came the first attempts at using the technique to determine beneficial ways in which to modify the structure, in order to enhance noise and vibration characteristics. Lalor (1979), Somkhanay (1982) and Lalor and Petyt (1982) used the static deflection of the cylinder block as a criterion by which to judge dynamic response. Numerical optimization techniques were used to vary parts of the block structure in order to minimize the static deflection of the crankcase skirt or main bearings. The static loading applied was that associated with peak pressure in a single cylinder, applied at the cylinder head and at the adjacent main bearings. Although possessing a number of advantages, such as the relatively low computing cost associated with static analysis, this method suffered certain disadvantages, which prevented it from gaining general acceptance as a practical optimization tool. Chief among these were the uncertainty as to whether the static deflection criterion was an acceptable measure of dynamic behaviour, and the inability of the static loading to represent the complex interaction of excitations imposed by the various cylinders of a multi-cylinder engine.

In order to predict the radiated noise of the engine structure directly, however, further developments were required to provide a more adequate representation of the major excitation mechanisms occurring within the structure, and to incorporate the calculation of radiated noise from surface vibration velocity. Martin and Law (1989) calculated the excitation at the cylinders and main bearings, due to cylinder pressure loads, using an interaction analysis which accounted for the flexibility of the crankshaft, the cylinder block and the separating oil films. Wilcox (1988) developed a dynamic optimization procedure which incorporated shape optimization of the cylinder block, although this approach again suffered from the use of a single-force excitation model. The importance of including correctly phased multiple-point excitation forces was demonstrated by Ogendo and Zhang (1989), who showed that a substantially different optimum design is reached if only single force excitations are imposed.

Recent work by Milsted, Zhang and Hall (1992, 1993) and Zhang (1992) has established a comprehensive noise analysis capability which includes determination of the excitation model and which performs explicit calculation of one-third octave band A-weighted radiated sound power. Initial attempts at performing optimization of concept-level finite element models with this noise prediction program were made using direct iterative procedures, but these proved largely unsuccessful for all but the simplest one-dimensional problems. Although compounded by the relatively large computing requirements of each analyser call, the main reasons for the poor performance of these numerical trials was the mismatch between the characteristics of the direct iterative approach and the requirements of the noise optimization problem which are outlined in Section 1.1. It is these shortcomings of the direct iterative approach to optimization which the present work aims to address.

1.4 The direct iterative approach to optimization

The majority of currently available FE-based optimization programs use iterative numerical optimization algorithms which obtain function information by calling an analyser routine directly. These routines seek to identify the combination of input variable levels which returns the best value of a particular function (the *objective function*), whilst keeping within the allowed range of the input variables (the variable *bounds*), and observing user-defined limits on a number of other functions of these variables (*constraint functions*). In order to locate the optimum design, a wide variety of alternative search strategies are employed. A number of the more efficient algorithms use approximation techniques, or use sensitivity analysis, embedded within the FE code, to supply local gradient information to the optimizer.

Although these methods have the potential for rapid identification of a single optimum, they make inefficient use of the computationally expensive finite element analyses which have been requested, in that many function evaluations are often performed in an optimization trial, of which the analyst will generally only be presented with the final iteration. An additional characteristic of this rapid search for a single point in the multi-dimensional variable space is that it gives little indication as to the relative effect of each variable on the functions of interest. Of much greater importance, however, is the fact that the variable combination which is located is optimum only for the particular set of bounds and constraints which are specified by the user prior to commencing the search.

If the nature of the problem being investigated is such that only one combination of constraints is possible, or feasible, then this characteristic of the direct iterative approach is of no consequence. When carrying out a practical optimization study, however, especially in the early 'concept' stages of an engine design, it is unlikely that the constraint levels which are to be imposed will be known with any great precision. In the case of noise optimization subject to a mass constraint, for example, the exact mass of an initial engine block specification is often not a precisely defined target. Thus, rather than being a rigid constraint which must be strictly adhered to, the mass of the structure can more appropriately be thought of as an additional variable in the optimization process. The goal of an optimization study, therefore, should not be restricted to the identification of one optimum, subject to a single mass constraint, but should be to provide some indication as to how the optimum noise level will vary over a range of possible constraints, so that the designer has the necessary information to enable him to select an appropriate trade-off between these conflicting requirements.

If one were to seek to provide this information using the direct iterative approach, however, it would be necessary to carry out a separate optimization trial for each constraint value, or for each combination of values if multiple constraint functions were being considered. Even if the number of function evaluations required for each optimization trial were very low, the number of different constraint criteria which could practically be investigated would still be extremely limited.

An additional problem which is encountered when using iterative search techniques is that it is not generally possible to identify whether an optimum which has been located is the globally optimum design within the region of interest, or is merely a local minimum. This is an important issue in the present application, since Ogendo and Zhang (1989) have shown that there is substantial potential for the occurrence of local minima in the noise function, due to the fact that resonance vibration of the engine structure is a significant component of the radiated noise. Although the nature of an optimum design can be further investigated by performing a number of optimization trials with different starting locations, each of these additional runs has a computational requirement similar to the initial trial, rendering this solution extremely costly.

1.4.1 Identification of precise optima

Of the three characteristics of the direct iterative approach to optimization which make it unsuitable for noise optimization purposes, as outlined in Section 1.1, the third concerns the relationship between the precision with which an optimum design is sought and the overall accuracy of the noise prediction process. This process necessarily involves a number of approximations and assumptions, derived from a variety of sources, which together result in a substantial margin of uncertainty in the function values which are computed.

The first of these sources of approximation concerns the accuracy with which the finite element model of an engine component, for example a cylinder block, represents the nominal definition of that part as defined by the relevant engineering drawing. The discrete nature of the finite element technique is such that even a highly detailed model can only approximate to the continuous nature of the component being modelled, with the accuracy of the approximation increasing as the fineness of the mesh, and hence its ability to represent localised geometric detail, is increased. (This of course assumes the ability of the analyst to construct a faithful representation of the engineering drawing, which, even with the latest generation of graphical preprocessing tools, is far from being a trivial task). When carrying out an optimization study, however, in which a substantial number of analyses must be performed, the model which is employed must necessarily be coarser than that which would be used to carry out a single dynamic study, in order to ensure computational viability. In such situations, significant reduction in the accuracy of the model is likely, although this can be mitigated to some extent by carrying out correlation with experimental test data, should hardware be available.

The second contributor to prediction uncertainty is the solution method used to analyse the finite element model. Whilst the calculation of natural vibration characteristics and frequency response functions is a well-established and reasonably accurate mathematical procedure, several alternative methods of calculating radiated sound power from surface vibration velocities may be used, offering different levels of compromise between numerical accuracy and speed of solution. In increasing order of complexity/accuracy/computational requirement, the three most

commonly used procedures are as follows:

i) representation of the surfaces of the structure as simple flat plate radiators
ii) the Rayliegh approximation method, idealising the structure as a set of simple sources
iii) the Helmholtz boundary integral method

Computational limitations often require the selection of either i) or ii), with the resulting numerical imprecision which this choice entails, and it is the first of the methods which is implemented within the noise analysis program used in the present study, as discussed by Milsted, Zhang and Hall (1993).

Perhaps of more fundamental importance is the source of inaccuracy associated with calculation of the excitation functions which are to be applied to the structure. Forces applied by the cranktrain are often calculated using an interaction analysis which takes account of the flexibility of the crankshaft, the cylinder block and the separating oil films. Whilst cranktrain forces are likely to be the dominant source of excitation for the majority of engines, however, a number of other significant mechanisms are often unaccounted for in current noise analysis practice, (such as piston slap, valve train impact forces, and gear and chain meshing excitations), with a subsequent reduction in overall accuracy. An issue related to the scope of the excitation is the scope of the finite element model itself, since a cylinder block, or block/head/sump assembly, never vibrates in isolation, but has a number of other components attached to it, whose mass and stiffness characteristics will necessarily modify the dynamic behaviour of the bare block. Although such issues have important consequences for the practical application of noise analysis and optimization, their investigation goes beyond the scope of the current work.

Even if the noise function returned by the analyser program could be relied upon as an accurate prediction of the performance of the nominal engine design, however, there remains a further source of variability within the noise analysis procedure, due to the effect of manufacturing tolerances. None of the engine structures which are manufactured will be an exact representation of the nominal design, and thus the finite element model is better considered as a typical member of a class of engines (which may or may not have the nominal design as its mean) than as a representation of all of these structures. This is an important characteristic, since the corollary is that only well defined trends in the variation of function values with variable modification can be relied upon as being representative of the class as a whole, with small-scale effects being particular to individual members of the class. (There are also implications for the correlation of finite element models when comparison is made with experimental data derived from a single member of this class). A further effect of manufacturing constraints, of course, is that there is no benefit in specifying component dimensions to a higher degree of precision than can be achieved by the manufacturing process.

When carrying out an analysis at the concept stage of an engine design there is still one further cause of uncertainty, in that many aspects of the nominal design itself may be undetermined, either due to lack of detailed component drawings or because decisions are awaited as to major design features. This 'fluid' nature of the design clearly has the potential to introduce the most significant amount of uncertainty into the analysis process.

All of these observations lead to the conclusion that the determination of precise optimum designs is inappropriate to the study of radiated sound power. It is potentially misleading, in that it may highlight variations in the noise level of the nominal design, with respect to variable modification, which cannot be expected to be representative of the engine population as a whole, (although this cannot of course be verified without conducting a detailed analysis, prototype manufacture and test program). Simply quoting the final optimum design to a lower degree of precision, or using coarser termination criteria in the search for this optimum, is unlikely to prove a satisfactory remedy, however, in that it does not address the manufacturing characteristic of the distribution of actual components about the nominal design. It also effectively wastes the computational effort which has been invested in generating the precise function values on which the optimization is based. The underlying problem is the (potentially) unrepresentative level of detail which is present in the definition of the objective function which is used for the optimization phase of the investigation.

1.4.2 Overview of the direct iterative optimization procedure

Figure 1.1 shows a flowchart representing the main features of the direct iterative method of optimization. This illustrates the fact that the large computing requirements associated with multiple dynamic analyses of the finite element model are incurred within the optimization loop, and that the number of these analyses which are required to locate a single optimum is generally unknown prior to entering the loop, thus preventing accurate planning of computer resource allocation. Whilst computing allocation presents little problem when iterations take a matter of seconds or minutes, considerable inconvenience may be caused when carrying out a noise optimization exercise, since a single function call may itself take many hours to perform.

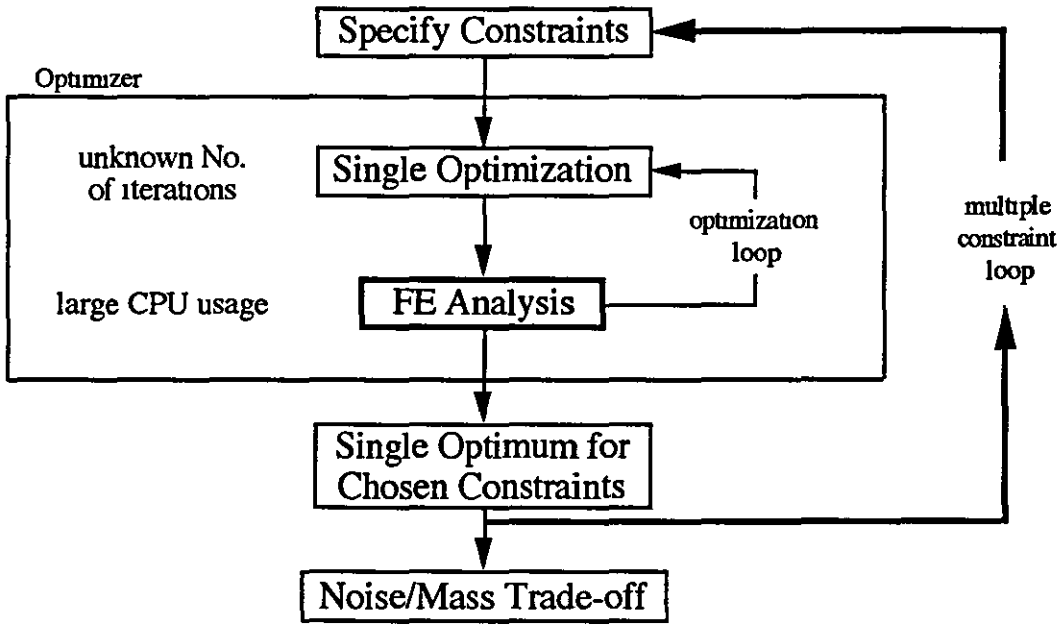

Figure 1.1 Flowchart showing direct iterative optimization procedure

The flowchart also shows that each investigation of a different set of constraint levels requires the execution of the optimization loop, with the large computing costs which this entails. As mentioned previously, this places severe restrictions on the number of different constraint combinations which it is feasible to investigate. It is also worth considering that the inner optimization loop may itself need to be executed several times when locating a single optimum, in order to establish that this minimum is global in nature.

In addition to the fundamental theoretical considerations discussed above, the use of direct iterative optimization techniques has a number of practical disadvantages when applied to noise optimization of engine structures, amongst which the following are important. Firstly, the general sensitivity information required to assess local gradients is not currently available for complex functions such as radiated noise. As a result, noise optimization cannot be carried out directly using algorithms which require this information. Secondly, commercial optimizers which use direct iterative methods are almost inevitably linked with a proprietary finite element analysis package, since, as shown in Figure 1.1, they must be able to modify input variables and call the analyser directly. The analyst is thus limited by the functionality of the particular finite element program used, which can cause particular problems in the case of noise optimization studies, since many codes do not have the capability to calculate complex dynamic responses such as radiated noise. A related constraint is that many of the currently available structural optimization packages are only able to modify a very limited range of variable types, such as material properties, or physical properties of shell elements.

1.5 A survey of currently available software packages

A number of software package are currently available for carrying out structural optimization of finite element models, several of which are incorporated within major finite element codes.

One example of an integrated package is the optimization capabilities contained in Version 66 of MSC/NASTRAN. This program implements the methods of Vanderplaats (1984), and its use is described by Vanderplaats and Blakely (1989). The optimizer uses closed-form response sensitivities and approximate analysis techniques in its search for an optimum design, and the program is claimed to be able to solve "real-world design optimization problems that contain many hundred design variables" (quoted from MSC/NASTRAN sales literature, 1992). The range of response functions and design variables which can be included within an optimization study are extremely limited, however, and may be summarised as follows:

Response functions:	weight or volume
	frequency or buckling load factor
	stress, displacement or force
	composite stress component or failure criterion
Design variables:	plate thickness
	beam properties (cross sectional area, moment of inertia, etc.)
	material properties

These capabilities clearly fall far short of the requirements of the noise optimization problem. When analysing a solid model, for example, the only characteristics of the structure which can be varied are the material properties. Vanderplaats and Blakely (1989) state that "development efforts are already under way to include geometric design variables - the locations of the grid points of the finite element model", together with steady-state dynamic response optimization. Neither of these capabilities has as yet appeared.

A very different implementation of the Vanderplaats methods has been carried out by RASNA, in their MECHANICA suite of programs. Finite element meshes are constructed from adaptive p-elements, and geometry is defined parametrically, allowing more convenient specification and modification of design variables. The range of design variables which can be addressed is far greater than in NASTRAN, and includes a shape optimization capability, allowing movement of mesh nodes. The range of response functions is still limited, however, resulting from the need for sensitivity information. Functions include mass, stress, frequency, displacement and temperature.

Geometric shape optimization is also offered by the SDRC I-DEAS Optimization module, which again is integrated with a comprehensive suite of finite element analysis

programs. Implementation of shape optimization would appear to be less convenient than the RASNA approach, however, because of the possibility of causing severe distortion in elements when using a traditional h-method formulation. Response function and design variable capabilities are otherwise similar to those of NASTRAN and MECHANICA.

In summary, the three popular commercial optimization packages described above do not appear to have the required functionality to carry out successful noise optimization of engine structures, particularly in terms of the limited range of responses which can be included as objective or constraint functions. These limitations are in addition to the characteristics of the direct iterative optimization approach which are discussed in Section 1.4.

1.6 An example of the use of iterative optimization techniques

As discussed above, optimization of radiated noise is not currently possible using the class of advanced optimization algorithms which require sensitivity information in order to assess local gradients. Direct iterative optimization can still be carried out however, by using optimization algorithms which use finite difference approximations to the function gradients. These general-purpose algorithms have a number of disadvantages when used to carry out noise optimization studies, however, amongst which the following are of importance:

1. Considerable computational effort is required to assess the local gradients at the starting point in order that a search direction can be identified.
2. The search algorithms used tend to be very conservative, only making small changes in the input variables at each call.
3. As a result, a large number of calls are often required to locate a minimum.
4. The termination criteria of these algorithms also tend to be conservative, in that many additional function evaluations are often performed, in order to verify that an optimum has been reached.
5. As with other iterative approaches, it is not possible to identify whether the optimum which has been located is local or global in nature without carrying out further optimization trials from different starting points.

As an example of the use of these iterative optimization routines, the following study was undertaken by Ogendo and Zhang (1989), and also discussed by Zhang (1992). This example uses the finite element model of the diesel engine cylinder block which is described in Appendix 1C, and investigates the effect on radiated noise of the three design variables which are listed in Table 1.1 and identified in Figure 1C.1.

	Variable	Thickness in mm lower	upper	mean	range
A	Crankcase skirt	6.0	14.0	10.0	8.0
B	Bearing panels 1 and 5	20.0	32.0	26.0	12.0
C	Bearing panels 2 and 4	20.0	32.0	26.0	12.0

Table 1.1 Variable values for the three variable iterative optimization study

The search history of the optimization study is shown in Figure 1.2, and demonstrates the following characteristics.

1. The central difference calculation used to assess local gradient information requires an initial (2n +1) = 7 function calls to identify a search direction before optimization commences.
2. When the search commences, only three substantial steps are taken before progress slows dramatically. This is followed by 9 steps in which the variable changes are extremely small before the next significant change is made (in variable C).
3. The design specification at call 20 is effectively the same as that at call 45. An extra 25 function evaluations have been carried out in order to verify that an optimum has been reached.

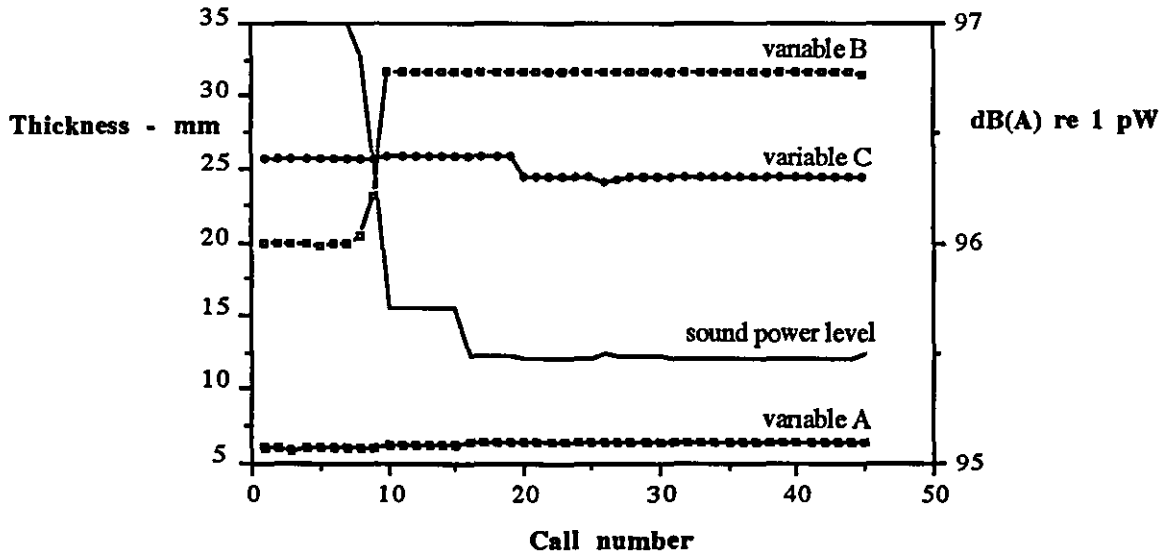

Figure 1.2 Search history - unconstrained optimization in three variables using direct iterative optimization

The iterative optimization algorithm has thus taken 45 function calls in order to identify an unconstrained optimum in three design variables. In comparison, Section 9.2 presents an example in which the response surface method is used to carry out an unconstrained optimization in seven variables with a similar number of function calls (43 calls). In addition, the response surface approach will also allow the identification of a constrained noise optimum subject to any feasible mass constraint, and the identification of an optimum mass design subject to any feasible noise constraint, without the need for further analyser calls. Direct iterative optimization studies involving greater than three design variables were not found to be viable with the available search algorithm, which generally failed to converge to an optimum value, even after an unacceptably large number of function calls.

Whilst it may be possible to improve the performance of the iterative optimization algorithm used above by, for example, using forward difference calculations instead of central differences, and modifying the tolerance of the termination criterion, this study still illustrates the three fundamental requirements for which the use of direct iterative optimization is inappropriate. Firstly, the algorithm aims to identify the optimum design with a higher level of precision than is appropriate to the formulation of the problem. Secondly, the method is not robust with respect to the occurrence of local minima, since a single trial cannot identify whether the optimum found is local or global. Multiple optimization trials are then required to verify the nature of the minimum, and to locate a global optimum, necessarily incurring a substantial computational cost. More importantly, however, a separate optimization trial must be carried out for each combination of imposed constraint levels which are of interest, with a consequent increase in computing requirements.

1.7 The response surface approach to optimization

An alternative procedure to the direct iterative technique described above is to build up a 'database' of knowledge concerning the system under investigation by calculating the values of the required functions at a number of selected points within the design variable space. In order to reduce the overall computational requirements, the sample points should be chosen such that a sufficiently accurate response surface can be generated with the minimum number of function evaluations. This objective may be achieved by using experimental design theory to determine the most efficient combination of sample points at which to test. A mathematical model can then be generated from the database, so that an approximating *response surface* is constructed using, for example, low order polynomials. Optimization may then be carried out using this surface as a representation of the original response, with the difference that the evaluation of each function and its gradients takes negligible computing time.

A response surface-based optimization study may be divided into four phases, as follows:

Phase 1. Experimental design selection

In this phase, experimental design theory is used to select the points in the design region at which to test. The only parameters which need to be supplied in order to generate the design are the number of variables to be investigated and the bounds within which each of these variables may be modified. The type of design which is selected is determined by the complexity of the mathematical model which is to be used to approximate the response function. If the response is to be approximated using a linear response surface, for example, then a design may be chosen which includes tests at combinations of just two values of each variable. If a quadratic surface is to be fitted, however, then each variable must appear at at least three different levels within the experimental design.

Phase 2. Analysis

In the analysis phase, the functions which are of interest are evaluated at each of the combinations of variable values which are specified in the experimental design. The analysis program does not need to be directly linked with the optimization program, and the order in which the tests are carried out is unimportant. Thus, several tests may be carried out in parallel, using multiple processors or multiple computers.

Phase 3. Model building

This phase may be subdivided into two sections, as follows:

Phase 3a: response surface fitting.

Surface fitting algorithms are used to estimate the coefficients of the mathematical model using the function values returned by the analyser program.

Phase 3b: validation of response surfaces.

For each response function, detailed lack-of-fit calculations are performed in order to assess how well the response surface is representing the variation of the original function throughout the whole of the design variable space. This stage is of vital importance to the

successful use of the response surface approach, since if the approximating mathematical model is providing an inadequate representation of the calculated function, the optimization process is likely to identify optimum designs which are substantially in error.

Phase 4. Optimization

In the optimization phase, the engineer can specify which of the calculated functions is to be the objective of the optimization, and which are to be constraints. In addition to minimizing radiated noise subject to a mass constraint, for example, the engineer can also optimize the design for low mass, subject to a noise constraint. It is also possible to carry out a *sweep* through the entire range of constraints, giving an explicit trade-off between different functional constraints without the need for further analyser calls.

In the current work, attention has largely been focused on determining the complexity of model that is required to represent the noise function (Phase 3b), and selecting an appropriate experimental design with which the model can be estimated using the minimum number of function calls (Phase 1). Accurate representation of the computed responses is fundamental to the successful use of the overall procedure, and Chapters 3 – 8 are devoted to this subject. The analysis phase of the procedure has been performed using the separate analyser program which is described in Appendix 1A, and estimation of model coefficients is carried out using standard surface fitting algorithms, as described in Section 2.13. The selection of an optimization algorithm to locate optimum designs on the approximating response surface is discussed in Chapter 9, where a range of numerical studies are presented which demonstrate the capabilities of the method.

1.7.1 Overview of the response surface-based optimization procedure

Figure 1.3 shows a flowchart depicting phases 2 and 4 of a response surface-based optimization study.

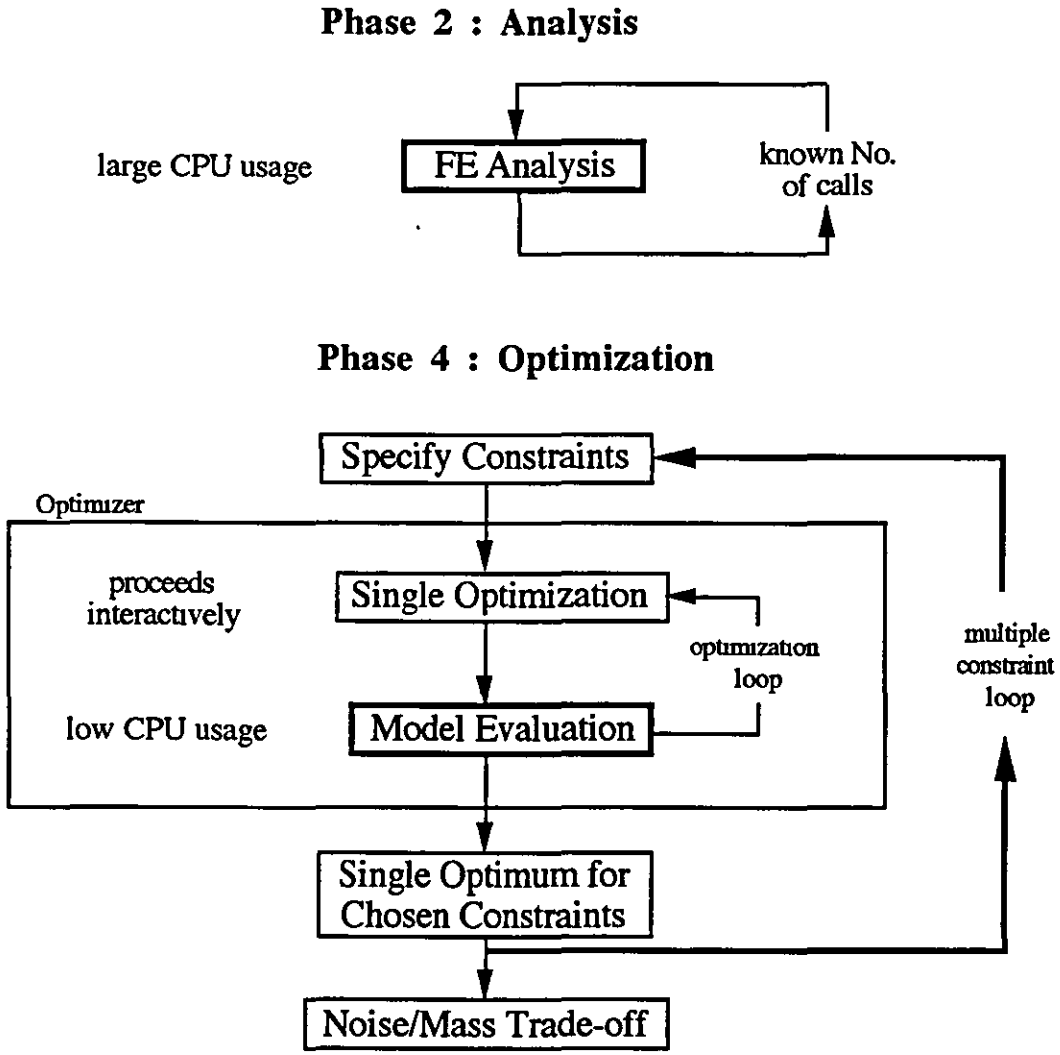

Figure 1.3 Flowchart showing phases 2 and 4 of the response surface-based optimization procedure

The flowchart for phase 4 of this procedure is identical in form to the flowchart of Figure 1.1 for the direct iterative approach. The difference lies in the fact that, instead of calling the finite element analysis program in order to calculate function values, this information is obtained by simply evaluating the low-order mathematical model of the original response. The computationally intensive finite element calculation of function values is now carried out outside the iterative optimization loop, in phase 2, and full use is made of all of the analyses which are performed. Investigation of multiple constraint criteria still requires the execution of the inner optimization loop, but this now incurs no significant computational cost.

1.8 Characteristics of the response surface method

The major advantages of the response surface approach to optimization are as follows:

- In the testing phase, all functions which may be of interest are calculated and stored in a database of knowledge concerning the system under investigation. These may be recalled during the optimization phase, and any combination of objective and constraint functions investigated interactively.
- The coefficients of the response surface models indicate the relative importance of each of the design variables throughout the whole design region, giving the designer a valuable insight into the nature of the system.
- Any continuous function of the design variables which can be calculated or estimated at the required sample points may be added to the database prior to optimization. This would, for example, allow non-FE functions such as cost to be included within an optimization study. Alternatively, the results of both static and dynamic analyses can be combined in a single optimization study. Similarly, a number of different measures for assessing the radiated noise level of running engines can be used, including various weighting curves and measures of noise quality. It is becoming increasingly important for an engine to perform well under a range of these different criteria, and the response surface approach will allow the identification of designs based on each of these required measures, or on some appropriate combination of them, without incurring excessive computational cost.
- Because the finite element analyses are carried out outside the iterative optimization loop, any available analysis package, or even a combination of different specialist packages, can be used to supply function information.
- If a dedicated analyser program is used to carry out function evaluations, the modification of design variables can be automated within this code, as is possible when using the direct iterative approach. Because analyses are performed outside the optimization loop, however, it is also possible to incorporate major modifications to the engine structure which are beyond the capability of automatic modification routines. These changes can be carried out either manually or semi-automatically.
- The number of finite element analyses which need to be performed are known in advance, allowing accurate prediction of computer usage.
- Because analyses do not need to be run consecutively, parallel running may be carried out on multiple processors or machines, drastically reducing the elapsed time required for the testing phase of the study.

Disadvantages of the response surface approach:

It is probable that an advanced software program for direct iterative optimization will be able to find a constrained optimum with fewer function evaluations than would be required to establish an approximating response surface of sufficient accuracy. Thus, in cases where the value of each functional constraint can be precisely determined prior to optimization, so that the only requirement is the efficient identification of a single optimum, the direct iterative approach is likely to be the best method to follow.

Restrictions on the use of the response surface approach:

An essential requirement of this method is that it must be possible to model the functions which are of interest using response surface approximations. Any function of the design variables can be used as either an objective or constraint function, as long as it fulfils the following three criteria :

1. Computable at the required test points.
2. Continuous function of the design variables.
3. Capable of being approximated by a low order polynomial.

One of the principal aims of the work is to establish that the noise response function fulfils these criteria. Additional functions of the input variables which may be of interest when carrying out dynamic analyses of engine and powertrain models are discussed in Section 9.11.1.

Appendix 1A

A finite element analysis procedure for calculation of radiated engine noise

1A.1 General approach

The finite element analysis program which has been developed in order to compute the sound power radiated from an engine structure has as its basis a much-modified version of the commercial finite element analysis code PAFEC. A general description of the procedure is given in Milsted, Zhang and Hall (1993), and a more detailed discussion of the underlying theoretical and physical considerations can be found in Zhang (1992) An introductory user guide to the program suite is provided by Hall and Zhang (1992).

Early work concerned with the tailoring of this code in order to perform a radiated noise calculation was carried out by Turner (1983), who wrote a set of forced response subroutines to accompany the PAFEC free vibration analysis routines. Also implemented at this stage was a dynamic substructuring capability, using a modified version of Kron's method (Turner, Milsted, and Hanks, 1986). Following this work, a detailed investigation into the physical aspects of the noise radiation problem was undertaken by Ogendo and Zhang (1983), which established A-weighted sound power as an appropriate objective function for optimization purposes This work also highlighted the need for a complex set of excitation forces which can accurately simulate the dynamic loading applied to the structure. Using the separate modules for generation of excitation forces and calculation of sound power from surface vibration levels which are described below, a theoretically sound method for predicting radiated engine noise had now been established. As part of a subsequent investigation, (Milsted Zhang and Hall, 1989), substantial effort was devoted to increasing the computational efficiency of the whole noise analysis procedure, in order to facilitate its use within a numerical optimization capability. This work centred on the replacement of the standard PAFEC eigensolution, based on a master-slave reduction method, with a Lanczos eigensolution algorithm (Sehmi, 1989) A sparse matrix approach to storage and computation was also implemented, together with a modified residual flexibility algorithm and a number of changes to the matrix assembly and general administration routines.

In its present form, one complete noise analysis of a single-structure model consists of four main stages, as shown in the simplified flow chart of Figure 1A.1. Each of these stages is discussed in the following sections

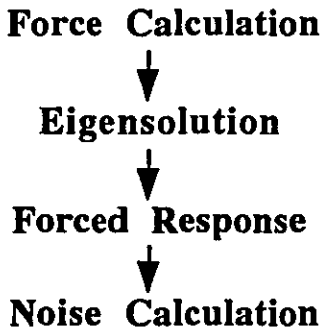

Figure 1A.1. Schematic diagram of noise analysis procedure

1A.2 Calculation of mechanical loads

In the first of these phases, the excitation forces which are to be applied to the structure are computed, using a separate program module. A more detailed account of the interaction analysis which is used to calculate these excitation forces may be found in either Martin and Law (1989) or Ogendo and Zhang (1989), from which the following summary is taken

"For the purpose of vibration and noise prediction, the mechanical loads arising from the combustion process can be usefully approximated by a set of discrete forces and moments acting on the cylinder block and head A suitable set is comprised of forces and moments at each main bearing, piston side thrust forces on each cylinder and the gas force on the cylinder head. The calculation of these forces can be approached at various levels of sophistication ranging from the elementary rigid crank train and rigid block procedure through to a full interaction analysis between the fixed and the moving parts of the engine using FE models of the block and crankshaft. The latter approach has been taken because it is the only means by which the significant bearing-to-bearing force differences can be predicted. . . [The procedure used] is a quasi-static interaction analysis of a system comprising the crank train, the block and various oil films which separate the two. Experimentally measured, frequency-averaged cylinder pressure diagrams are used as the excitation. This analysis involves the sequential solution of the pertinent structural and hydrodynamic equations which describe the important features of the engine. The structural equations are formulated in terms of influence coefficients derived from representative substructured finite element models of the crankshaft and engine structures. The hydrodynamic operation of the main bearings is modelled by a 'mobility' method which provides explicit relationships for the translational velocities of the crankshaft journals relative to the bearing sleeves. The load so obtained is a function of the crankshaft angle and must therefore be transformed to the frequency domain for subsequent use in the vibration analysis. In this procedure the inertial loadings, the flexibility of the block and the crankshaft, as well as the nonlinear oil film effects in the journal bearings are properly accounted for, but piston slap, gear and valve train impacts are not yet included. Although the omission of these impact forces does not affect the establishment of the numerical optimization procedure, it is clear that their presence would increase the accuracy of the noise prediction."

1A.3 Solution of the eigenproblem

The second portion of the noise analysis procedure is the solution of the undamped eigenproblem

$$(\mathbf{K} - \lambda \mathbf{M})\, \mathbf{x} = 0 \qquad (1A.1)$$

where **M** and **K** are the mass and stiffness matrices respectively. This is the most computationally intensive part of the procedure, and that to which the most development effort has been devoted.

The engine structures for which a noise calculation is to be carried out will often be unrestrained. In order to ensure the positive definiteness of the stiffness matrix under such conditions, it is necessary, both here and in the forced response calculation, to use a shifted stiffness matrix of the form

$$\tilde{\mathbf{K}} = \mathbf{K} + \delta \mathbf{M} \qquad (1A.2)$$

The corresponding shifted spectral matrix is then

$$\tilde{\Lambda} = \Lambda + \delta \mathbf{I} \qquad (1A.3)$$

The shift parameter δ takes a positive value small enough in magnitude with respect to the eigenvalue of the lowest non-explicit mode that the residual flexibility approximation outlined in Section 1A 4 remains valid For systems in which no rigid-body modes are present, a value of $\delta = 0$ is used.

As discussed above, a Lanczos eigensolver with partial re-orthogonalisation has been implemented, (Sehmi, 1989), which also takes advantage of the sparse banded nature of the stiffness and mass matrices. The solution algorithm for a restrained structure may be summarised as follows, (Milsted, Zhang and Hall, 1993) The general eigenproblem is first transformed to the standard form

$$(\mathbf{H} - \theta\mathbf{I})\mathbf{y} = 0 \tag{1A.4}$$

The stiffness matrix is factorised as

$$\mathbf{K} = \mathbf{L}\mathbf{L}^T \tag{1A.5}$$

using Choleski decomposition, so that

$$\mathbf{H} = \mathbf{L}^{-1}\mathbf{M}\mathbf{L}^{-T},\ \mathbf{y} = \mathbf{L}^T\mathbf{x} \text{ and } \theta = 1/\lambda \tag{1A.6}$$

The Lanczos method is then employed to solve the standard eigenproblem for the first k modes, giving a modal matrix $\mathbf{Y}_k$ and a spectral matrix Θ_k. The solution of the original eigenproblem is then obtained from

$$\mathbf{L}^T\mathbf{X}_k = \mathbf{Y}_k \text{ and } \Lambda_k = \Theta_k^{-1} \tag{1A.7}$$

1A.4 Forced response

Using the free vibration information generated within the second phase of the program, a modal forced response calculation is then carried out to give the steady-state vibration response of the structure. This may be summarised as follows (Milsted, Zhang and Hall, 1993).

The complete set of eigenvalues and mass-normalised eigenvectors are first assembled into a diagonal spectral matrix Λ and a modal matrix $\mathbf{X}$, respectively. The forced response at a frequency ω can then be written in the form of a receptance matrix as

$$\alpha = \mathbf{X}\,[\,\Lambda(\,\mathbf{I} + \mathrm{j}\,\eta\,) - \omega^2\,\mathbf{I}\,]^{-1}\,\mathbf{X}^T \tag{1A.8}$$

where η is a diagonal matrix of the modal loss factors. The mode set is truncated at a frequency which is large enough to ensure accurate representation of all resonant behaviour within the frequency range of interest. A residual flexibility approximation to the discarded modes is used in order to take some account of the contribution of those modes which have not been explicitly included in the calculation, so that equation (1A 8) becomes,

$$\alpha = \mathbf{X}_k\,\{[\Lambda_k(\,\mathbf{I} + \mathrm{j}\eta_k) - \omega^2\mathbf{I}]^{-1} - \tilde{\Lambda}_k^{-1}\}\,\mathbf{X}_k^T + \tilde{\mathbf{K}}^{-1} \tag{1A.9}$$

where the subscript k denotes matrices of order equal to the number of modes explicitly included (kept) in the analysis.

1A.5 Radiated sound power calculation

The final phase of the noise analysis procedure is the calculation of radiated noise from the surface vibration levels, which is carried out by a separate program module. The following summary of the procedure used is taken from Milsted, Zhang and Hall (1993).

"Calculation of radiated sound power from the surface vibration levels can be carried out with varying degrees of sophistication In increasing order of complexity the three well established procedures are.

(i) Idealisation of the engine structure as a set of simple flat plate radiators
(ii) Rayleigh's method of idealisation of the structure into simple sources
(iii) The Helmholtz boundary integral method

"The choice is not particularly critical to the program structure as any of the three methods can be viewed simply as a post-processor operating on the surface vibration levels whose calculation is the core of the procedure Using (i), the A–weighted sound power, summed from one-third octave band contributions, is evaluated from

$$W = \rho c \sum_{i} w_i \left(\sum_{j} \sigma_{ij} A_j \langle v_{ij}^2 \rangle \right) \qquad (1A.10)$$

where ρc is the characteristic impedance of air, A_j is the effective sound radiating area, σ_{ij} is its associated radiation efficiency; and $\langle v_{ij}^2 \rangle$ is the space averaged mean-square velocity normal to the sound radiating surface. The subscript i identifies the one-third octave band with A–weighting w_i and the subscript j denotes a panel region of the noise radiating surface "

1A.6 Variable types

The following classes of variables can be modified by the current version of the noise analysis program:

Physical properties of beam elements
Physical properties of shell elements
Material properties of 2D and 3D elements
Nodal movement which does not require mesh alteration

Appendix 1B

The computer program *optrsm*

As an integral part of the current investigation into the use of response surface techniques for the noise optimization of engine structures, a computer program has been developed in order to aid the application of these methods. This response surface methodology and optimization code forms part of a larger suite of programs which together allow a complete noise optimization study of an engine structure to be carried out using finite element methods. The other main component of this suite of programs is the noise analysis program, described in Appendix 1A, which is used to simulate the steady-state vibration of the engine and hence predict radiated noise. Further details of this program, together with a description of the way in which the two programs are used together to carry out noise optimization, are given by Hall and Zhang (1992).

The response surface methodology / optimization program *optrsm* is composed of two independent modules, the first of which is used to select an appropriate experimental design prior to testing, whilst the second is used to process the results which are returned by the analyser program, and to perform the required modelling and optimization tasks. Each of the modules is highly interactive, with only a minimal requirement for previously constructed input data files The results presented within subsequent chapters of the present work have been generated using this program, with the noise analysis function evaluations being provided by the separate analyser program mentioned above

The flow diagrams shown in Figures 1B.1 and 1B 2 depict the suggested procedure for carrying out a complete noise optimisation study using both the experimental design selection / optimization program *optrsm* and the separate noise analysis program This procedure may be summarised as follows.

Phase I

The first phase of an analysis involves the construction of the base FE model, the specification and selection of the design variables and the necessary checks of the data. For the noise analysis problem this includes the assembly of the two data files which must be supplied to the main analyser program, with the tangible results of this initial phase being the finite element analysis data file and the analyser data file.

Phase II

This phase constitutes the optimization process proper. The preprocessing mode of *optrsm* is first used to generate the test specifications for which function values are required (STEP 1) For each required test, the specification consists simply of the value of each of the design variables This list of test specifications is then inserted into the analyser data file (STEP 2), and the separate noise analysis program executed (STEP 3). The result of each run is extracted from the output file (STEP 4), and the postprocessing mode of *optrsm* is then used to analyse the results and perform the optimization (STEP 5). The execution of steps 2, 3 and 4 is described by Hall and Zhang (1992). The use of *optrsm* at step 5 to process the results supplied by the analyser program is detailed by Hall (1992), together with a description of the results which are generated during step 5, and guidance as to their interpretation.

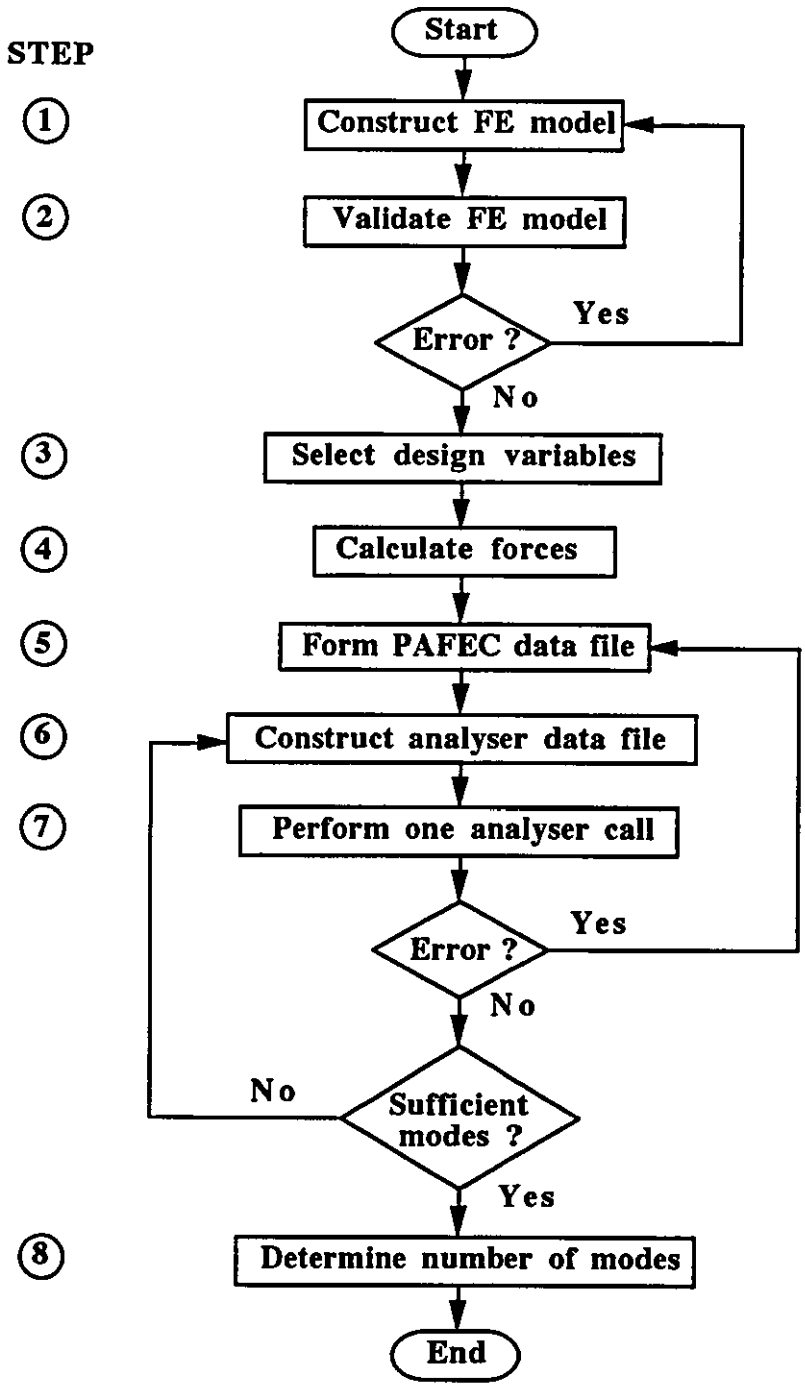

Figure 1B.1. Phase I : Model Building and Data Validation

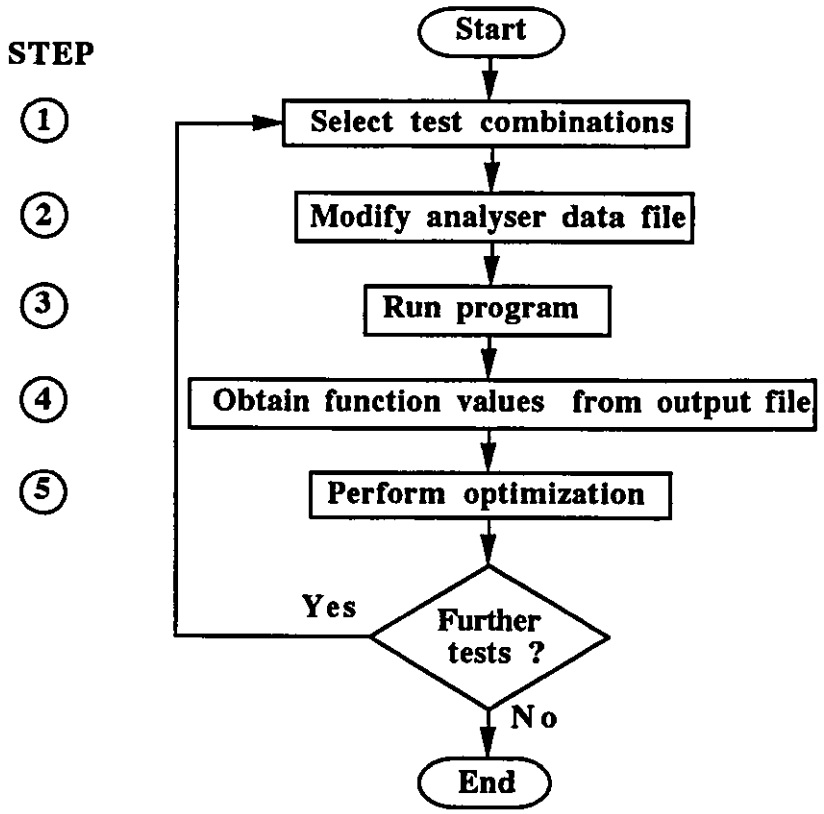

Figure 1B.2. Phase II : Analysis and Optimization

Appendix 1C

Finite element model of a representative engine structure

In order to evaluate the various techniques which are discussed in subsequent chapters, a series of numerical studies have been carried out, using the optimization program which has been developed as an integral part of the present work (see Appendix 1B). These studies have been carried out using a finite element model of the cylinder block of a four-stroke direct injection diesel engine, which has a four-cylinder, in-line configuration, a capacity of 3 86 litres, and a rated speed of 2800 rev/min. The cylinder block is manufactured from cast iron, with dry liners, and a highly ribbed water jacket exterior. The finite element model of the cylinder block is shown in Figure 1C.1. The structure of the engine block is a modified version of a standard production unit, with a number of design alterations having been made in order to reduce the level of radiated noise In particular, it can be seen that three methods of main bearing support are employed; a pair of longitudinal stiffeners connecting the sides of each main bearing bulkhead, a bearing beam tying together the four bearing caps, and lateral supports connecting each bulkhead to the skirts. Only the first of these is found in the standard version of the engine block (Turner, 1983).

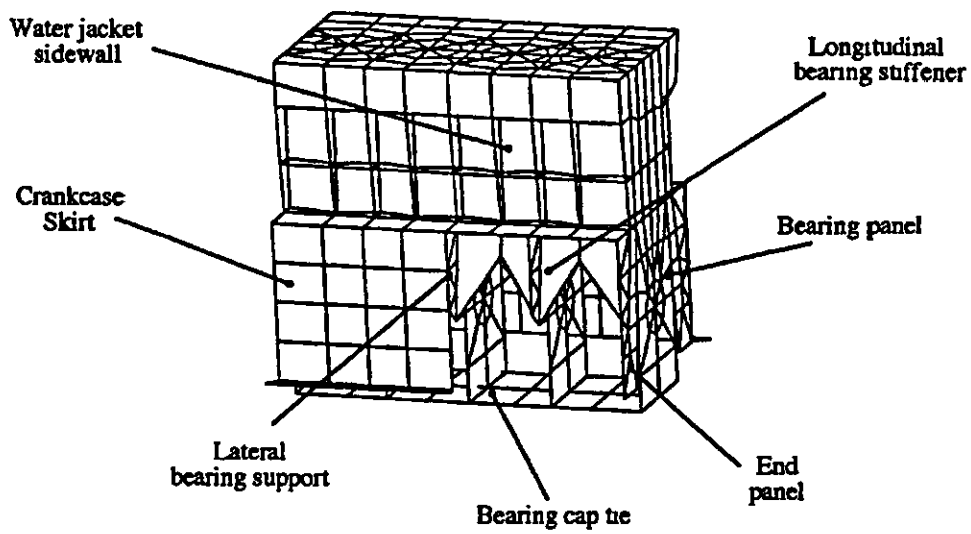

Figure 1C.1. Finite element model of engine block

The finite element model of the engine block is constructed entirely from shell elements, with 840 elements being used, giving 3409 degrees of freedom. Excitation forces, calculated using a non-linear coupled crank / block analysis, as described in Appendix 1A, are applied at 40 degrees of freedom within the structure; x and y components of force and moment at each of the main bearings, two normal forces on each cylinder sidewall face and one head force at each cylinder. Responses at 198 surface coordinates are used to calculate the surface-averaged velocities from which radiated sound power is computed.

Shell elements were chosen to model the engine structure, in preference to solid elements, for two main reasons. Firstly, modification of the structural features of the engine block model is substantially easier to carry out within the finite element analysis program if shell elements are used. As an example, the particular design variables which are addressed by the studies included in the present work involve modification of the wall thickness of various parts of the engine block casting When using a shell model, this can be accomplished by altering a single physical property definition for each element involved, whereas, if solid elements were to be employed, the position of several nodes on each element would have to be modified. Even if this could be achieved without altering the pattern of nodal connectivity, the amount of computational effort required to implement such a change would be significantly greater than that required for the corresponding shell model.

A second reason for preferring a shell model to a solid model is that the size of the analysis problem is significantly smaller, resulting in a reduced solution time for a single noise calculation This is particularly important in the present work, since, when carrying out an optimization study, a substantial number of function calls are often required. Reductions in the time needed for a single analysis lead to more effective use of the available computing resource, either by allowing a more detailed model of the design variable space to be constructed, or by enabling more variables to be included in the study.

It is recognised that the finite element model described above is extremely coarse in comparison with current industry practice. In a commercial environment, a model used for a dynamic analysis might typically be constructed entirely from solid elements. It would exhibit a much greater level of structural detail, and hence a higher mesh density, and might contain over ten times the number of degrees of freedom as the model used here It should be borne in mind, however, that the reasons for this level of model refinement are often commercial as well as technical, and that such a model would be used either for a single dynamic analysis, or to examine the effect of a small number of manually implemented design changes A full formal optimization study using a model of this size would currently require a level of computing resource not readily available to the majority of analysis groups.

Within a commercial environment, a numerical optimization study of a concept-stage engine might thus be carried out using a suitable compromise between these two levels of modelling detail. Such a model might, for example, be constructed predominantly from shell elements, with an intermediate level of mesh density, in order to gain some of the advantages possessed by such models, as outlined above. Solid elements are to be preferred for main bearing bulkheads and bearing caps, however, since it is widely recognised that the accurate representation of mass and stiffness distribution in this area is essential to the accurate prediction of dynamic behaviour (see, for example, Zhang, 1992) Recent studies, for example Ott, Kaiser and Meyer (1990), have shown that mixed shell and solid models of this type can give good correlation with experimental modal analysis results up to at least 2000 Hz.

The justification for using an all-shell model in the current work, however, is that the purpose of the present investigation is to establish the validity of the theoretical approach outlined in subsequent chapters, rather than to draw particular conclusions concerning the example engine block which is being used as a test case. What is important within this context is that the finite element model used is a representative example of a possible engine design, rather than an accurate reflection of the original structure which it has been constructed to simulate, and that the time required for a single analysis is low enough to allow program development to proceed at a convenient pace. Use of the finite element model described above is further supported by a large body of test work which has been carried out in order to provide a correlation with numerical results. An experimental modal analysis programme was carried out using just the bare cylinder block, and a frequency correlation exercise performed with the analytical natural frequencies (Turner, 1983). The results of this investigation were then used to modify the original finite element model in order to improve correlation. Analysis of this updated model showed that, in the range to 2000 Hz, all but one of the experimental natural frequencies were being simulated with a frequency accuracy of ±10%. Although the accuracy of the model above 2000 Hz has not been verified, due to lack of experimental data, it is considered that inaccuracies in this range are unlikely to substantially alter the conclusions drawn from numerical trials, since the noise spectrum is dominated by contributions from one-third octave bands in the range 400—2000 Hz, as shown in Figure 1C 2. Extensive comparisons have also been made between calculated sound power levels and noise levels measured from a running engine (Coulson and Southall, 1976).

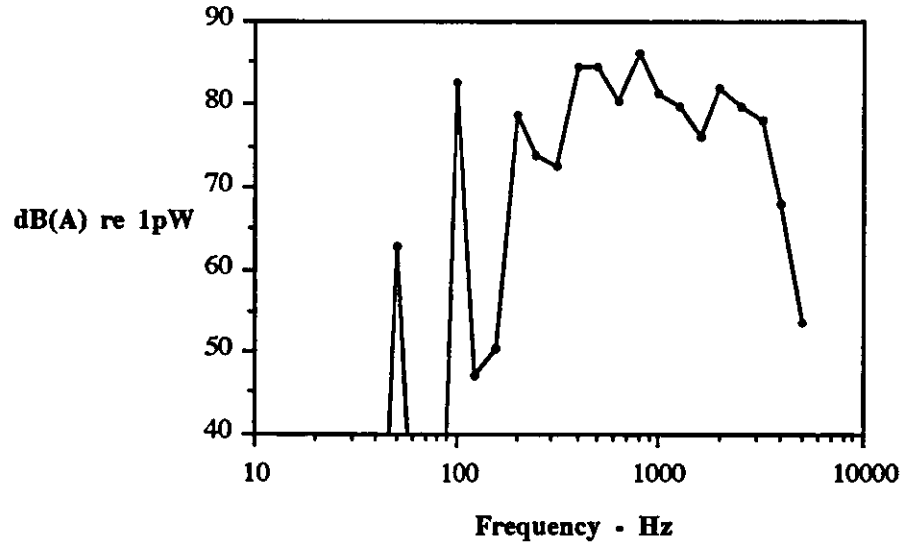

Figure 1C.2. One-third octave band sound power spectrum for initial engine design

Seven design variables have been selected for investigation in the present work, each of which represents the thickness of a part, or parts, of the cylinder block structure. These are identified in Figure 1C.1, and the bounds between which they may vary are given in Table 1C.1.

	Variable	Thickness in mm lower	upper	mean	range
A	Crankcase skirt	6.0	12.0	9.0	6.0
B	Bearing panels 1 and 5, end panels	20.0	32.0	26.0	12.0
C	Bearing panels 2, 3 and 4	20.0	32.0	26.0	12.0
D	Longitudinal stiffener	4.0	14.0	9.0	10.0
E	Lateral bearing support	10.0	25.0	17.5	15.0
F	Bearing cap tie	6.0	12.0	9.0	6.0
G	Water jacket sidewall	6.0	12.0	9.0	6.0

Table 1C.1 Seven design variables of the four-cylinder engine block

2. Response Surface Methodology

The first section of this chapter reviews the development of the techniques of response surface methodology when used for general experimental work. Following this, applications of the technique to the area of 'computer experimentation' are reviewed. The remaining sections of the chapter provide an introduction to a number of the concepts, definitions, conventions and methods of analysis which underlie the work of Chapters 3 - 8.

2.1 A review of RSM

2.1.1 Early work in the use of response curves

The family of techniques known collectively as Response Surface Methodology are generally considered to have had their beginnings in work carried out in the field of agricultural research, particularly with regard to crop growth, in the late 1920's and 1930's. Here the need for planned experimentation was paramount, since a complete growing season was required to carry out a single study, with limited scope for iteratively building a detailed experiment, due to the large seasonal fluctuations in, for example, weather conditions. Wishart (1938,1939) used orthogonal polynomials to approximate the growth rate of pigs, whilst Winsor (1932) used a 'functional' or 'mechanistic' model to investigate a situation in which relative growth was thought to decrease exponentially with time.

Although each of these studies involved the use of one-dimensional growth curves, multi-dimensional response surfaces were also being used in this period to investigate crop yields. An early example is provided by Mitscherlich (1930), and Crowther and Yates (1941) used response surfaces to evaluate the effect of different fertilisers on the yield of arable crops. In such investigations, the need to design an experiment for the specific purpose of fitting response surfaces was seldom addressed explicitly, although an important paper by Yates (1935) used factorial experimental designs to collect response data, and provided the groundwork for much of the later research into the design and selection of experimental schemes. This work also described a simple tabular means of calculating the magnitudes of the various linear and interaction terms of first-order response surface equations (referred to as orthogonal contrasts).

2.1.2 The development of RSM in the 1950's

The popularity of response surface methods increased markedly in the early 1950's, as did interest in the theoretical principles underlying the design of experiments. This popularisation was due, to a large extent, to a number of papers published by G.E.P. Box and his associates at ICI in Manchester. The dominance of this group was such that Hill and Hunter (1966) opened their review paper on RSM with the statement that "Response surface methodology was initially developed and described by Box and Wilson (1951)". Although later workers and reviewer have acknowledged previous work in the field, such as that described in Section 2.1.1, this era was clearly an important turning point in the general acceptance of the technique.

In the first major paper of this period, Box and Wilson (1951) discussed a number of experimental designs, whose aim was to identify an optimum location on the response surface with the minimum number of observations. They assumed that the responses could be modelled using polynomial functions, and stressed the importance of orthogonal designs for estimating model parameters (see Section 2.7). It was also in this paper that Box and Wilson introduced the Central Composite Design (CCD), which has become one of the most widely used second-order designs, and is the subject of Chapters 6 and 7 of the present work.

The topics addressed by Box and Wilson were further discussed and extended by Box (1952, 1954), Davies (1954), and Box and Youle (1955), who also stressed the iterative nature of an experimental investigation. Box and Hunter (1957) judged competing experimental designs on the basis of prediction variance within the design variable space, and from this developed the important concept of rotatability of a design (see Section 2.8). A further important development in the comparison of experimental designs was presented by Box and Draper (1959), who discussed the robustness of a design to model misspecification, and introduced the bias criterion.

In addition to the work of Box *et al.*, a number of other major areas of research into RSM were attracting attention within this period, amongst which was the development of the theory of optimal design. This work was mainly undertaken by Kiefer (1958, 1959, 1960, 1962a, 1962b) and Kiefer and Wolfowitz (1959, 1960), who developed a number of different criteria by which to judge experimental designs. These measures of design optimality are important when generating new types of experimental design, but are of less relevance to a practical use of RSM if standard designs are employed.

In parallel with these developments in the theoretical fundamentals of RSM, a number of papers were aimed at popularising the methods, presenting practical applications of the theory which could more easily be assimilated by industrial statisticians. Among these, the series of papers by Bradley (1958), and Hunter (1958, 1959a, 1959b) are notable.

2.1.3 Designs for linear models

One of the earliest two-level designs to be used for fitting linear models is the full factorial design, in which tests are carried out at all combinations of the high and low bounds of each variable (see, for example, Yates, 1935). This design allows estimation of the main effect terms of a linear model, as well as the interaction effects between all possible combinations of linear factors. It has the disadvantage, however, that the number of tests required grows quickly as the number of variables increases. In order to address this drawback, a class of designs known as fractional factorial designs was introduced by Finney (1945), in which a 2^{-p} fraction of the full factorial design is used, selected so as to yield the maximum information concerning interactions between variables. The importance of these designs was quickly recognised, and examples of their use were discussed by Davies and Hay (1950) and Daniel (1959), amongst others. These designs are still widely used for estimating linear models, and are discussed in detail in Chapter 3.

An alternative method of reducing the number of tests required by the full factorial design was developed by Plackett and Burman (1946). For a Plackett-Burman design in n variables, the number of tests required is equal to n + 1, and is a multiple of 4 rather than a power of 2. One disadvantage of this class of design is that it is not available for all values of n; Plackett and Burman presented designs for n = 3, 7, 11, ..., 99 variables.

In 1957 the National Bureau of Standards, of the U.S. Department of Commerce, published a collection of fractional factorial designs, in which a selection of experimental arrangements requiring 2^{n-p} tests were presented for n = 5, ..., 16 and p = 1, ..., 8. This publication provided a useful reference for those making practical use of response surface methods within industry, and its use is discussed by Zelen and Connor (1959). Alternative methods of formulation and analysis of fractional factorial designs were presented by Box and Hunter (1961a, 1961b). Discussion of many aspects of two level designs, together with a number of variations on the standard methodology, was undertaken in a series of papers by Addleman (1961, 1962a, 1962b, 1963, 1964, 1969).

Recent work concerning first-order designs has concentrated on modifications to the existing methodology in order to obtained improved estimation efficiency, (reducing the number of tests required to estimate a particular model), or to address specific industrial applications, e.g. Box and Jones (1989).

2.1.4 Second-order models and designs

Second-order experimental designs have attracted a large amount of attention, both by researchers and by practitioners of RSM. They offer an attractive compromise between the

simplicity, and often inadequacy, of designs for linear models, and the large number of tests required by higher order designs. Three-level versions of the factorial and fractional factorial designs used for fitting linear models may be employed, but these tend to be large and inefficient unless the number of design variables is extremely small. A selection of three-level fractional factorial designs was published by the National Bureau of Standards (1959). These designs comprise 3^{n-p} tests, and are presented for n = 4, ..., 10 and p = 1, ..., 5. The use of these designs was discussed by Zelen and Connor (1959).

The scope for constructing different experimental schemes is far greater for second-order designs than for linear designs, and this is reflected in the number of competing arrangements which have been put forward. The family of Central Composite Designs was introduced by Box and Wilson (1951), in which a two level fractional factorial design is augmented with an additional (2n + 1) tests, in order to allow estimation of the pure quadratic components of a response surface. Hartley (1959) suggested a class of small composite designs, which are a variation on the CCD. A number of three-factor, three-level designs were discussed by DeBaun (1959), some of which were found to perform better than the full factorial design. Box and Behnken (1960) introduced a class of incomplete three level fractional factorial designs, which are generally rotatable, or almost rotatable, and which block orthogonally (see Section 2.10). Box-Behnken designs are, however, only available for certain problem sizes (n = 3–7, 9–12, 16).

A class of saturated designs, in which the number of parameters to be estimated is equal to the number of tests in the design, was introduced by Box and Draper (1971), for n = 2 and n = 3 variables. These designs were generalised for $n \geq 4$ by Box and Draper (1974). A family of economical designs for fitting the strict quadratic model was developed by Hoke (1974), generalising earlier designs by Rechtschaffner (1967). These designs are based on partially balanced irregular fractions of the 3^n factorial, and are valid for any number of variables $n \geq 3$. Although only requiring a small number of tests, Hoke showed that his designs compared favourably with both Box-Behnken designs and the CCD's of Hartley (1959). A further investigation by Lucas (1976) compared the Hoke designs with the CCD and Box-Draper designs, and found that the Hoke designs performed better than the saturated designs of Box-Draper, and nearly as well as the CCD, although requiring far fewer tests.

Doehlert (1970) and Doehlert and Klee (1972) introduced uniform shell designs, which are generated from the points of a regular simplex, such that the points lie on concentric spherical shells. The disadvantages of these designs are that they require a large number of test points and variable levels. Lucas (1976) found that they did not perform as well as either the Box-Behnken design or the CCD. Hybrid designs were introduced by Roquemore (1976), and are constructed from a CCD of dimension n–1, augmented with an extra row. The specification of this extra test is determined in such a way as to achieve a similar degree of orthogonality as the CCD, whilst also being nearly rotatable.

Among the many other schemes which have been put forward, an early example of a family of saturated designs is given by Koshal (1933), from which it is possible to construct specific designs of any order d in n variables. Designs which are based on irregular fractions of factorials, and are very nearly saturated, were introduced by Westlake (1965) for n = 5, 7 and 9 variables. As is the case for first-order designs, much recent work has aimed at reducing the number of tests which need to be performed, whilst maintaining an acceptable level of accuracy in the estimation of parameters. Examples of this approach are the small composite designs of Draper (1985) and Draper and Lin (1990a). A set of saturated designs which are constructed using tests which form part of the three level factorial design were developed by Notz (1982).

An additional area of research concerns the number of centre points which need to be included in a second-order experimental design, and work in this area has been carried out by Lucas (1977) and Draper (1982), both of whom gave guidelines for the selection of such points.

2.1.5 Designs of higher dimensionality

Experimental designs of order greater than two have received relatively little attention by RSM researchers and practitioners, mainly due to the increased number of test points which are required in order to estimate the parameters of a higher order model. When carrying out experiments in which random experimental error is of importance, it is often better to use any additional tests to obtain better estimates of the parameters of a second order model than to fit a model of higher dimensionality.

Third-order designs are discussed by Gardiner, Grandage and Hader (1959), Draper (1960b, 1960c, 1961, 1962), Herzberg (1964) and Huda (1987). Fourth-order designs and models are discussed by Huda and Shafiq (1987), Arap Koske (1987) and Arap Koske and Patel (1987).

2.1.6 Measures of design optimality

Various criteria have been developed for judging the performance of an experimental design, many based on the optimal design theory work carried out by Kiefer and Wolfowitz, dating back to the late 1950's. The most widely used measure of optimality is that of D-efficiency, with a D-optimal design being one in which the generalised variance of the coefficients is a minimum. A related measure is that of G-efficiency, in which the maximum prediction variance within the design variable space is minimized, and both of these criteria were used by Lucas (1974) to compare four types of composite designs. Lucas (1976) also used

these two measures to compare a number of standard experimental designs, including the CCD (Box and Wilson, 1951), Box-Behnken designs (1960), uniform shell designs (Doehlert, 1970), the designs of Hoke (1972), Box-Draper saturated designs (1971), and a class of D-optimal designs proposed by Pesotchinsky (1975). Further comparisons, using these and other measures of optimality are given by Silvey (1980), Bandemer (1980), and Atkinson (1982).

A related area of research has been the development of algorithms to construct D-optimal designs. Early work in this field was carried out by Wynn (1970) and Federov (1972), and computer programs have been developed for the purpose by Federov (1972) and Mitchell (1974a, 1974b). In an improved version of Mitchell's algorithm, Galil and Kiefer (1980) presented a family of computer-search methods to identify optimum designs. Cook and Nachtsheim (1980) review a number of algorithms which are available for this purpose.

2.1.7 Locating the optimum

Although obviously of great importance in the practical use of an integrated RSM-based approach to optimization, techniques for locating the unconstrained or constrained optimum of a function are not the main focus of attention for those involved in the use of response surface methods. These methods form a separate body of knowledge which will not be reviewed in detail here. Within the present work a standard numerical optimization algorithm from a commercial subroutine library has been used to carry out optimization studies (see Chapter 9). In order to give a historical perspective to the search for optimum conditions, however, it is useful to consider a number of the techniques which were in general use before advanced numerical optimization routines became available.

Early work on optimization was concerned mainly with the identification of stationary points on a response surface (unconstrained optimization), and often relied on graphical methods to visualise the shape of the response surface. The use of such techniques is, of course, extremely limited when the number of design variables exceeds three. Two techniques which have been widely used for identifying optima are the method of steepest ascent, and ridge analysis.

The method of steepest ascent is described by, amongst others, Davies (1954) and Khuri and Cornell (1987), and involves locally approximating the response surface as a (linear) hyperplane, and using this equation to determine the direction in which the response is expected to improve most rapidly. Experiments are performed along this "line of steepest ascent" until curvature of the experimental response surface causes the result to deviate from the prediction of the hyperplane. At this point, a new hyperplane approximation is derived, and the procedure continues iteratively until no further improvement is gained. The chief disadvantage of this

approach would appear to be that it is highly susceptible to local gradient information, and can thus only be expected to perform well on smooth surfaces. Ridge analysis was introduced by Hoerl (1959) and formalised by Draper (1963). Its purpose is to locate the maximum value of a response surface when no stationary point lies within the bounds of the design variable space.

2.1.8 Other research directions

Among the many other areas of research which have been investigated by RSM workers, the ability to deal efficiently with multiple response functions is of some considerable importance, since "nearly all practical RSM problems are truly *multiple* response in nature" (Myers, Khuri and Carter, 1989). If the responses which are of interest can be assumed to be independent, then standard experimental designs may readily be employed, and each of the resulting response surface generated in the usual way. This approach is followed in the present work. If, however, there is significant interaction between responses, as may occur, for example, in chemical processes (see for example Ziegel and Gorman, 1980), then an alternative methodology may be adopted, such as that of Box and Draper (1965), which accounts for these effects. A more recent example of this approach is given by Bates and Watts (1985).

Whichever method is followed to generate the analytical response surfaces for each function, some means of performing a multi-response optimization is required. This branch of optimization is much less advanced than the methods used for single-response optimization, with only a small number of commercially available algorithms and computer routines. Some early attempts at multi-response optimization, such as those by Myers and Carter (1973) and Biles (1975) simply carried out a constrained optimization of a single objective function. A different technique, based on the concept of a *desirability function* was used by Harrington (1965), and more recently by Derringer and Suich (1980). This achieved a *compromise* optimum, the location of which depended on the weighting, or desirability, ascribed to each objective function. In a similar manner, Khuri and Conlon (1981) used a *distance function* to describe how far each response was from its optimal value.

Another area of research which is gaining in importance is the ability to deal with large numbers of design variables. The standard experimental designs are usually quoted for ≤ 10 or 12 design variables, and after this become impractically large in size. In many applications, however, the number of potential design variables may be many tens or even hundreds, of which only a relatively small number may have a significant effect on the response functions of interest. In such cases it may be impractical to carry out even a simplified two-level design in order just to estimate main effects. One possible strategy to deal with this situation is the *Group Screening Method*, discussed by Watson (1961). Using this technique, the design variables are arranged in a number of groups, and a standard experimental design is carried out, in which the

variable levels of each member of a group are varied together. If certain assumptions are made regarding the nature of the effect of each of the variables, then any group which is found to have an insignificant effect on the response can be considered to have no significant variables. Groups which do have an effect on the response functions are divided into smaller groups, and the process repeated until either the number of potentially active variables is reduced to a manageable size, or the group size is reduced to one. Similar approaches have been taken by Srivastava (1975) and Morris (1987)

An alternative approach to screening was taken by Welch, Buck, Sacks, Wynn, Mitchell and Morris (1989), who present an example in which the six important factors are identified from a 20-dimensional input, curvature and interactions are detected, and a predictive equation generated, with just 30 tests.

2.1.9 The application of RSM to computer "experimentation"

The use of response surface methods to investigate the effect of varying the inputs to a computer simulation program has received very little attention in the published literature, although its use appears to have become more widespread in recent years. An early example is given by McKay, Conover and Beckman (1979), who used Latin hypercube sampling to select values of input variables to a computer code. Iman and Helton (1988) used a fractional factorial design to investigate a computer model, fitting the response surface using a least squares criterion. They found that the linear response surface model was often inadequate to represent the complexity of the computer model throughout the range of the input variables, but was useful in ranking the effects of these variables. Further investigation of the analytic response surface was carried out using Latin hypercube sampling and Monte Carlo sampling.

Sacks, Welch, Mitchell and Wynn (1989) modelled the deterministic output of a computer code as the realisation of a stochastic process, and drew a number of important distinctions between computer experiments and physical experimentation. A similar approach was taken by Welch, Buck, Sacks, Wynn, Mitchell and Morris (1989), who used a method known as kriging (Matheron, 1963), and compared the classical frequentist approach with a Bayesian method of prediction.

Sacks, Schiller and Welch (1989) use this technique to model departures from a first- or second-order model, and present examples in which the test point coordinates of a given design are optimised to minimise the integrated mean square error (IMSE) over the design variable space. They present an example in which a seven dimensional CCD requiring 79 tests is used to investigate a methane combustion process.

2.1.10 Additional references

Reviews of RSM have been published by Hill and Hunter (1966), Mead and Pike (1975), and Myers, Khuri and Carter (1989). The three main texts on the subject are Myers (1971), Box and Draper (1987), and Khuri and Cornell (1987).

2.2 RSM concepts and definitions

In the remaining sections of this chapter, a number of the concepts, definitions, conventions and methods of analysis of RSM are introduced. More detailed descriptions of many of the topics are available in the published literature, and suitable sources are referenced where appropriate. The method of coding variable values is first described (Section 2.3), together with the effect which this has on the shape of the design variable space (Section 2.4). Following this, the terminology used to describe mathematical models and experimental designs is introduced (Section 2.5 and 2.6). The properties of orthogonality, rotatability and permutation invariance are then described (Sections 2.7–2.9), as well as the practice of arranging tests in blocks (Section 2.10). The characteristics of the *moment matrix* are described in some detail in Sections 2.11 and 2.12, since reference to these results is made in the discussions of Chapters 3–8. The final two sections cover the method of least squares (Section 2.13) and methods of testing lack of fit (Sections 2.14 and 2.15).

2.3 Coding of variables

When the set of input variables to be used in a particular investigation has been identified, it is often found that the values of the bounds between which the variables may be varied is significantly different for each of the factors involved. As an example, Table 2.1 shows the variable bounds in the original units, together with the mean and range, used in an investigation of the behaviour of worsted yarn under cycles of repeated loading, reported by Box and Draper (1987, p.28)

	Variable	bounds lower	bounds upper	mean	range
A	Length of test specimen (mm)	250.0	350.0	300.0	100.0
B	Applied load (g)	40.0	50.0	45.0	10.0
C	Amplitude of load cycle (mm)	8.0	10.0	9.0	2.0

Table 2.1 Variable values for the worsted yarn experiment

It can be seen that in this example the largest and smallest variable ranges differ by a factor of 50, with the result that, even if the effect of each variable on the response function were equal, the resulting coefficient values would differ significantly. This difference leads to difficulties in interpreting the relative effect of each of the design variables, and may also lead to problems when using iterative optimization methods to locate a minimum on the fitted surface, since the search direction is often determined by local gradient information, calculated in variable units.

In order to eliminate this illusory difference in variable effects it is usual to scale the units of each variable such that standardised, or 'coded', variable bounds are obtained, which are the same for each factor. These coded levels are obtained by applying the transformation shown in (2.1), which results in values of ±1 for the bounds of each variable, and a value of 0 for each variable mean. For each variable i, the original variable value X_i is converted into the dimensionless value x_i.

$$x_i = \frac{X_i - M_i}{\Delta_i} \qquad (2.1)$$

where $M_i = \frac{X_{hi} + X_{li}}{2}$ is the mean of high and low bounds for variable i

$\Delta_i = \frac{X_{hi} - X_{li}}{2}$ is the half-range of variable i

= the distance from the mean to either of the variables bounds

X_{li} and X_{hi} are the lower and upper variable bounds respectively

As an example, the upper bound of variable A, from Table 2.1, is coded as

$$x_{hA} = \frac{X_{hA} - M_A}{\Delta_A} = \frac{350.0 - 300.0}{50.0} = 1$$

The important property of equal coded range for each factor may be obtained solely by division by the variable half-range, with subtraction of the mean value giving the convenient additional feature of distribution about a zero value. An additional advantage of using this standardised notation is that, when each of the variables is tested only at its mean or bounded values, the appearance of integer values in the design matrix leads to increased simplicity in the estimation of model parameters. Note that, in some experimental designs, tests may not be carried out exactly at variable bounds, in which case the 'test range' will not be equal to the 'bound range' for a given variable, although if the design is *permutation invariant* (see Section 2.9), then this test range will be the same for each variable. Note also that if the tests to be performed are not symmetric about the mean level M_i of each variable, then the mean, over the N tests, of the coded test values

$$\bar{x}_i = \frac{1}{N}\sum_{u=1}^{N} x_{ui}$$

will not be equal to the mean of the coded bounds, $m_i = 0$ (see also Section 2.12). Further details of the use of coded variables can be found in Khuri and Cornell (1987) p.46, and Box and Draper (1987) p.20.

2.4 The effect of variable coding on the shape of the design region

Any particular design configuration, defined as a combination of n variable values, can be visualised as lying within an n-dimensional design variable space. If each of the variables can independently vary between its bounded values, then the shape of the design variable space is that of a hyper-rectangle. The effect of the variable coding described in the previous section is to normalise the bounds on each of the variables to ±1, so that the coded design variable space is a hypercube. A normalised design region in three design variables is shown in Figure 2.1.

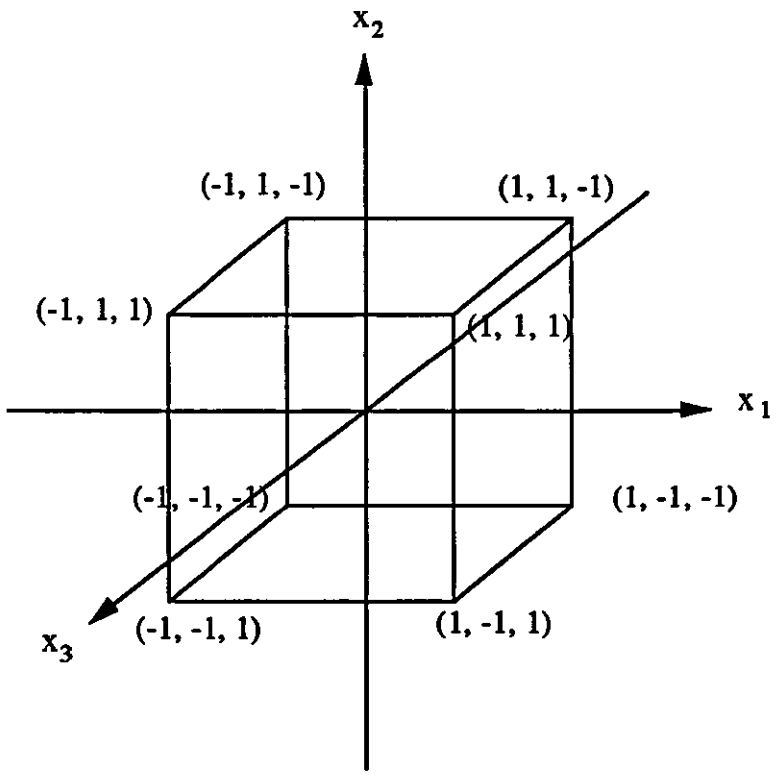

Figure 2.1. A cuboidal design space in three dimensions

2.5 Terminology for mathematical models and experimental designs

A general polynomial model in n variables may be regarded as consisting of T terms, each of the form

$$\gamma_t x_1^{\alpha_{1t}} x_2^{\alpha_{2t}} \cdots x_n^{\alpha_{nt}} \qquad (t = 1, ..., T) \tag{2.2}$$

where the $x_1, x_2, ..., x_n$ are the n variable values, α_{it} $(i = 1, ..., n ; t = 1, ..., T)$ is the index of the i^{th} variable in the t^{th} term, and γ_t is the coefficient of the t^{th} term. In order to complete the definition of a statistical model, an error term ε is added to (2.2). A model containing terms for which the α_{it} $(i = 1, ..., n)$ are zero for all t, for example, would be as shown in (2.3)

$$Y = \sum_{t=1}^{T} \gamma_t + \varepsilon = \beta_0 + \varepsilon \tag{2.3}$$

Such a model, however, is of little use as a predictive tool, and in subsequent chapters a number of more complex mathematical models are used to represent the variation of the response functions throughout the design variable space. In order to distinguish between models of similar specification, the following convention has been adopted.

A *linear model* is one in which no term contains a variable which has an index other than 0 or 1. Such a model may also be termed a first-order model. The simplest of these models is one in which each term contains no more than one variable which has a non-zero index, and an example of this is shown in equation (2.2). This is referred to as either a *strict linear* model or a *main effects* model.

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon \tag{2.4}$$

In contrast with the main effects model, a linear model which contains a number of terms in which more than one variable has a non-zero-index is described as a *linear + interactions* model. Each of the terms in which two variables have a non-zero index may be described as mixed quadratic terms, with each term containing three variables being a mixed cubic term, and so on. However, each of these terms represents an interaction between linear effects in each of the variables represented (see Section 3.2), and thus may also be described as a *linear interaction term.* Equation (2.5) shows a linear + interactions model which contains all possible interaction terms, from the [n(n-1)/2] two-way interactions to the single n-way interaction.

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \beta_{ijk} x_i x_j x_k + \sum_{i=1}^{n-3} \sum_{j=i+1}^{n-2} \sum_{k=j+1}^{n-1} \sum_{l=k+1}^{n} \beta_{ijkl} x_i x_j x_k x_l + \cdots\cdots \beta_{123\ldots n} x_1 x_2 x_3 \cdots x_n + \varepsilon \quad (2.5)$$

A model containing terms in which variables may have an index of 0, 1 or 2 is termed a *quadratic*, or *second-order*, model. A *strict quadratic* model is one in which, for each term, the sum of indices of the variables is not greater than 2. The terms which may appear are thus the mean, the main effect terms x_i, (i = 1, n), the two-way interaction (or mixed quadratic) terms $x_i x_j$, (i = 1, n, j = i, n), and the pure quadratic terms x_i^2, (i = 1, n). Such a model is shown in equation (2.6).

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j=i}^{n} \beta_{ij} x_i x_j + \varepsilon \quad (2.6)$$

In addition to the terms of equation (2.6) a *quadratic + interactions* model may contain any other term in which no variable has an index other than 0, 1 or 2. These include the remaining k-way, $(2 < k \leq n)$, interaction terms of (2.5), as well as terms of the form

$$x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} x_4^{\alpha_4} \cdots x_n^{\alpha_n} \qquad \text{where } \sum_{i=1}^{n} \alpha_i > 2$$

These latter terms represent all possible interactions between quadratic effects, and between these quadratic effects and linear effects. A model containing all such terms is shown in equation (2.7)

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n-1} \sum_{j=i}^{n} \sum_{\substack{k=j \\ k>i}}^{n} \beta_{ijk} x_i x_j x_k + \sum_{i=1}^{n-1} \sum_{j=i}^{n-1} \sum_{\substack{k=j \\ k>i}}^{n} \sum_{\substack{l=k \\ l>j}}^{n} \beta_{ijkl} x_i x_j x_k x_l + \cdots\cdots \beta_{112233\ldots nn} x_1^2 x_2^2 x_3^2 x_4^2 \cdots x_n^2 + \varepsilon \quad (2.7)$$

The above convention can, of course, be extended to models of order greater than two, with, for example, a *strict cubic* model being one in which, for each term, the sum of indices of the variables is not greater than 3, and a *quartic + interactions* model being one which may contain any term in which each variable has an index in the range 0 - 4. In the present work, however, only models of order one and two will be considered in detail.

The above conventions also apply to experimental designs, with an *n^{th}-order* design being one which, as a minimum requirement, allows for the estimation of all the parameters of a strict n^{th}-order model. An n^{th}-order design may also allow for estimation of some or all of the additional parameters of an n^{th}-order + interactions model. Note that a design which allows the estimation of a full n^{th}-order + interactions model may also be used to fit a full k^{th}-order + interactions or strict k^{th}-order model, $(0 \leq k \leq n-1)$, whilst a design which allows estimation of a strict n^{th}-order model may be used to fit any strict k^{th}-order model, $(0 \leq k \leq n-1)$. Due to a desire to minimize the number of function evaluations which need to be performed, however, the ability of an experimental design of a given order to yield models of a lower order is seldom of practical benefit.

2.6 Saturated designs

The *saturation ratio* of a model/design combination may be defined as the ratio of the number of terms to be fitted to the number of tests in the design. A design in which the number of tests is equal to the number of coefficients which are to be estimated is termed a *saturated design.*

2.7 Orthogonality

An orthogonal design is one in which the parameter estimates of the terms in the fitted model are uncorrelated with one another. The prediction variance at any point in the design variable space is then expressible as a weighted sum of the variances of the parameter estimates in the model. The importance of this property lies in the fact that the independence of the parameter estimates facilitates the appraisal of the parameter values obtained, as well as the comparison of relative precision of the parameter estimates for different terms of the model. In the present work, this independence is fundamental to the application of the probability plot technique, described in Appendix 4C, for testing the statistical significance of individual parameters.

In practice, a small degree of correlation between parameter estimates will not normally invalidate conclusions drawn from either probability plots or variance comparisons, and a requirement for exact orthogonality is often somewhat relaxed in cases where its attainment would lead to a large increase in the number of function evaluations required. A related benefit of orthogonal designs is that the precision of the parameter estimates which are obtained is greater than that for comparable non–orthogonal designs (Box and Draper (1987), p.79).

2.8 Rotatability

A rotatable design is defined as one in which the the prediction variance is the same at all points which lie at the same distance from the centre of the design variable space, so that surfaces of constant prediction variance form concentric hyperspheres. One of the advantages of a rotatable design is the fact that the variances of individual parameters, and the covariance effects between them, are unaltered if the design is rotated relative to the variable axes. Orthogonality is thus maintained under design rotation, so that prediction precision will not be adversely affected if the principal axes of the response surface are not aligned with the variable axes (Box and Draper (1987), p.484). For first-order designs, the requirements for orthogonality and rotatability are, in fact, identical (Box and Draper (1987), p.483). When investigating a cuboidal design region, however, concentric hyperspheres of constant prediction variance would not appear to be of any particular benefit. The identification of a constrained optima on low order polynomial surfaces, for example, often identifies optimum variable combinations which lie at, or close to, one of the vertices of the design variable space (see Chapter 9), where prediction variance would be highest.

A much more useful design, when carrying out optimization within the design variable space, would be one which resulted in the prediction variance being approximately constant within the region, since the maximum variation in prediction variance is arguably of more importance than the exact shape of the variance contours. This requirement for approximately constant variance can be met by employing a special type of rotatable design known as a *uniform precision* design, in which the prediction variance at points lying on a hypersphere of radius 1.0 is the same as that at the design centre. Second-order uniform precision designs are discussed in Section 5.3, below, where it is shown that strict conditions must be fulfilled in order to obtain them, often requiring many extra function evaluations, and that they cannot also be orthogonal. For these reasons the property of exact rotatability is often compromised in the search for designs which have a small test requirement and are also orthogonal. Methods are available which can be employed to assess the degree of rotatability of a design (Khuri, 1985), although their use is not investigated here.

2.9 Permutation invariance

A further characteristic which is automatically possessed by rotatable designs is that the precision of the parameter estimates for each class of term are invariant under permutation of the factors in the model (permutation-invariant, for short, see Hoke, (1974), p.376), so that the estimates of the parameters do not change if the order of the variables is changed. Thus, all the parameters of the same form are measured with equal precision; e.g. the variance of parameter

β_{ii} is independent of the value of i, and similarly for β_i and β_{ij}, etc..

This is clearly an important property, since the ability accurately to compare different variables depends heavily on the availability of equally precise information concerning the effect of each of the factors. Additionally, it is often not possible to estimate in advance which of the input variables is likely to have the greatest effect on the response functions, which would necessarily lead to difficulties in determining the appropriate order of variables if the design to be used were not permutation-invariant. This problem is compounded by the fact that certain of the variables may have a large effect on one or more of the computed responses, but a smaller, or even negligible, effect on other responses. Indeed, it is shown in Section 9.5 that, when carrying out an optimization study, it is advantageous to include some variables which possess just this property.

Although the majority of the commonly used experimental designs are inherently permutation-invariant, a number of the more advanced schemes for estimating the parameters of second-order models deviate substantially from this ideal, in the search for designs which reduce the number of function evaluations to the theoretical minimum (see Chapter 8). Due to the requirement for mathematical models which will allow optimization to be carried out within the design variable space, experimental designs which are not permutation-invariant have not been investigated within the current work.

2.10 Blocking of tests

When carrying out experimental work of any nature, it is important that all test runs carried out as part of an experimental design are conducted under controlled conditions, so that variations between test runs only occur in those variables which are being actively investigated. Studies involving the sampling of output from a manufacturing process, for example, should ideally test products made by the same machine from the same batch of raw materials, and manufactured on the same day. If this is not possible for some reason, such as insufficient raw material being available in a single batch, then the resulting inconsistency between tests becomes an additional variable in the experimental design.

The effect of an additional variable due to differing experimental conditions is often of no interest in itself, but is of some considerable concern in that these variations may bias the estimates of the effects of other variables. A number of experimental designs have been developed in order to address this problem; see for example Box and Behnken (1960), Khuri and Cornell (1987), Box and Draper (1987). Using these schemes, the required tests are divided into a number of groups, or 'blocks', with uncontrolled variables, such as raw material batch or day of manufacture, only differing between these blocks, and not within them. The division of tests into blocks is carried out in such a way that the effect of any uncontrolled

operating conditions is orthogonal to the effect of any of the variables being investigated, and hence will not bias the estimation of the parameter coefficients. This method of orthogonal blocking ensures that estimation of design variable parameters is independent of any unwanted effect which may vary between blocks.

When carrying out an "experimental" programme using computer simulation, however, all relevant conditions are exactly repeatable. In this context, blocking of tests is not required, and hence the use of an experimental design scheme which allows orthogonal blocking of tests is of no particular advantage. Indeed, a number of these designs have a significant disadvantage, since additional function evaluations are often required in order to ensure that the tests block orthogonally. In selecting suitable experimental designs with which to investigate the engine noise problem, the ability to carry out orthogonal blocking of tests has thus not been a consideration, and the evaluation of designs which do posses this feature, as in Section 8.1, has been conducted solely on the basis of other relevant characteristics of the design.

2.11 The D, X and M matrices

The specification of an experimental design can be expressed in terms of the *design matrix*, **D**, which contains the combination of variable values to be used for each run. The design matrix for a three-level full factorial design in one dimension would thus be

$$\mathbf{D} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{2.8}$$

Each row of the *regressor matrix*, **X**, is formed by calculating the values of each of the terms of the model equation which result from the appropriate test combination. Thus, if the coefficients of the model

$$Y = \beta_0 + \beta_1 x + \beta_{11} x^2 \tag{2.9}$$

were to estimated from the above design, the appropriate regressor matrix would be

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{2.10}$$

The *moment matrix* is then defined as $\mathbf{M} = N^{-1}\mathbf{X}'\mathbf{X}$, where N is the number of tests in the design, and $\mathbf{X}'$ represents the matrix transpose of **X**. For further details see Khuri and Cornell (1987), p. 54.

2.12 Characteristics of the moment matrix

The discussion in this section is based on that of Khuri and Cornell (1987, pp. 54–57, 60–61). An experimental design may be assessed for rotatability by examining the elements of the moment matrix. For a first-order model in n variables, for example, the **X** matrix in the coded variable levels is of the form $(\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$. The moment matrix may then be written as follows.

$$\frac{1}{N}\mathbf{X'X} = \begin{array}{c} \\ 1 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & x_1 & x_2 & x_3 & \cdots & x_n \end{array} \\ \left[\begin{array}{cccccc} 1 & [1] & [2] & [3] & \cdots & [n] \\ & [11] & [12] & [13] & \cdots & [1n] \\ & & [22] & [23] & \cdots & [2n] \\ & & & [33] & \cdots & [3n] \\ & \text{symmetric} & & & \ddots & \vdots \\ & & & & & [nn] \end{array}\right] \end{array} \tag{2.11}$$

where, for $i, j = 1, ..., n, \quad i \neq j$,

$$[i] = \frac{1}{N}\sum_{u=1}^{N} x_{ui} \qquad [ii] = \frac{1}{N}\sum_{u=1}^{N} x_{ui}^2 \qquad [ij] = \frac{1}{N}\sum_{u=1}^{N} x_{ui}x_{uj} \tag{2.12}$$

It can be seen that the value [i], called the first-order moment in the i^{th} variable, is equal to the mean of the test levels x_{ui} of the i^{th} variable over the N tests. Similarly, the pure second-order moment [ii] is the mean of the squared variable values x_{ui}^2. The value [ij], $i \neq j$, is termed the mixed second-order moment. In order to simplify the moment matrix, it is usual to apply a scaling transformation to the coded variables, chosen such that the mean of the test values in each variable is zero, and hence [i] = 0. Clearly, an additional criterion of [ii] = 0 is unattainable unless all the x_{ui} are equal to zero, but a simplification of the moment matrix may be obtained by scaling the variable values such that [ii] = 1.

The first of these criteria may be fulfilled by subtracting from each of the uncoded test values X_{ui}, u = 1, ..., N, the mean value of the i^{th} variable over the N tests, such that

$$x_{ui} = X_{ui} - \overline{X_i} \tag{2.13}$$

$$\text{where } \overline{X_i} = \frac{1}{N}\sum_{u=1}^{N} X_{ui} \tag{2.14}$$

Note that, if the test specifications are symmetric about the mean of the variable bounds M_i, then the mean of the test values will be equal to zero.

It can be seen from (2.12) that the requirement [ii] = 1 is met when

$$\sum_{u=1}^{N} x_{ui}^2 = N$$

and this may be obtained by applying a scaling factor c_i to each of the N test values, such that

$$\sum_{u=1}^{N} (c_i X_{ui})^2 = N \tag{2.15}$$

Thus,

$$c_i = \left[N / \sum_{u=1}^{N} X_{ui}^2 \right]^{1/2}$$

giving

$$x_{ui} = \frac{X_{ui}}{\left[\frac{1}{N} \sum_{u=1}^{N} X_{ui}^2 \right]^{1/2}} \tag{2.16}$$

The two criteria [i] = 0 and [ii] = 1 can be met simultaneously by first transforming the variable values to a zero mean using (2.13), and then scaling to give unit mean square by (2.16). The following relationships can then be formed:

From (2.13),

$$x_{ui}^{(1)} = X_{ui} - \overline{X_i} \tag{2.17}$$

By substitution into (2.16),

$$x_{ui} = \frac{x_{ui}^{(1)}}{\left[\frac{1}{N} \sum_{u=1}^{N} (x_{ui}^{(1)})^2 \right]^{1/2}}$$

$$x_{ui} = \frac{X_{ui} - \overline{X_i}}{\left[\frac{1}{N} \sum_{u=1}^{N} (X_{ui} - \overline{X_i})^2 \right]^{1/2}} \tag{2.18}$$

or

$$x_{ui} = \kappa_i \left[X_{ui} - \overline{X_i} \right] \tag{2.19}$$

where

$$\kappa_i = \left[\frac{1}{N} \sum_{u=1}^{N} (X_{ui} - \overline{X}_i)^2 \right]^{-1/2}$$

It can be seen from inspection of this equation that, in general, the attainment of these two criteria will not yield equal coded ranges for each variable, since the value of the scaling factor κ_i is dependent on the particular distribution of test points in the i^{th} variable. Thus, the adoption of this coding convention does not ensure the fulfilment of the requirement of Section 2.3, that the range of each of the variables be equal, as specified in equation (2.1):

$$x_{ui} = \frac{X_{ui} - M_i}{\Delta_i}$$

However, if the normalised distribution of test values is the same for each variable, which is a necessary condition for the design to be permutation invariant, then the value of the scaling factor κ_i will be the same, as a proportion of the variable range, for each variable. Thus, for the class of designs which are invariant with respect to permutation of the variable order, (which includes all designs considered within the present work), the coded test range of each factor is identical, although not generally equal to ±1, and the requirement of Section 2.3 is fulfilled. Note that (2.18) is of a similar form to equation (2.1), with the mean of the actual test values replacing the mean of the variable bounds, and the denominator providing a measure of the spread of the test points in the direction of the X_i-axis. Indeed, if the test points are symmetrically distributed about the mean of the variable bounds, with only bounded values being tested (as for the factorial and fractional factorial designs introduced in Chapter 3), then the two expressions are identical.

As discussed above, the effect of the scaling convention of (2.18) is to simplify the moment matrix by producing values of [i] = 0 and [ii] = 1, so that the moment matrix of equation (2.11) is of the form

$$\frac{1}{N}\mathbf{X'X} = \begin{array}{c} \\ 1 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & x_1 & x_2 & x_3 & \cdots & x_n \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ & 1 & [12] & [13] & \cdots & [1n] \\ & & 1 & [23] & \cdots & [2n] \\ & & & 1 & \cdots & [3n] \\ & \text{symmetric} & & & \ddots & \vdots \\ & & & & & 1 \end{bmatrix} \end{array} \qquad (2.20)$$

In the general case of a design in n variables, the elements of the moment matrix (also known as *design moments*), take the following form:

$$[1^{\delta_1} 2^{\delta_2} \ldots n^{\delta_n}] = \frac{1}{N} \sum_{u=1}^{N} x_{u1}^{\delta_1} x_{u2}^{\delta_2} \ldots x_{un}^{\delta_n} \tag{2.21}$$

where each of the indices δ_i (i = 1, n) is a non-zero integer. A design moment for which the sum of the indices $(\delta_1 + \delta_2 + \ldots + \delta_n) = \delta$ is termed a design moment of order δ, and, in general, a design of order d will have design moments of order $\delta = 0, 1, 2, \ldots 2d$. Thus, a second-order design will have design moments of order up to 4. For such a design, the **X** matrix in the coded variable levels is of the form $(1, x_1, x_2, \ldots, x_n, x_1^2, x_2^2, \ldots, x_n^2, x_1x_2, x_1x_3, \ldots, x_{n-1}x_n)$, so that, for a two-dimensional problem:

$$\frac{1}{N}\mathbf{X'X} = \begin{array}{c} \\ 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_2^2 \\ x_1x_2 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \end{array} \\ \left[\begin{array}{cccccc} 1 & [1] & [2] & [11] & [22] & [12] \\ & [11] & [12] & [111] & [122] & [112] \\ & & [22] & [112] & [222] & [122] \\ & & & [1111] & [1122] & [1112] \\ & \text{symmetric} & & & [2222] & [1222] \\ & & & & & [1122] \end{array}\right] \end{array} \tag{2.22}$$

The notation for array elements follows that of (2.11); thus, in addition to (2.12), for the general case of an n–dimensional design, and for i, j = 1, ..., n, $i \neq j$,

$$[iii] = \frac{1}{N}\sum_{u=1}^{N} x_{ui}^3 \qquad [iij] = \frac{1}{N}\sum_{u=1}^{N} x_{ui}^2 x_{uj} \qquad [iiii] = \frac{1}{N}\sum_{u=1}^{N} x_{ui}^4$$
$$[iiij] = \frac{1}{N}\sum_{u=1}^{N} x_{ui}^3 x_{uj} \qquad [iijj] = \frac{1}{N}\sum_{u=1}^{N} x_{ui}^2 x_{uj}^2 \tag{2.23}$$

As for a first-order design, adoption of the scaling convention of (2.18) leads to a simplified matrix:

$$\frac{1}{N}\mathbf{X'X} = \begin{array}{c} \\ 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_2^2 \\ x_1x_2 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \end{array} \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & 1 & 1 & [12] \\ & 1 & [12] & [111] & [122] & [112] \\ & & 1 & [112] & [222] & [122] \\ & & & [1111] & [1122] & [1112] \\ & \text{symmetric} & & & [2222] & [1222] \\ & & & & & [1122] \end{array}\right] \end{array} \tag{2.24}$$

Moment matrices for designs of order greater than two are constructed following a procedure analogous to that outlined above.

Khuri and Cornell (p.56, p.60) show that, if a design is rotatable, its moment matrix will be of a particular form. For a rotatable first-order design in n variables, for example, the moment matrix must be as follows:

$$\frac{1}{N}\mathbf{X'X} = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \lambda_2 \mathbf{I}_n \end{bmatrix} \tag{2.25}$$

where $\mathbf{I}_n$ is the n × n identity matrix. The sub-matrix $\lambda_2\mathbf{I}_n$ represents the [ii] entries of the moment matrix, which, under the scaling convention of (2.18), each have a value of 1. Thus, the scale factor λ_2 takes a value of 1. Additionally, since the moment matrix is diagonal, first-order rotatable designs are also orthogonal.

For a second-order rotatable design in n variables, the moment matrix is of the following form:

$$\frac{1}{N}\mathbf{X'X} = \begin{array}{c|cccc|cccc|cccc}
 & 1 & x_1 & \cdots & x_n & x_1^2 & x_2^2 & \cdots & x_n^2 & x_1x_2 & x_1x_3 & \cdots & x_{n-1}x_n \\ \hline
1 & 1 & \mathbf{0}' & & & \lambda_2 & \lambda_2 & \lambda_2 & \lambda_2 & & & & \\
x_1 & \mathbf{0} & & & & 0 & 0 & 0 & 0 & & \mathbf{0} & & \\
\vdots & & & \lambda_2 \mathbf{I}_n & & \vdots & \vdots & \vdots & \vdots & & & & \\
x_n & & & & & 0 & 0 & 0 & 0 & & & & \\ \hline
x_1^2 & \lambda_2 & 0 & \cdots & 0 & & & & & & & & \\
x_2^2 & \lambda_2 & 0 & \cdots & 0 & & \lambda_4(2\mathbf{I}_n + \mathbf{J}_n) & & & & \mathbf{0} & & \\
\vdots & \lambda_2 & 0 & \cdots & 0 & & & & & & & & \\
x_n^2 & \lambda_2 & 0 & \cdots & 0 & & & & & & & & \\ \hline
x_1x_2 & & & & & & & & & & & & \\
x_1x_3 & & \mathbf{0} & & & & \mathbf{0} & & & & \lambda_4 \mathbf{I}_p & & \\
\vdots & & & & & & & & & & & & \\
x_{n-1}x_n & & & & & & & & & & & &
\end{array} \tag{2.26}$$

where $\mathbf{J}_n$ is an n × n matrix of ones, and p = n(n-1)/2. Thus, for a rotatable second-order design, the only non-zero elements in the moment matrix are [ii] = λ_2, [iiii] = $3\lambda_4$ and [iijj] = λ_2. As for first-order designs, the effect of the scaling convention is to set $\lambda_2 = 1$. The value of the second parameter, λ_4, may be chosen to achieve certain other design characteristics, such as orthogonality. The selection of a suitable value for this parameter will be discussed further in Chapters 5 and 6.

2.13 The method of least squares

Once the required function values have been evaluated at each of the test points of the experimental design, a method is required which will estimate the terms of the chosen mathematical model. In selecting a method by which to carry out this fitting process, it should be remembered that the mathematical model will ultimately be used in the identification of an optimum variable combination within the design variable space, and that it is generally not known in advance in which area of the experimental region an optimum is likely to lie. Indeed, if a number of optimization procedures are to be carried out with differing objective functions, constraint functions or constraint values, then many areas of the design variable space may be of importance. In such a case, it is appropriate to use a method which will ascribe equal importance to the function values calculated in all parts of the design variable space. The method used in the current work, widely used for fitting equations to data, is the method of least squares, more detailed descriptions of which may be found in, for example, Press et al. (1986), p.499 et seq., Khuri and Cornell (1987), pp. 23-28, Box and Draper (1987), Chapter 3. This technique also has the advantage that a large number of computer routines are readily available, both from commercial software libraries (e.g. NAG, 1983) and from other sources (e.g. Press et al., 1986). The least squares method seeks to identify values of the unknown parameters of a model such that they minimise the sum, over all of the test points, of the squares of the errors between the predictions based on the fitted model and the calculated function values. A summary of the method is given in Appendix 2A.

2.14 Testing lack of fit

Following the construction of an approximating response surface, it is important that the next stage of any investigation be the validation of this mathematical model, in order to ensure that it is a sufficiently accurate representation of the original function which it seeks to approximate. This validation is carried out by first selecting a number of locations within the design variable space, and comparing the response calculated by the original analysis program at each of these points with the value predicted by the approximating model. The difference between these two values is referred to as the 'lack of fit' of the response surface at the point in question. The lack of fit which is found at the chosen points may be expressed in a number of ways, among which one of the most common is the analysis-of-variance (ANOVA) method.

The ANOVA method is a standard statistical approach to evaluating the quality of fit of a predictive model (see, for example, Box, Hunter and Hunter, 1978). Using this technique, a comparison is made between the amount of variation about a mean value which occurs in the computed function values and the amount of variation which occurs in the predicted values,

calculated using the approximating mathematical model. A summary of the method is given in Appendix 2B. Perhaps one of the greatest disadvantages of the ANOVA approach is that the results which are produced are often hard to interpret, especially when used by engineers or designers who have no formal training in statistical methods. Additionally, the measures of fit which are obtained give little indication as to which areas of the response surface are being accurately represented, and in which regions significant discrepancies are occurring. Such information would be particularly useful in determining whether substantial improvements in the accuracy of the approximating response surface could be gained by moving to a higher order model.

To avoid these concerns, an alternative approach has been taken in comparing the various experimental designs and predictive models which are discussed in Chapters 4 - 8. A measure of lack of fit is developed which presents the lack-of-fit data in a form which is directly related to the units of each of the original response functions, and is more amenable to interpretation by non-specialist users.

2.15 Assessment of lack of fit for optimization purposes.

Calculation of lack of fit is carried out for each function at all test points for which function value information has been computed. For the purposes of assessing lack of fit at different locations within the design variable space, these test points may be divided into a number of groups, such as, for example, the design points used to generate the model, and those extra points which have been specifically selected to provide additional lack of fit data. The categories of test points used in the present investigation are described in Section 4.1.1, for first-order designs, and Section 6.5.1 for second-order designs. The lack-of-fit error E at each point is calculated, as for the analysis-of-variance method, as the difference between the computed function value and the predicted value at this point. Thus,

$$E = (Y_u - y_u) \tag{2.27}$$

where Y_u is the computed function value at the u^{th} test

y_u is the predicted function value at the u^{th} test

This absolute error value may be quoted in the original function units, and provides a readily understandable measure of the performance of the predictive model at each point within the design region. Although this absolute error value is instructive in assessing the suitability of the mathematical model, a more useful measure would be one which seeks to normalise the error term in some appropriate fashion, and an initial choice might be to calculate the error as a percentage of the measured value. When carrying out an optimization study, however, an important consideration is the way in which the response surface would be distorted by such an

error, since it is this distortion which leads to incorrect identification of optimum variable combinations. A suitable measure of the error occurring at a test point would therefore seem to be one which links this error to the range over which the function in question varies throughout the design variable space. The error may thus be expressed as a percentage of the function range, as follows:

$$E_R = \frac{E}{F_h - F_l} \tag{2.28}$$

where the absolute error E is as defined above, and F_h and F_l are the extreme values of the function which occur within the design region.

The use of this normalised measure of lack of fit has the additional advantage that it enables general recommendations to be made as to when the fit of a model should be considered adequate. Errors which constitute less than, say, 5% of the function range are unlikely to cause substantial distortions in the overall shape of the response surface, and will hence not lead to large errors in the determination of an optimum. Errors of 15-20% or more would, however, usually give cause for serious concern, and in such cases the investigator should make further inquiries into the distribution of these errors, and consider either conducting a further set of computational function evaluations in order to provide additional data, or use a predictive model of a higher order. Between these ranges, the level at which errors become unacceptable is largely a matter of judgment, and will often depend upon experience of a particular application.

Appendix 2A

The method of least squares

Consider a case in which the measured response Y is representable as a strict quadratic model in two variables, plus some error term ε, as shown in equation (2A 1)

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \qquad (2A.1)$$

At each of the n test points, a measured function value Y_u, (u=1, . ., N), is obtained, and the error term ε_u at each point is thus

$$\varepsilon_u = Y_u - \beta_0 - \beta_1 x_{u1} - \beta_2 x_{u2} - \beta_{11} x_{u1}^2 - \beta_{22} x_{u2}^2 - \beta_{12} x_{u1} x_{u2} \qquad (2A.2)$$

The sum of the squares of the errors, ε_s^2, is then as shown in equation (2A.3), and it is this value which the least squares algorithm seeks to minimise.

$$\varepsilon_s^2 = \sum_{u=1}^{N} (Y_u - \beta_0 - \beta_1 x_{u1} - \beta_2 x_{u2} - \beta_{11} x_{u1}^2 - \beta_{22} x_{u2}^2 - \beta_{12} x_{u1} x_{u2})^2 \qquad (2A.3)$$

In order for the least squares estimates to be maximum likelihood estimates, a number of conditions must be placed on the error term ε in the statistical model (2A.1). These conditions on ε are as follows,

1. They are statistically independent.
2. They have zero mean and a common variance, σ^2.
3. They are normally distributed.

The validity of these conditions when carrying out computer 'experimentation' is discussed in Appendix 4C, although in any case the least-squares algorithm gives sensible results under a much wider range of conditions.

For the full set of N tests, the model of equation (2A.1) can be expressed in matrix notation as follows

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (2A.4)$$

where **Y** is a vector of measured function values, **X** is the regressor matrix, **β** is the vector of coefficients and **ε** is a vector of error terms, as follows;

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}^2 & x_{22}^2 & x_{21}x_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & x_{N1}^2 & x_{N2}^2 & x_{N1}x_{N2} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{12} \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

It can be shown that the least square estimates of the elements of β may be found by solving the normal equations

$$\mathbf{X'Xb} = \mathbf{X'Y} \tag{2A.5}$$

where $\mathbf{b} = (b_0, b_1, ..., b_{12})'$ are the least squares estimates of the coefficients $\beta_0, \beta_1, . ., \beta_{12}$, respectively, and may be calculated as

$$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'Y} \tag{2A.6}$$

The covariance of the parameter estimates is then given by equation (2A.7), where each diagonal element C_{ii} represents the variance of the i^{th} element of **b**, and each off-diagonal element C_{ij} represents the covariance between b_i and b_j.

$$\text{Cov}(\mathbf{b}) = \mathbf{C} = (\mathbf{X'X})^{-1}\sigma^2 \tag{2A.7}$$

The standard error of each b_i is equal to the positive square root of the variance value of that b_i, $| \sqrt{\text{Var}(b_i)} |$. The response $y(\mathbf{x}_m)$ which is predicted by the model at any point $\mathbf{x}_m = (x_{m1}, x_{m2})'$ is given by equation (2A 8),

$$y(\mathbf{x}_m) = \mathbf{X}_m'\mathbf{b} \tag{2A.8}$$

where $\mathbf{X}_m' = (1, x_{m1}, x_{m2}, x_{m1}^2, x_{m2}^2)$ is of the same form as a row of the **X** matrix of (2A 4). Finally, the variance of $y(\mathbf{x}_m)$, which is a measure of the precision of the prediction, may then be calculated as

$$\text{Var}[y(\mathbf{x}_m)] = \text{Var}[\mathbf{X}_m'\mathbf{b}] = \mathbf{X}_m'(\mathbf{X'X})^{-1}\mathbf{X}_m\sigma^2 \tag{2A.9}$$

Note that, for an orthogonal design (see Section 2.7), the matrix **X'X** is diagonal, so that the prediction variance at the point $\mathbf{x}_m$ is equal to the sum over i of $(X_{mi})^2 b_i$. Additionally, if the maximum value of each variable within the design region is 1 0, then the maximum prediction variance is equal to the sum of the variances of the individual coefficient estimates.

Appendix 2B

Analysis of Variance

The analysis-of-variance method compares the amount of variation about a mean value which occurs in the computed function values with the amount of variation which occurs in the predicted values, calculated using the approximating mathematical model.The variation which occurs in the computed function values is termed the total sum of squares (SST) and is calculated as follows:

$$SST = \sum_{u=1}^{N} (Y_u - \overline{Y})^2 \qquad (2B.1)$$

where Y_u is the computed function value at the u^{th} test

$\overline{Y}$ is the mean function value over the N tests, $\overline{Y} = (Y_1 + Y_2 + \cdots + Y_N) / N$

The number of statistical degrees of freedom associated with SST is N - 1. The variation which occurs in the predicted values at the same test points is referred to as the sum of squares due to regression (SSR). For a model which contains a total of p parameters (including the mean value), the number of statistical degrees of freedom associated with SSR is p - 1. The sum of squares due to regression is calculated as

$$SSR = \sum_{u=1}^{N} (y_u - \overline{Y})^2 \qquad (2B\ 2)$$

where y_u is the predicted function value at the u^{th} test

The lack of fit at each of the u =1, ., N selected points is calculated as the difference between the computed function value Y_u, and the predicted function value y_u. The sum of the squares of these values is termed the sum of squares of the residuals (SSE), and may be expressed as

$$SSE = \sum_{u=1}^{N} (Y_u - y_u)^2 \qquad (2B.3)$$

The number of statistical degrees of freedom associated with the sum of squared residuals is equal to the difference in number of degrees of freedom between SST and SSR, and is thus (N - 1) - (p - 1) = (N - p) These three values, together with their degrees of freedom are usually assembled in tabular form, as shown in Table 2B.1. This table highlights the fact that the total sum of squares variation is partitioned into two components. Firstly, the SSR represents the amount of the variation in data values which is being accounted for by the predictive model. The SSE then gives the remaining variation in the computed function values which is unaccounted for by the fitted model.

Source of Variation	Statistical Degrees of Freedom	Sum of Squares	Mean Square
Regression	p -1	SSR	SSR / (p - 1)
Residual	N - p	SSE	SSE / (N - p)
Total	N - 1	SST	

Table 2B.1 Analysis-of-Variance Table

The above quantities may be used to assess the accuracy of the fitted model by testing the null hypothesis that all model coefficients except β_0 are equal to zero. If the test shows this hypothesis to be untrue, then this allows acceptance of the alternative hypothesis; that at least one of the model coefficients is non-zero, in addition to the mean. In order to test the null hypothesis, the sum of squares information calculated above is

first used to compute an F–statistic, which is of the form:

$$F = \frac{\text{Mean Square Regression}}{\text{Mean Square Residual}} = \frac{SSR / (p - 1)}{SSE / (N - p)} \qquad (2B.4)$$

This value is then compared with standard tables, for the appropriate number of degrees of freedom, in order to assess whether the hypothesis is to be accepted or rejected at a given level of significance. An additional statistic which may be used to determine the quality of fit of the predictive model is the coefficient of determination, defined as

$$R^2 = \frac{SSR}{SST} \qquad (2B\ 5)$$

The coefficient of determination reflects the proportion of the total variation of the computed function values about their mean value which is being explained by the fitted model.

3. Two level factorial designs

3.1 Introduction

The general aim in the use of response surface techniques is to construct a mathematical representation of the measured data which establishes a connection between the values of the input variables and the level of the response functions of interest. This analytical model is then used to predict the function values at combinations of input variable levels other than those initially tested. The simplest such representation is a linear model, also known as a 'main effects' model, the general form of which is as follows.

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i \tag{3.1}$$

In this equation, $X_1, X_2, ..., X_n$ are the levels of the n input variables, Y is the value of a measured response function, and $\beta_0, \beta_1, ..., \beta_n$ are unknown coefficients which link the input variables and the resulting response. The aim of the current chapter is to present a number of strategies for selecting variable combinations at which to carry out computational trials, in order to estimate the values of the unknown parameters with the greatest efficiency, whilst maintaining an acceptable level of accuracy. In the following chapter the results of a number of experimental trials will be presented, in order to establish the suitability of these methods to the engine noise application.

3.2 A simple method of estimating main effects

Considering a simple one-variable example, the linear model is of the form

$$Y = \beta_0 + \beta_1 X_1 \tag{3.2}$$

in which there are two unknown parameters to be estimated. At least two tests must therefore be carried out in order to obtain sufficient function information to be able to distinguish between the effects of the two terms. Having decided the range of variable values over which predictions of the response function are required, an appropriate choice for the two values of the variable at which to carry out tests would seem to be the low and high bounds of this range. This gives the advantage that all predictions over the required range are obtained by interpolation.

If a further variable is added to the investigation, then the model becomes

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad (3.3)$$

in which three parameters are to be estimated. The minimum test requirement is now for three trials, with the further provision that each of the variables must appear at a minimum of two different levels, with a suitable choice of settings being the low and high bounds of each variable, as above. A further parameter is added to the model with each extra variable which is to be considered. For the general case of n variables, the minimum number of trials is thus n + 1, with each variable appearing at at least two distinct levels.

This requirement for test data may be met by carrying out a basic 'one at a time' test procedure, in which one trial is performed with all of the variables at, say, their low level, followed by n further trials, in each of which one of the variables is set to its high value, with all others at the low level. Even if no pure quadratic effects occur in the response, however, this method can still lead to inaccuracies, since it assumes that the effect of each variable will be independent of the value of all others. This may not be the case, as it is possible that the effect on the response function of a small change in the value of a variable A may be different when another variable B is at its high level to the effect when B is at its low level. The occurrence of this connection between variables is termed *interaction*, with the present example representing AB interaction. An interaction effect is represented in the predictive model by a cross term involving the interacting variables, so that (3.3) would become

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 \qquad (3.4)$$

If many such effects are present in the variation of measured response, or a small number occur with large magnitude, then there is substantial scope for error in the prediction of the response throughout the region of interest. In addition, the magnitude of this two-way interaction term may itself be different at different levels of some third variable C, leading to a three-way (or third order) interaction term ABC. In general, in fact, there exists the possibility of interactions of increasing complexity up to n^{th} order for a problem in n variables, although their magnitudes are likely to decrease rapidly with increasing order.

The limitations of the 'one at a time' approach may be demonstrated diagrammatically with reference to a three variable problem. Figure 3.1 shows the three-dimensional design variable space, together with four variable combinations which might be tested using such a strategy. The low and high levels of each variable are here represented as -1 and (+)1 respectively. The distribution of test points throughout the experimental region clearly leaves much to be desired, with the constructed model being based on information collected from only a small portion of the total space. Although predictions at points which fall on the edges of the

cube between two adjacent points are likely to be fairly reliable, function evaluation at all other locations requires a varying degree of extrapolation. Two-way interaction effects may lead to inaccuracies at, for example, the points (-1, 1, 1), (1, -1, 1) and (1, 1, -1), with the point (1, 1, 1) being further subject to the possibility of a three-way interaction effect.

An equally important defect of the 'one at a time' method is that, in addition to ignoring all interaction effects in the construction of a predictive model, it provides no means of assessing whether these effects are present or not. If the user is aware of these limitations, then the 'one at a time' approach can serve a useful purpose as an initial survey of a design variable space, in order to ascertain, for example, whether any of the selected input variables have a negligible effect on the response function of interest. More detailed investigation may then be posed in a reduced dimensional variable space, with a resultant saving in the cost of experimentation.

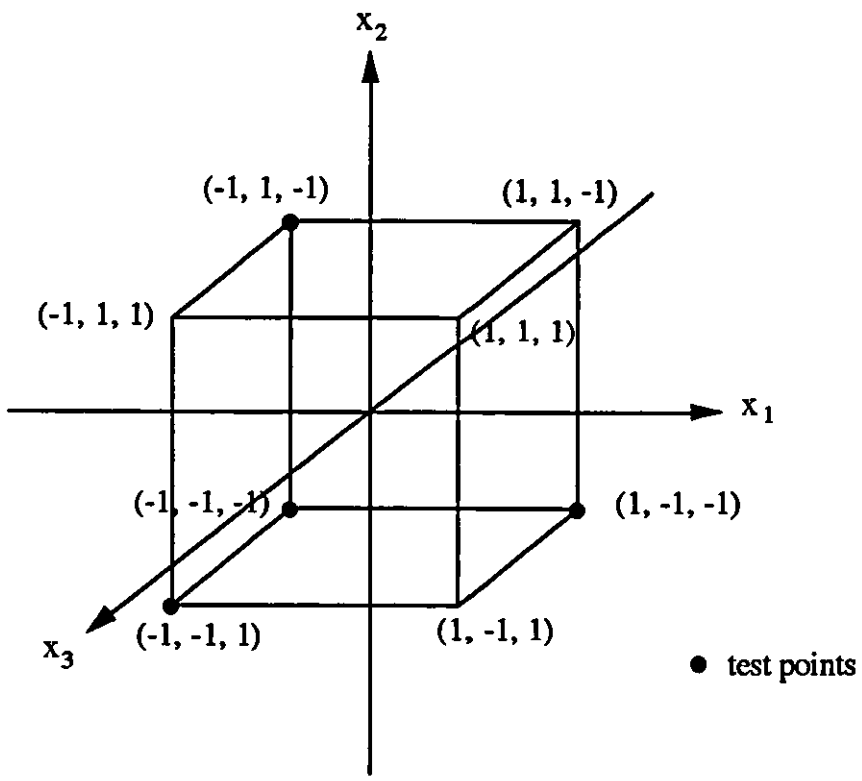

Figure 3.1. A 'one at a time' test in three dimensions

The adequacy of the model constructed using the 'one at a time' method may be assessed by carrying out further tests at one or more points within the experimental region, such as at (1, 1, 1) in Figure 3.1, and comparing the result with the value predicted by the model of equation (3.1). The magnitude of the difference between these two results provides some

indication of the degree to which the model is failing to represent the true response. An initial improvement to the constructed model may be obtained by improving the distribution of test points throughout the design variable space by, for example, testing at each of the vertices of the cube in Figure 3.1. Although this will reduce bias towards a particular area of the region, the representation of interaction effects is still precluded by the simple nature of the main effects model of (3.1). Additionally, eight tests are now being conducted to estimate just four parameters, and this ratio of model terms to test points will fall sharply if the number of design variables is increased. It is at this stage that a more formalised approach to the selection of an appropriate experimental design is required.

3.3 The factorial design

The factorial design is a widely used general purpose design for fitting a first-order model. It is valid for any number of input variables n, and has the further advantages that it is both orthogonal and rotatable. A factorial design in n variables is constructed by performing tests at all combinations of the high and low levels of each of the variables. The total number of such combinations is 2^n, and the test points may be visualised as lying at the vertices of an n–dimensional hypercube. The test points which are required for a three-dimensional case are shown in Figure 3.2.

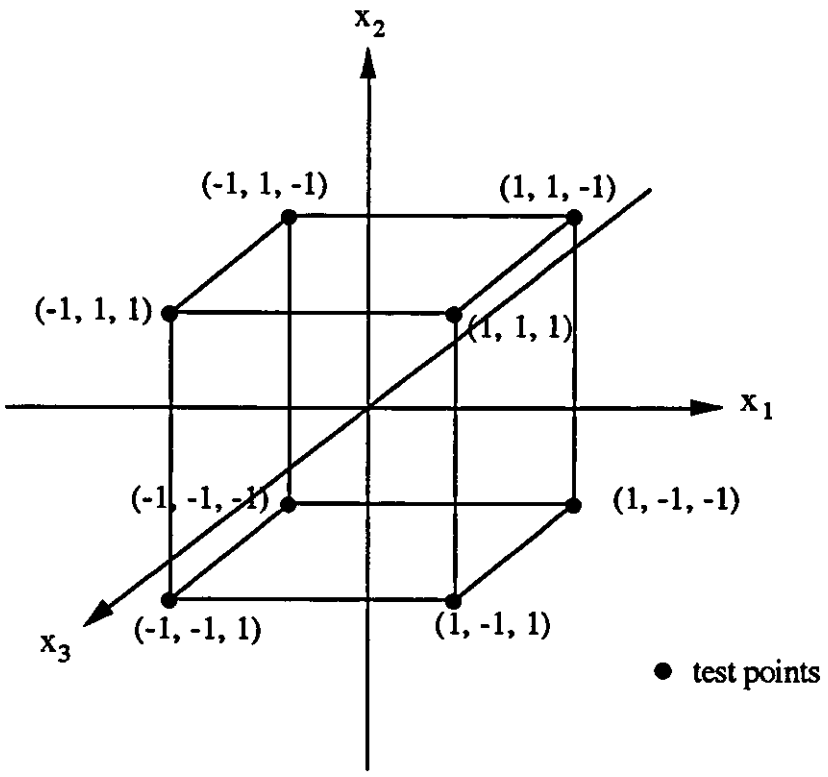

Figure 3.2. A two level factorial design in three dimensions

The variables are scaled in accordance with the convention described in Section 2.12, so that for each of the i = 1, ..., n variables in each of the j = 1,..., N tests, the value X_{ij} is converted to the standardised level x_{ij} by the transformation

$$x_{ij} = \frac{X_{ij} - \overline{X_i}}{\Delta_i} \tag{3.5}$$

The effect of this scaling is to normalise the values of the low and high bounds of each variable to -1 and + 1 respectively, as shown in Figure 3.2.

The design matrix for this three-variable design is then

$$\mathbf{D} = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \end{array}$$

This leads to the following moment matrix $\mathbf{M} = N^{-1}\,\mathbf{X}'\mathbf{X}$

$$\mathbf{M} = \begin{array}{c} \\ 1 \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

thus demonstrating that the design meets the conditions for both orthogonality and rotatability as outlined in Sections 2.7 and 2.8. All of the two level factorial designs introduced in the present chapter display both of these properties.

Since 2^n tests have been performed, it is possible to fit a mathematical model which has up to 2^n unknown parameters to be estimated. From Section 3.1, it is known that $n + 1$ of these terms are the mean plus main effect terms. The remaining $2^n - (n + 1)$ terms are then the interactions between the n variables, and the factorial design thus allows the construction of a 'linear + interactions' model of the form shown in equation (3.6).

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \beta_{ijk} x_i x_j x_k + \sum_{i=1}^{n-3} \sum_{j=i+1}^{n-2} \sum_{k=j+1}^{n-1} \sum_{l=k+1}^{n} \beta_{ijkl} x_i x_j x_k x_l + \cdots\cdots \beta_{123\ldots n}\, x_1 x_2 x_3 \cdots x_n \qquad (3.6)$$

It should be noted that only linear interactions appear in (3.6); that is, terms in which no variable has an index greater than 1. The two-way interactions may alternatively be considered as mixed quadratic terms, and the three-way interactions as mixed cubics, etc..

The advantages of the full factorial design are that it is easy to specify, and allows for the estimation of all possible linear interaction terms. This latter characteristic may be particularly useful when investigating a specific problem of an unfamiliar nature, in which the susceptibility to interaction effects is unknown. The main disadvantage, however, is that the number of terms in the model of equation (3.6) successively doubles as the number of variables is increased, as does the number of tests which must be performed in order to estimate these parameters. This test requirement is easily met for problems containing up to, say, 4 or 5 variables ($2^n = 16$ and 32 tests respectively), but quickly becomes excessive when the number of design variables exceeds 6 or 7. An investigation into the effect of 10 variables would require 1024 trials to be conducted, whilst with 15 variables this number increases to nearly 33,000.

It is unlikely, however, that all of these terms will have a significant effect on the predictive ability of the fitted model, and in general the magnitude of coefficients is likely to decline as the order of interaction increases. The proportion of significant terms in the model may thus fall extremely rapidly as the number of variables increases. If in a twelve variable study, for example, only effects up to and including fourth-order interactions were contributing to the model, then the number of significant terms would be 794 out of a total of $2^{12} = 4096$, or just over 19%. If the fourth-order interactions were negligible, this figure would drop to 299 terms (7.3%), and a linear plus two-way interactions model would contain just 79 parameters; less than 2% of those appearing in the original model. A similar model in 15 variables would contain less than 0.4% significant terms, and even for 7 variables the ratio is under 23%.

In such cases the full factorial design is an inefficient means of acquiring the necessary test data, and an alternative strategy is required in which the number of tests to be performed may be reduced in line with the expected number of significant terms.

In practice a full factorial design would only be considered if one or more of the following conditions were true:

i) The number of variables is small (say < 7)
ii) The cost of each test is very low
iii) All linear interaction terms are known to contribute substantially to the predictive accuracy of the model.

3.4 Economic first-order designs

In reducing the number of experimental test runs which needs to be carried out in order to estimate the required parameters of the mathematical model, a design is sought which will provide the necessary accuracy, and which will also maintain, if possible, the desirable properties of orthogonality and rotatability possessed by the full factorial design. Two

orthogonal designs for fitting first-order models which may be employed are the Simplex Design (Box, 1952) and the Plackett-Burman design (Plackett and Burman, 1946).

A simplex design in n variables uses n + 1 test points, positioned at the vertices of an n–dimensional regular figure (a simplex) in order to asses the mean value and main effects represented by the linear model of equation (3.1). The disadvantage of this design, however, is that the extremely low number of test points does not allow for the estimation of any of the interaction effects which may occur between variables.

The Plackett-Burman design is also largely aimed at fitting a main effects model, although some interaction terms may be included. Its main drawback as a general technique, however, is that it is not available for every value of n, the number of design variables.

3.5 Fractional factorial designs

One class of designs which does meet the above criteria is the family of fractional factorial designs, in which a 2^{-m} fraction ($1 \leq m \leq n-1$) of the complete factorial design in n variables, as outlined in Section 3.3, is used to collect the necessary function information. The fractional factorial design is valid for all values of $n > 1$, and has the additional advantage that the value of m can be selected to successively halve the number of tests to be carried out. The price to be paid for this reduction in test data, however, is a corresponding reduction in the number of terms of equation (3.6) which can be estimated. It is clearly important that the terms which remain in a reduced model are those which contribute substantially to the accuracy of prediction, and this is determined by the manner in which the original 2^n tests are subdivided to produce the reduced experimental design. Before proceeding further, a brief discussion of notation is required.

3.5.1 Notation for fractional factorial designs

Although the elements of a design matrix and points within a design variable space are most conveniently described using the ±1 notation used in the previous sections, an alternative method of specifying design variable test combinations is useful in outlining the construction of a fractional factorial design.

Under this convention, used by many authors, e.g. Davies (1954), each of the variables $X_1, X_2, \ldots, X_n$ is denoted by a capital letter (A, B, C,... etc.). Each treatment combination is then described by a string of k lowercase letters ($k \leq n$), in which the letter relating to a particular variable appears in the string if that variable is to be tested at its high level, but is absent if the low level is to be tested. Thus a test in five variables at the point (1, -1, -1, 1, 1), in

which variables A, D and E are tested at their high level, with B and C at the low level, would be notated as 'ade'. Special provision is made for the combination in which all variables are tested at their low level, and this is written as '(1)'.

Additionally, terms in the mathematical model are described using the appropriate capital letters in place of the $X_1, X_2, ..., X_n$ of equation (3.5). Thus the main effect terms are written as A, B, C, etc., and two-way interactions as AB, AC, BE, DF, etc.. In general, each interaction of order k ($k \leq n$) will be notated as a string of k capital letters denoting the variables between which the interaction effect occurs.

3.5.2 Construction of designs with m = 1

A method of specifying a 2^{-m} fraction of a 2^n factorial, following that of Khuri and Cornell (1988), is outlined in the following sections, and is valid for all $n > 1$. In order to aid description of this method, the simplest case of $m = 1$ is first discussed. Following this, the procedure is extended to designs employing a 2^{-2} fraction, and then to the general methodology for any value of m in the range $1 \leq m \leq n-1$.

For a design employing a 2^{-1} fraction of a full factorial design, the original 2^n tests must be divided into two fractions. The first step is to nominate one p-way interaction ($p \leq n$) which is thought to have a negligible influence on the model, and whose effect may be 'sacrificed'. This interaction should normally be of a high order, and, since only one is required, a logical choice is the single n-way interaction.

Of the total of 2^n test combinations, notated by a string of lowercase letters, as in the previous section, half will have an even number of letters in common with this nominated interaction, whilst half will have an odd number in common. This is the basis for dividing the original tests into two fractions. The choice as to which of these two fractions to use as the design is entirely arbitrary, as it cannot be known prior to testing whether one will yield better results than the other. In practice it is usual to use the 'principal fraction', which is defined as the fraction which has an odd number of letters in common with the nominated interaction if the nominated interaction contains an odd number of letters, and an even number in common with it if it is even.

It will be convenient, especially when considering designs for which $m > 1$, to denote the fraction whose tests have an even number of letters in common with a nominated interaction containing an even number of letters, or an odd number of letters in common with one containing an odd number of letters, by the symbol '+'. Conversely, the fraction whose tests have an even number of letters in common with a nominated interaction containing an odd number of letters, or an odd number of letters in common with one containing an even number

of letters, is denoted by the symbol '-'. Clearly, each of the tests within a given fraction will display the characteristics of the fraction as a whole, and may also be given the appropriate symbol.

As an example, consider the selection of a 2^{-1} fraction of a five variable factorial design containing $2^5 = 32$ tests. The full set of test combinations is as follows.

(1)	d	ad	be	abc	ace	bde	abde
a	e	ae	cd	abd	ade	cde	acde
b	ab	bc	ce	abe	bcd	abcd	bcde
c	ac	bd	de	acd	bce	abce	abcde

Table 3.1. Full set of 2^n tests for a five variable problem

If the five-way interaction ABCDE is selected as the term to be 'sacrificed', then the tests which have an even number of letters in common with this term are

(1)	ac	ae	bd	cd	de	abce	acde
ab	ad	bc	be	ce	abcd	abde	bcde

Table 3.2. The first of two 2^{n-1} fractions for a five variable problem

whilst the following have an odd number of letters in common with ABCDE

a	c	e	abd	acd	ade	bce	cde
b	d	abc	abe	ace	bcd	bde	abcde

Table 3.3. The second of two 2^{n-1} fractions for a five variable problem

Since ABCDE itself has an odd number of letters, it is this second fraction which, under the convention given above, is the principal fraction, denoted by '+'.

Having decided which tests to perform, it is now necessary to determine which of the original 2^n terms may be retained in the reduced model. Twice as many terms were present in the original model as there are now test results, with the result that the effect of pairs of terms will be indistinguishable from each other. Such terms are said to be 'confounded'. Any fitted model must include only one term of each pair, although the parameter estimate will in fact

represent the sum of the two effects. This may clearly lead to significant error if both of the terms have a substantial influence on the function to be approximated. If, however, the terms are paired such that only one of each set is significant, then little information will be lost. Since the reduced model contains the same number of coefficients as there are tests in the fractional design, an exact fit at all test points will always result, whichever term of each confounded pair is used. The effect of specifying an insignificant term of a pair will only become apparent when the model is used to predict the response at points other than those used to construct the model.

The effect of using the nominated interaction, in the above case ABCDE, to divide the factorial design into two fractions is that in all the tests of one fraction ABCDE is at its high level, whilst in the other fraction all tests have ABCDE at the low level. The result of this is that, whichever fraction is used, the ABCDE effect is only tested at one level, and is hence indistinguishable from the mean value, or identity effect, denoted as I_e. This confounding with the identity effect may be written as

$$ABCDE = I_e \tag{3.7}$$

and is termed the 'defining relation' for this particular design. This relation provides the basis for determining the confounding between all other terms in the full model. To find which other effect a particular term is confounded with, the term must be multiplied with the nominated interaction; the square of any letter being removed from the resulting character string. The term produced by this process is then the one with which the original term is confounded.

Returning to the above example, the term with which the main effect A is confounded is:

$$A \times ABCDE = A^2BCDE = BCDE \tag{3.8}$$

whilst the two-way interaction term CE is confounded with

$$CE \times ABCDE = ABC^2DE^2 = ABD \tag{3.9}$$

The full set of confounded parameters for the above example is then as given in Table 3.4, below. Each term in the first column is confounded with the term appearing on the same row of column two.

mean	ABCDE
A	BCDE
B	ACDE
C	ABDE
D	ABCE
E	ABCD
AB	CDE
AC	BDE
AD	BCE
AE	BCD
BC	ADE
BD	ACE
BE	ACD
CD	ABE
CE	ABD
DE	ABC

Table 3.4. Pattern of confounding for a five variable problem

This design appears to provide a promising basis for accurately representing the true response, since all of the main effect and two-way interaction terms may be estimated independently of each other. The fitted model would contain those terms which appear in the left-hand column. The predictive ability of the constructed model will depend on the size of the three-way and four-way interaction coefficients with respect to the two-way interaction and main effect terms respectively.

3.5.3 Construction of designs with m = 2

Designs which use a 2^{-2} fraction of the full factorial require that the number of tests be reduced to one quarter of the original number. In order to achieve this, two interactions must be nominated, to successively halve the number of tests. The choice of suitable nominated interactions, however, is now not so apparent. For the five variable example introduced above, an initial selection might be the five-way interaction ABCDE and one of the four-way interactions, say ABCD.

As for the previous example, half of the 2^n tests will have an even number of letters in

common with the first of these nominated interactions, whilst half will have an odd number in common. The same is true for the second interaction. Since the original parameter estimates of the nominated interactions were orthogonal, the division of tests will also be orthogonal, so that each half fraction created by either interaction will be further halved by the other. If the symbols '+' and '–' are again adopted, then the complete set of 2^n tests will be divided into four fractions, as follows.

Interaction 1	Interaction 2
+	+
+	-
-	+
-	-

The principal fraction is now defined as that fraction for which '+' symbols are obtained with respect to both defining relations.

For the five variable example considered earlier, the principal fraction for the 2^{-2} design with defining relation ABCDE and ABCD may be found by further dividing the tests of Table 3.3, and selecting those which have an even number of letters in common with ABCD. These are listed in Table 3.5.

e	abe	ace	ade	bce	bde	cde	abcde

Table 3.5. Principal fraction for 2^{n-2} design in five variables

Since both of the nominated interactions are now confounded with the mean effect, there are two defining relations of the form

$$ABCDE = ABCD = I_e \tag{3.10}$$

By multiplying each term of the full model by each of the two interactions, two terms are generated with which the original effect is confounded. Since only one quarter of the terms of the full model may now be represented, however, confounding must occur in sets of four parameters. A third defining relation is therefore needed in order to produce the fourth member of each set, and this is obtained by multiplying together the two nominated interactions already defined, to form what may be termed a *generalised interaction* between the two. For the present example, this would produce the term

$$ABCDE \times ABCD = A^2B^2C^2D^2E = E \tag{3.11}$$

so that the defining relations are now

$$ABCDE = ABCD = E = I_e \tag{3.12}$$

The pattern of confounding is thus as shown in Table 3.6. Each term in a given column is confounded with each of the three terms appearing on the same row of the table.

mean	ABCDE	ABCD	E
A	BCDE	BCD	AE
B	ACDE	ACD	BE
C	ABDE	ABD	CE
D	ABCE	ABC	DE
AB	CDE	CD	ABE
AC	BDE	BD	ACE
AD	BCE	BC	ADE

Table 3.6. Table of confounding for a five variable problem

This is clearly a very poor design, as the mean value is confounded with one of the main effects, each of the other main effects are confounded with one two-way interaction, and the six remaining two-way interactions are confounded in pairs. On examination of the above process it becomes apparent that if significant parameters are to be confounded only with higher order interactions, and not with each other, the number of letters in each of the defining relations must be kept as large as possible. More specifically, the worst case of confounding will be determined by the shortest of these relations, and so the original interactions should be nominated in such a way as to maximise the length of the shortest nominated or generalised interaction. In forming the third defining relation, it is clear from equation (3.12) that it will consist only of letters which occur in just one of the original terms. These terms must therefore be selected in order to contain as high a ratio as possible of unique letters to common letters.

If, for example, the nominated interactions were ABCD and BCDE, then the third term, formed by multiplication, would be AE, giving a minimum length of just two letters. An alternative choice of ABCD and CDE, however, would yield ABE, and a minimum length of three letters. This second choice would then result in the following confounding table.

mean	ABCD	CDE	ABE
A	BCD	ACDE	BE
B	ACD	BCDE	AE
C	ABD	DE	ABCE
D	ABC	CE	ABDE
E	ABCDE	CD	AB
AC	BD	ADE	BCE
AD	BC	ACE	BDE

Table 3.7. Revised table of confounding for a five variable problem

This is certainly an improvement over Table 3.6, in that the mean value is now only confounded with interactions of order three and four. It will still lead to substantial errors in the fitted model, however, if the two-way interaction terms contribute significantly, since confounding occurs between each main effect and a two-way interaction, and in pairs between the remaining two-way interaction terms. This, however, is the best design which can be obtained when using a quarter fraction of a five variable factorial design. This is as may be expected when it is borne in mind that the total number of the mean + main effect + two-way interaction terms is 16, whereas the number of test points is just 8.

3.5.4 Design resolution

In the previous section it was seen that a design which can distinguish between the mean value and each of the main effects is more desirable than one in which confounding occurs between a number of these terms. This ability to distinguish between the various parameters of the model is termed the *resolution* of a design, with a full factorial being of *full* resolution since it can independently estimate each of the 2^n terms of the complete linear + interactions model.

The resolution of a fractional factorial design is defined as the number of letters in the shortest term which appears in the defining relations, with a higher value indicating a better design. The value of the design resolution is usually written in Roman numerals as, for example, resolution VII or resolution II. The upper limit on such a value is clearly n, the number of design variables, since this is the maximum number of letters available. The 2^{-1} design with defining relation ABCDE = I_e which produced the confounding pattern of Table 3.4 was thus of resolution V. In contrast, the first attempt at constructing a 2^{-2} design with defining relations ABCDE = ABCD = E = I_e was only of resolution I, with the improved version using ABCD = CDE = ABE = I_e being of resolution III.

In practice, attention is often focused on the accurate estimation of mean, main effect and two-way interaction terms, and as a result the following three design resolutions are of particular importance.

Resolution III. In this design the mean is only confounded with interactions of order three or higher. Main effects are distinct from each other but confounded with one or more two–way interactions, plus higher terms. Two-way interaction terms may also be confounded with each other.

Resolution IV. Main effects are now confounded only with third-order interactions and higher, although two-way interactions are still confounded among themselves. The mean is independent of all interaction terms of order less than four.

Resolution V. The mean value is confounded only with interaction terms of order five and above. Each of the main effects is confounded with interactions of order four and above. Two–way interactions are now independent of each other, but are confounded with third-order interactions.

3.5.5 Construction of general fractional factorial designs

Having introduced the special cases of fractional factorial designs involving 2^{-1} and 2^{-2} fractions of the full 2^n factorial, the general method of constructing a 2^{-m} fraction may be described as follows.

1. Nominate m interactions which are to be confounded with the mean effect.

2. For each of the m nominated interactions, those tests which have an even number of letters in common with a nominated interaction containing an even number of letters, or an odd number of letters in common with one containing an odd number of letters, are given the symbol '+'. Conversely, those tests which have an even number of letters in common with a nominated interaction containing an odd number of letters, or an odd number of letters in common with one containing an even number of letters, are given the symbol '-'.

3. The principal fraction, which constitutes the required design, consists of all those tests which have been assigned a '+' sign with respect to each of the m nominated interactions.

4. The m nominated interactions are then multiplied together in all possible combinations of between 2 and m terms, to give an additional $2^m - m - 1$ generalised

interactions. Together with the original m interactions, these form the $2^m - 1$ parameters which are confounded with the mean effect, producing $2^m - 1$ defining relations. The nominated interactions should be selected in such a way as to maximise the number of letters in the shortest of these defining relations.

5. The table of confounding is produced by multiplying each term of the original model by each of the $2^m - 1$ defining relations. This then groups the 2^n terms into 2^{n-m} sets of 2^m terms, and the new model is constructed by selecting that term from each of the 2^{n-m} sets which is expected to make the most significant contribution to the performance of the model. This will usually be the term which consists of the fewest letters, although if two or more terms are of equal length it is often not possible to make an informed choice. In such cases it may prove useful to assess a number of possible models.

As an example, consider the construction of a 2^{-3} fraction of a 2^6 factorial array, with six variables A to F. If the three nominated interactions are the three-way interaction terms ABC, CDE and AEF, then the $2^n = 64$ tests are divided into eight fractions as shown in Table 3.8.

	ABC	CDE	AEF		ABC	CDE	AEF
(1)	-	-	-	bcd	-	-	-
a	+	-	+	bce	-	-	+
b	+	-	-	bcf	-	+	+
c	+	+	-	bde	+	-	+
d	-	+	-	bdf	+	+	+
e	-	+	+	bef	+	+	-
f	-	-	+	cde	+	+	+
ab	-	-	+	cdf	+	-	+
ac	-	+	+	cef	+	-	-
ad	+	+	+	def	-	-	-
ae	+	+	-	abcd	+	-	+
af	+	-	-	abce	+	-	-
bc	-	+	-	abcf	+	+	-
bd	+	+	-	abde	-	-	-
be	+	+	+	abdf	-	+	-
bf	+	-	+	abef	-	+	+
cd	+	-	-	acde	-	+	-
ce	+	-	+	acdf	-	-	-
cf	+	+	+	acef	-	-	+
de	-	-	+	adef	+	-	+
df	-	+	+	bcde	-	+	+
ef	-	+	-	bcdf	-	-	+
abc	+	+	+	bcef	-	-	-
abd	-	+	+	bdef	+	-	-
abe	-	+	-	cdef	+	+	-
abf	-	-	-	abcde	+	+	-
acd	-	-	+	abcdf	+	-	-
ace	-	-	-	abcef	+	-	+
acf	-	+	-	abdef	-	-	+
ade	+	-	-	acdef	-	+	+
adf	+	+	-	bcdef	-	+	-
aef	+	+	+	abcdef	+	+	+

Table 3.8. Division of 2^6 tests into 2^3 fractions

The principal fraction, containing those tests of Table 3.8 which are accompanied by three '+' signs, is then as follows.

ad be cf abc aef bdf cde abcdef

Table 3.9. Principal fraction for a 2^{6-3} design

The generalised interactions are then formed by multiplying the three nominated interactions together in all combinations, to give

$$ABC \times CDE = ABDE$$
$$ABC \times AEF = BCEF$$
$$CDE \times AEF = ACDF$$
$$ABC \times CDE \times AEF = ABDE \times AEF = BDF$$

so that the defining relations are

$$ABC = CDE = AEF = ABDE = BCEF = ACDF = BDF = I_e$$

Since the shortest of these terms consists of three letters, the design is of resolution III, in which the mean value and each of the main effects are distinguishable from each other, but in which the main effects are confounded with one or more two-way interactions. The pattern of confounding for this design is as shown in Table 3.10, in which each entry in a particular column is confounded with all other terms appearing on the same line of the table. A suitable model might then be composed of each of the terms in the first column of this table, although substitution of either BE or CF for the term AD may prove to be more appropriate.

mean	ABC	CDE	AEF	ABDE	BCEF	ACDF	BDF
A	BC	ACDE	EF	BDE	ABCEF	CDF	ABDF
B	AC	BCDE	ABEF	ADE	CEF	ABCDF	DF
C	AB	DE	ACEF	ABCDE	BEF	ADF	BCDF
D	ABCD	CE	ADEF	ABE	BCDEF	ACF	BF
E	ABCE	CD	AF	ABD	BCF	ACDEF	BDEF
F	ABCF	CDEF	AE	ABDEF	BCE	ACD	BD
AD	BCD	ACE	DEF	BE	ABCDEF	CF	ABF

Table 3.10. Pattern of confounding for a 2^{6-3} design

3.5.6 Selection of an appropriate fractional factorial design

As has been demonstrated in the previous sections, the size of a fractional factorial design may be selected in order to give a particular design resolution, enabling the fitted model to distinguish between certain of the possible linear + interaction terms. What is not yet clear, however, is which of the possible values of m is appropriate for specifying a design of a desired resolution in a given number of variables. The maximum resolution which can be obtained for a particular value of m depends on the successful selection of m interactions which, when multiplied together, will yield the maximum number of letters in the shortest of the defining relations. This process of selecting interactions becomes increasingly difficult as the values of both m and n increase, and is largely a matter of experimentation. Once a satisfactory nomination of interactions has been made for a given combination of m and n, however, these choices will be applicable to all future designs having the same values of m and n. It is therefore convenient to derive suitable interactions for all commonly used combinations of m and n, and to use the known values of design resolution for each as the basis for selection of an appropriate design when undertaking a particular investigation. The maximum design resolution which can be obtained for an n variable problem using a 2^{-m} fraction containing N tests is listed in Table 3.11 for $1 \leq n \leq 6$, and in Table 3.12 for $7 \leq n \leq 12$, together with an appropriate choice of nominated interactions. Each of these combinations of nominated interactions has been individually derived, although in many cases the choice of interactions which gives the maximum resolution is not unique, and equivalent combinations may be found in the published literature.

n	m	N	nominated interactions			resolution
1	0	2				FULL
2	0	4				FULL
2	1	2	AB			II
3	0	8				FULL
3	1	4	ABC			III
3	2	2	AB	BC		II
4	0	16				FULL
4	1	8	ABCD			IV
5	0	32				FULL
5	1	16	ABCDE			V
5	2	8	ABC	CDE		III
6	0	64				FULL
6	1	32	ABCDEF			VI
6	2	16	ABCD	CDEF		IV
6	3	8	ABC	CDE	AEF	III

Table 3.11. Fractional factorial designs for $1 \leq n \leq 6$

n	m	N	nominated interactions				resolution
7	0	128					FULL
7	1	64	ABCDEFG				VII
7	2	32	ABCD	CDEF			IV
7	3	16	ABCD	CDEF	ACFG		IV
7	4	8	ABD	BCE	ACF	ABCG	III
8	0	256					FULL
8	1	128	ABCDEFGH				VIII
8	2	64	ABCDE	DEFGH			V
8	3	32	ABCD	CDEF	EFGH		IV
8	4	16	ABCD	CDEF	EFGH	ACEH	IV
9	0	512					FULL
9	1	256	ABCDEFGHI				IX
9	2	128	ABCDEF	DEFGHI			VI
9	3	64	ABCD	CDEF	EFGH		IV
9	4	32	ABCD	CDEF	EFGH	ACHI	IV
10	0	1024					FULL
10	1	512	ABCDEFGHIJ				X
10	2	256	ABCDEF	EFGHIJ			VI
10	3	128	ABCDE	DEFGH	AGHIJ		V
10	4	64	ABCD	CDEF	EFGH	GHIJ	IV
11	0	2048					FULL
11	1	1024	ABCDEFGHIJK				XI
11	2	512	ABCDEFG	EFGHIJK			VII
11	3	256	ABCDE	DEFGH	GHIJK		V
11	4	128	BCDEH	ABCEGI	ABDEFJ	AEFGK	V
12	0	4096					FULL
12	1	2048	ABCDEFGHIJKL				XII
12	2	1024	ABCDEFGH	EFGHIJKL			VIII
12	3	512	ABCDEF	DEFGHI	GHIJKL		VI
12	4	256	ABCDE	DEFGH	GHIJK	ADGKL	V

Table 3.12. Fractional factorial designs for $7 \leq n \leq 12$

The above tables show that the advantages of using a fractional factorial design rather than a full factorial grow substantially as the number of variables is increased. If, for example, a design of resolution V is required, then any study in less than five variables will still require that a full factorial be used. With 5, 6, or 7 variables, a half fraction can be used, with the necessary fraction diminishing to 1/4 for 8 or 9 variables, 1/8 for 10 variables, and 1/16 for investigations involving 11 or 12 factors. Even a 1/16 fraction of a 2^{12} design still requires 256 tests to be carried out, and it is perhaps at this point that a preliminary investigation would be useful in order to ascertain whether all of the variables are indeed having a significant effect on the variation of the function. A main effects model using a design of, say, resolution III would be useful in such a case. Alternatively, if the number of potential variables is extremely large, e.g. 100+, then a specific strategy may be required for screening the initial input in order to

identify those variables which are worthy of further investigation. A number of such strategies have been advanced, (see for example Welch et al. (1989) and Watson (1961)), but further investigation has not been undertaken within the present work. Attention has been focused on designs suitable for studies in up to, say, 12 variables.

A further characteristic of designs involving 7 to 12 variables is that the cost of obtaining a high resolution design becomes prohibitively large as n increases, whilst it is probable that only a fairly small fraction of the terms in the model will be of significance. As a result, an initial investigation in this number of factors would rarely be performed using a full factorial design or a fractional factorial with resolution greater than, perhaps, resolution V. Only if such a test had shown there to be significant underspecification in the model due to missing higher order interaction terms would larger designs be considered. The nature of the fractional factorial design then allows the data already collected to be incorporated into the new design, thus minimising the number of additional tests which need to be carried out.

Tables 3.11 and 3.12 provide information as to which fractional factorial design to choose in order to obtain a particular resolution for the number of design variables under investigation. In order to specify a particular design, however, it is first necessary to determine the level of resolution which is appropriate to the particular nature of the application to be investigated. This is equivalent to establishing the degree of interaction, or possible interaction, which is expected to occur within the model. Since this is entirely dependent on the characteristics of the specific process which is to be studied, no general recommendation is likely to be entirely satisfactory in every situation. When investigating an unfamiliar application in which the user has no prior knowledge as to the relative significance of interaction effects, two alternative approaches are possible, as follows.

i) Carry out an initial test on a representative study within the general category of the problem to be addressed using a high resolution design, with the aim of assessing the magnitude of high order interaction terms. This test may be conducted using a relatively small number of design variables in order to minimise the number of tests which must be performed.

ii) Use a design of a lower resolution to investigate the particular problem under consideration on the assumption that certain of the interaction terms will be negligible. The validity of this assumption may be tested by carrying out additional trials within the design variable space in order to assess the ability of the fitted model accurately to predict the true response.

If a number of investigations are to be carried out in related fields, then the first of these methods is perhaps the most useful, since it provides the investigator with a more comprehensive view of the nature of the problem being addressed, which may be of substantial

advantage in analysing the results of future studies. If only a single investigation is to be conducted, then the quantity of additional information gained by this approach may not justify the extra testing requirement.

When carrying out noise optimization studies on internal combustion engines, the individual engine block or engine block/sump systems of different installations share a large number of features in common. Additionally, the categories of variable which may be used in an analysis/optimization study will also be similar. It is thus reasonable to expect that this similarity will extend to the general type of mathematical model which is needed in order to represent the noise function adequately. This being the case, it is probable that detailed studies of a particular engine system will lead to significant insights into the ways in which design variables interact, and influence the noise function, which will be broadly applicable to a wide range of potential applications. Such a series of studies for the finite element model of Appendix 1C is presented in the next chapter.

4. Experiments using first-order designs

In this chapter, the results of a number of experimental studies using first-order designs are presented. These numerical trials have been conducted in order to assess the suitability of the response surface models of the previous chapter to the computer prediction of radiated engine noise, and are carried out using an FE model of a representative engine structure. A description of the four-cylinder engine block model used, and of the seven design variables which are used in the present investigation, may be found in Appendix 1C.

4.1 Full factorial design in seven variables

The first design to be investigated is the full factorial design of Section 3.3. For a problem involving seven design variables, $2^7 = 128$ tests are carried out. This design is used to estimate the parameters of the full linear + interactions model of equation (3.6), in order to assess whether radiated noise can be adequately represented by a linear model. Mathematical models of both the noise function and the structural mass of the engine are obtained using the surface fitting techniques outlined in Chapter 2. The full factorial is an orthogonal design, in which no covariance effects occur between any model terms, and the variance value obtained for each term is identical. A value of 7.8125×10^{-3} was obtained for each.

4.1.1 The mass function

In order to provide an introduction to the use of response surface techniques, the analysis of the mass function is first presented. This is a very simple function, since for each variable the quantity being modified is the thickness of the relevant component part of the engine block. The mass of the variable is directly proportional to this thickness, and the effect of variations in the thickness of one variable will not be affected by the thickness of any other component, so that no interaction effects occur. It is expected that the variation of mass throughout the design variable space can be accurately described using just the main effects model of equation (3.1).

The 20 largest coefficients of the mass function model are listed in Table 4.1. The coefficients are arranged in order of magnitude and are normalised for variable values which are scaled to lie in the range ± 1.

1.	MEAN	72.104	11.	BCEFG	9.3749×10^{-10}
2.	F	4.8496	12.	BEG	7.8126×10^{-10}
3.	D	3.2267	13.	BEFG	6.2502×10^{-10}
4.	A	3.1984	14.	BF	6.2500×10^{-10}
5.	B	3.1901	15.	EG	6.2498×10^{-10}
6.	G	3.1151	16.	DEG	4.6876×10^{-10}
7.	C	1.9894	17.	BDF	4.6875×10^{-10}
8.	E	1.6699	18.	BDEFG	-4.6875×10^{-10}
9.	BC	-1.3829×10^{-5}	19.	EFG	4.6875×10^{-10}
10.	CEFG	-1.4063×10^{-9}	20.	BDEG	3.1251×10^{-10}

Table 4.1 Mass coefficients for full factorial test (n=7)

It is clear from this table that the model is dominated by the largest 8 coefficients, which are the mean and main effect terms. The remaining 120 terms in the model are very close to zero, with the small estimates obtained being due to round-up errors within the solution algorithm. This lack of interaction terms within the model is as expected, due to the nature of the particular function and variables involved.

In order to confirm that the full 128-term model is accurately representing the mass function throughout the whole of the design variable space, the lack of fit between the predictive model and the experimental response may be calculated at a number of points within the region of interest. Two categories of test point have been used to test this lack of fit, as follows.

i) Design points. The lack of fit is calculated at each of the 2^n test points which were used to construct the analytical model. This indicates the ability of the chosen parameters to account for the variations in the function value which occur at the points specified by the full factorial design. Since the design contains the same number of tests as there are coefficients in the full 128 term model (a saturated design), it is expected that the model will reproduce the original data exactly, to machine precision, with no lack of fit at any of these points.

ii) Model under-specification due to the omission of higher order terms, such as, for example, quadratic or cubic terms, can best be investigated by calculating the lack of fit which occurs at points in the design space which are distant from those used to construct the model. $2n + 1$ extra tests have been conducted in order to test this potential source of lack of fit. One of these lies at the centre of the design variable space $(0, 0, \ldots, 0)$, with the other $2n$ points lying at the centre of each of the $n - 1$ dimensional hypercubes formed by successively setting one of the variables to either its high or low bound. These $2n$ tests are located at the points $(\pm1, 0, \ldots, 0)$, $(0, \pm1, \ldots, 0)$, ... , $(0, 0, \ldots, \pm1)$. They lie at the vertices of a cross-polytope, or 'star', and are often termed 'star points'. For the seven variable problem currently being

investigated, the number of extra tests performed is thus 15. The exact specification of these points, in the original variable units, is given in Appendix 4A.

The lack of fit which occurs in each of these two categories of test point is summarised in Table 4.2.

Category	No. of tests	Maximum lack of fit kg	Maximum lack of fit E_R	Average lack of fit kg	Average lack of fit E_R
i) Design points	128	4.1211×10^{-13}	9.7018×10^{-13}	1.1102×10^{-13}	2.6136×10^{-13}
ii) Higher order	15	4.0534×10^{-5}	9.5422×10^{-5}	2.9342×10^{-5}	6.9074×10^{-5}
Average over all 143 tests				3.0778×10^{-6}	7.2455×10^{-6}

Table 4.2 Summary of lack of fit calculations for 128 coefficient mass model using full factorial design with 128 tests (n=7)

As expected, the lack of fit at any of the design points is approximately zero, due to the saturated nature of the design. Calculation of lack of fit at the additional 15 points also shows a negligible error in mass prediction, both in terms of the maximum and the average lack of fit at these points. The increase in error over that at the design points is due to the small non-zero estimates for the insignificant model terms, which, although they exactly cancel at the design points, do not necessarily do so at other points within the design variable space. These results confirm that the full model gives an almost exact prediction of mass variation throughout the region of interest, and since only the mean and main effect terms contribute substantially to this, they further suggest that exact prediction may also be attained using just the main effects model of equation (3.1). If mass were the only function to be assessed, then this observation would allow the use of a very low resolution fractional factorial design (resolution III), or even the 'one at a time' approach of Section 3.2. Since the variation in noise of the system is also required, however, with the two functions being calculated concurrently, no such saving in experimental effort would be possible unless the noise function were also representable using just a main effects model. As will be seen below, this is unlikely to occur in practice.

Due to its simple nature, the mathematical representation of the mass function need not be analysed further.

4.1.2 The noise function

For the noise function model, the full set of parameter estimates obtained is listed in Table 4B.1 of Appendix 4B. The coefficients are again arranged in rank order and normalised for variable values which are scaled to lie in the range ±1.

The lack of fit which occurs between the experimental noise level and the value predicted by this model, for each of the two categories of test point described in the previous section, is summarised in Table 4.3.

Category	No. of tests	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
i) Design points	128	4.8317×10^{-13}	6.4575×10^{-12}	1.2512×10^{-13}	1.6722×10^{-12}
ii) Higher order	15	1.5486	20.697	1.1890	15.891
Average over all 143 tests				0.12472	0.16669

Table 4.3 Summary of lack of fit calculations for 128 coefficient noise model using full factorial design with 128 tests (n=7)

This again shows that the saturated nature of the design leads to virtually zero lack of fit at any of the design points. Calculation of lack of fit at the additional 15 points, however, shows that the error occurring at locations away from the factorial points is substantial, indicating that there is a significant amount of curvature in the actual response surface. This is a cause of considerable concern, as the maximum error at any of these additional points is over 1.5 dB(A), which represents over 20% of the range through which the noise function varies within the experimental region. The average error at these points is nearly 16%. If these values are typical of the predictive ability of the model at locations away from the vertices of the n–dimensional hypercube, then the use of this model is likely to prove extremely misleading in the identification of favourable areas of the design variable space.

In order to improve the fit at these points it is necessary to use an enhanced predictive model which will allow for the effect of higher order terms. This in turn requires a more complex experimental design, in order to obtain the additional function information which is needed. Such designs and models are the subject of subsequent chapters. If further parameters are to be added to the present 128 terms of the fitted model, however, then a design for fitting a

higher order model must also include at least 128 tests to allow all of the linear + interaction terms to be estimated. It is important to establish, therefore, whether all of the terms of the present model are contributing significantly to the overall predictive ability, or whether, as has been found to be the case for the mass function, a number of these can be dispensed with without adversely affecting the ability of the model to provide an accurate representation of the true response surface.

4.1.3 Assessment of model parameters

The parameter listing of Table 4B.1 shows that the dominance of the main effect terms is much less pronounced than was the case for the mass function. Eleven of the interaction terms have parameter estimates which are within an order of magnitude of the smallest of the main effect terms, with a further 90 terms being within two orders of magnitude. In contrast with the mass function, no prior knowledge of the nature of the response is available to aid interpretation of these parameter estimates. Since the noise response is known to be a highly complex function of the design variables, however (see, for example, Zhang, 1992), it is probable that the linear + interactions model is insufficient to provide an accurate representation of the variation in noise throughout the design variable space. This view is supported by the lack-of-fit results of Table 4.3. It is also possible that a number of the terms of the present model are in fact confounded with higher order terms (quadratic, cubic etc.), the effects of which cannot be independently assessed when tests are carried out at just two levels of each variable. If the higher order terms which have been neglected do not have a dominant influence on the overall shape of the response surface, then their presence may be expected to have relatively little effect on the estimation of the larger coefficients of the present model. Since the model contains as many parameters as there are tests in the design, however, all variations in the response at the test points must be attributed to those terms which are available. Confounding between negligible linear interaction terms and the effect of unknown higher order terms may thus result in non-zero parameters being ascribed to terms which in fact have no significant influence on the true response. The inclusion of these spurious effects will then lead to misleading predictions of the response function at other points within the region of interest.

If the perturbations in the parameter estimates of the linear + interactions model which are caused by underspecification of the fitted model can be considered as being randomly distributed among the model terms, then this effect may be treated as a source of random error, or statistical noise, in the construction of the mathematical model. If this is the case, then those parameter estimates which are dominated by this effect will be expected to be indistinguishable from a set of normally distributed random data. In order to determine which of the terms represent real characteristics of the response surface, and which are due to confounding caused

by underspecification of the fitted model, some means is required of testing the hypothesis that all parameters are indeed just random values. The test which has been used within the present work is the probability plot method, which is described in Appendix 4C. The following sections give examples of how probability plots have been used to assess which coefficients of the linear + interactions model are statistically significant.

4.2 Use of probability plots in identifying significant parameters

A normal probability plot for the full set of 127 parameters in the model of Section 4.1.2 (the mean is not included in a probability plot analysis) is shown in Figure 4.1. The method used to calculate the 'Z score' value, which is a normalised measure of the rank order of the coefficients, is described in Appendix 4C.

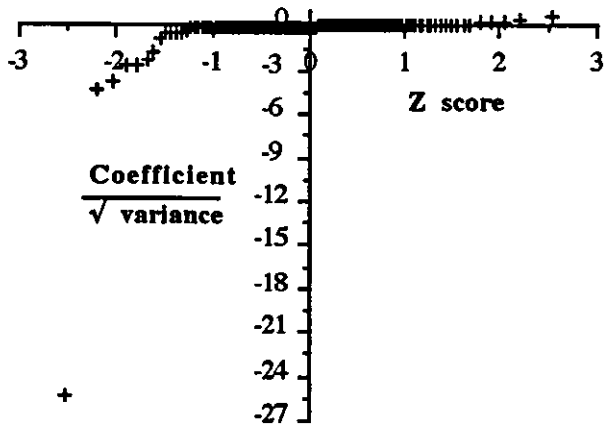

Figure 4.1 Normal probability plot for full factorial test in seven variables
Function : noise
No. of coefficients = 127 of 127

It can be seen from this graph that, of the 127 parameters plotted, the majority appear to lie on an approximate straight line through the origin of the graph, suggesting that the smaller coefficients (in magnitude terms) of the model may indeed be normally distributed, and with an approximately zero mean value. The only parameters which deviate significantly from this line are the seven coefficients with the largest negative magnitudes, and these must therefore be removed from the plot, since they are clearly not due to random statistical noise. It seems probable that a number of other terms may also lie away from the projected 'statistical noise line', but these are extremely hard to identify from this initial plot, due to the compressed scale

resulting from the magnitudes of the largest coefficients. With the seven largest coefficients removed, the resulting graph, with the Z score values recalculated to reflect the reduced amount of data, is shown in Figure 4.2.

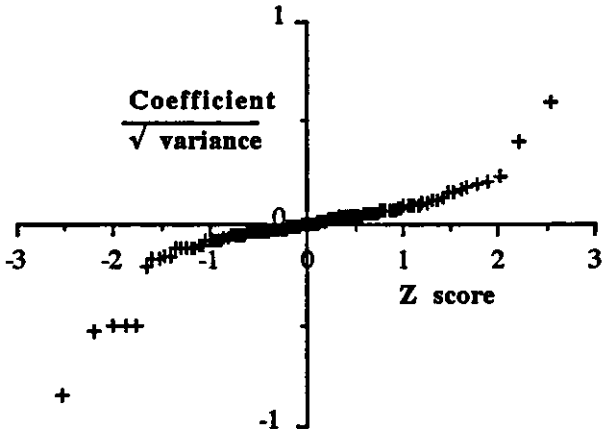

Figure 4.2 Normal probability plot for full factorial test in seven variables
Function : noise
No. of coefficients = 120 of 127

This graph allows the identification of a further seven parameters which have magnitudes inconsistent with an assumption of normal distribution; five negative and two positive. These seven are removed from the analysis, and the remaining 113 terms plotted against revised Z scores, as shown in Figure 4.3.

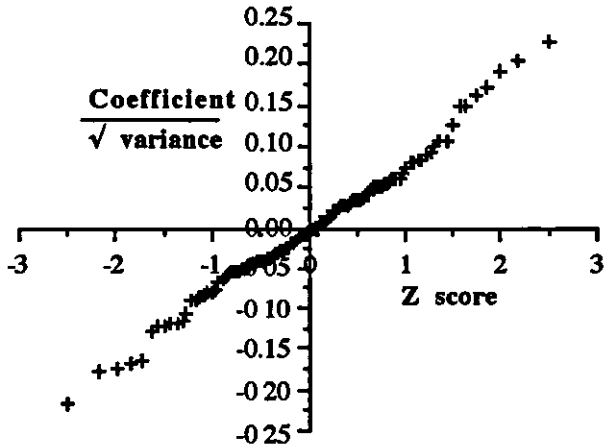

Figure 4.3 Normal probability plot for full factorial test in seven variables
Function : noise
No. of coefficients = 113 of 127

All of the parameters plotted in Figure 4.3 now lie on an approximate straight line through the origin of the graph, with any slight deviations from linearity being attributable to sampling error (see Appendix 4C). This result suggests that the parameter values of Figure 4.3 are indistinguishable from a set of normally distributed random data, and allows the investigator to accept the hypothesis that the remaining 113 parameters of the model are due to 'random' effects, rather than to any underlying characteristic of the system under investigation.

It is interesting to note from the plot that the 'statistical noise line' passes almost exactly through the origin of the graph, with just four of the negative parameters having positive Z scores, and that the two points lying at the ends of the distribution are of very similar magnitude (-2.184 and 2.258). This provides further reassurance as to the validity of the method when applied to the present deterministic environment, since the parameters which have been found to be statistically insignificant, and are hence assumed to be due to a random error effect, have also been shown to have an approximately zero mean.

4.2.1 The half-normal probability plot

Upon reconsidering the above process, by which the significant terms of the model have been identified, it is clear that no distinction has been made between terms which are of equal magnitude and opposite sign. In determining which of the parameters to remove from the probability plot analysis, it was purely the magnitude of the relevant parameters which was of interest. This is of course due to the fact that the degree of impact which a given parameter, whether positive or negative, has on the predictive ability of the fitted model is determined simply by its magnitude. This being the case, the parameter estimates may alternatively be displayed using the half-normal probability plot, or Daniel plot, as described in Appendix 4C.

To compare the two methods, a half-normal probability plot of the original 127 parameters is shown in Figure 4.4.

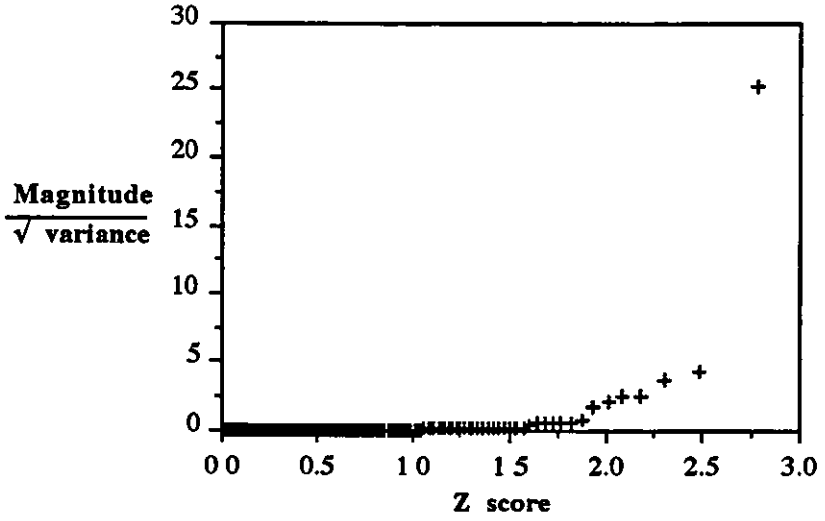

Figure 4.4 Half-normal probability plot for full factorial test in seven variables
Function : noise
No. of coefficients = 127 of 127

The scale of this plot is again compressed by the magnitude of the largest term, but seven parameters are still identifiable as deviating substantially from the approximate straight line formed by the smaller coefficients. Following the same procedure as used with the normal probability plot, these seven terms are removed from the analysis and the Z scores recalculated accordingly. The resulting plot is shown in Figure 4.5.

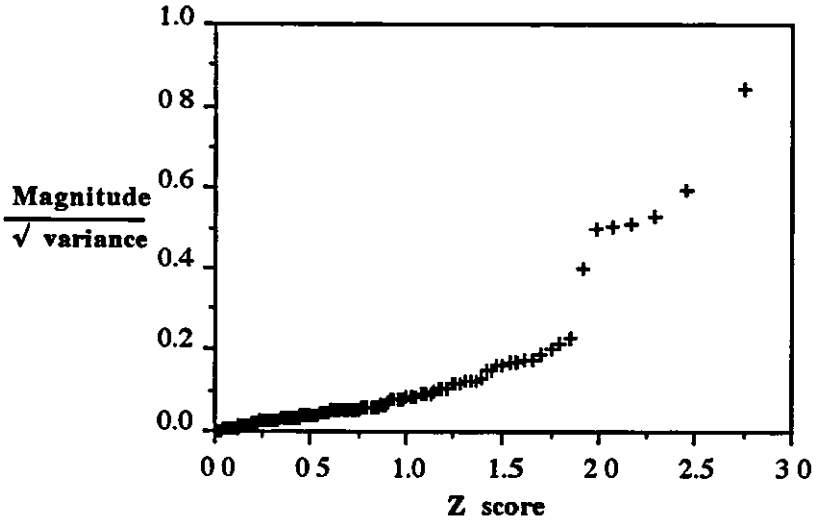

Figure 4.5 Half-normal probability plot for full factorial test in seven variables
Function : noise
No. of coefficients = 120 of 127

This shows that a further seven terms lie significantly above the projection of the 'statistical noise line', with all remaining parameters appearing to lie on a straight line through the origin. This may be confirmed by removing the seven significant terms from the analysis and replotting, to give the graph of Figure 4.6.

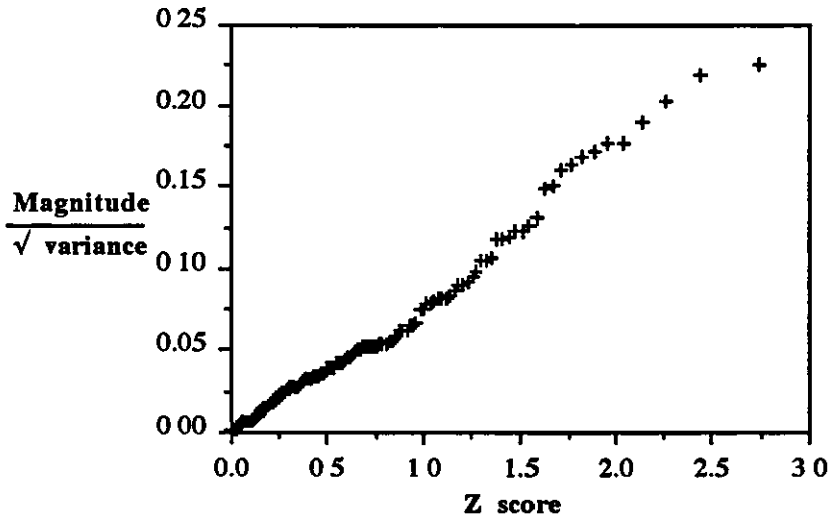

Figure 4.6 Half-normal probability plot for full factorial test in seven variables
Function : noise
No. of coefficients = 113 of 127

All parameters now lie on an approximate straight line, thus leading to the same conclusion as was reached using the normal probability plot analysis; that 113 of the original 127 parameters are indistinguishable from normally distributed random data. Examination of the data used in each analysis reveals that, as is expected, the parameters which have been found to be significant are the same in each case. The advantage of using the half-normal probability plot, however, is that all significant parameters deviate from the 'statistical noise line' in the same direction, and at just one side of the graph. This allows the magnitude of all terms to be directly compared, irrespective of sign, and may aid the rapid identification of significant parameters. As the above example has demonstrated, however, equivalent results may be obtained using either of these methods.

4.2.1 Are probability plots necessary ?

Before studying further the reduced model which would result from removal of the insignificant terms, it is perhaps worth considering whether the above identification of significant parameters could have been successfully accomplished without recourse to the method of probability plot analysis. An obvious alternative would be simply to plot the magnitudes of the coefficients against their rank order, and try to differentiate between significant and insignificant terms purely on the basis of their magnitudes. Figure 4.7 shows a plot of coefficient magnitude against rank order for the full set of 127 parameters (the mean is again removed prior to plotting).

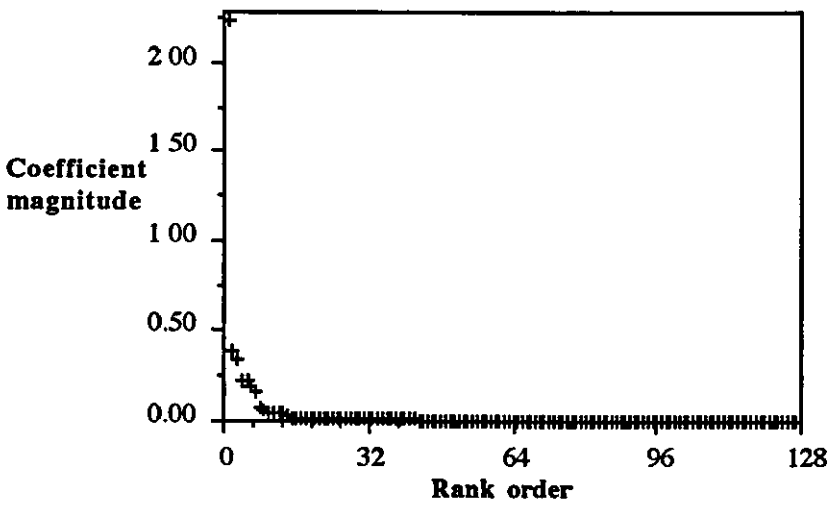

Figure 4.7 Model coefficients for full factorial test in seven variables
Function : noise
No. of coefficients = 127 of 127

So far the method appears to be promising, since seven terms are again easily identified as being substantially larger than the others. Further terms are hard to assess due to the scaling imposed by the magnitude of the larger coefficients. Re-scaling the y-axis to exclude the seven largest terms produces the plot of Figure 4.8

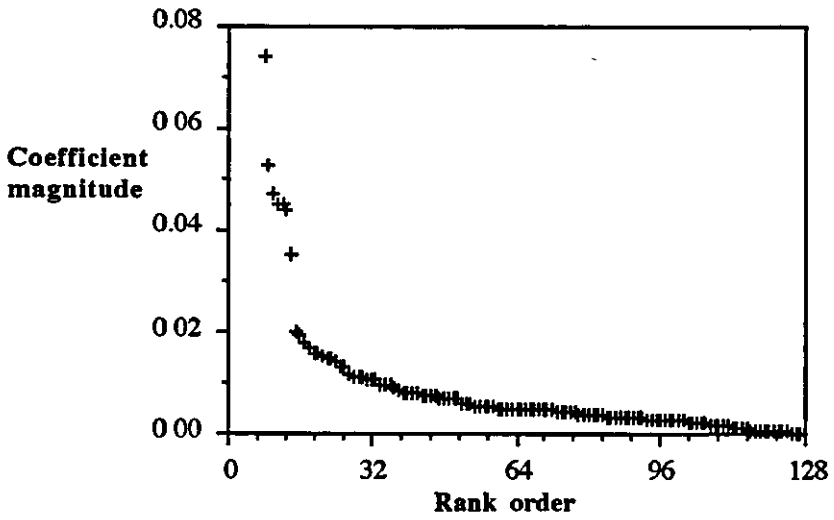

Figure 4.8 Model coefficients for full factorial test in seven variables
Function : noise
No. of coefficients = 120 of 127

Again, a further seven terms have been revealed as being distinct from the remaining parameters. Since the probability analysis has already shown there to be fourteen significant terms, one might conclude that an equivalent result had been obtained using this much simplified method. If the method were used without this prior knowledge, however, the next step of the procedure would necessarily be to re-scale the plot again to remove the seven largest terms. This yields the graph of Figure 4.9.

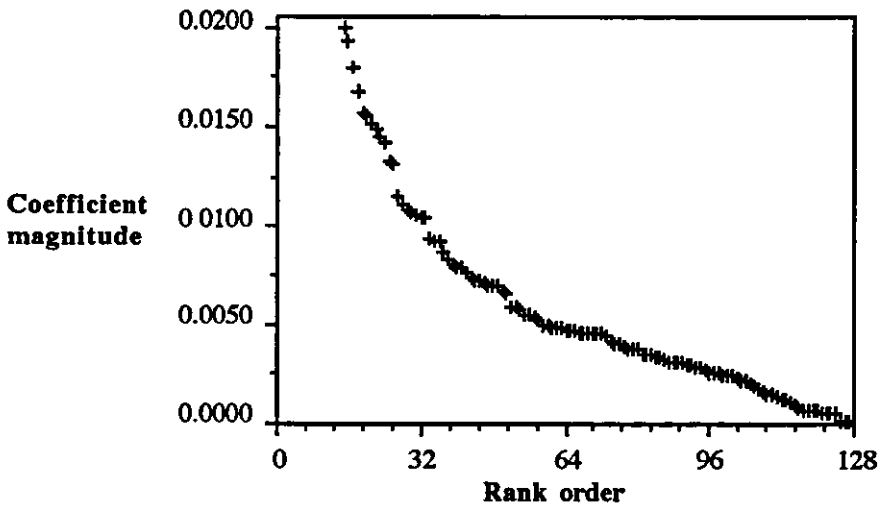

Figure 4.9 Model coefficients for full factorial test in seven variables
Function : noise
No. of coefficients = 113 of 127

This graph now suggests that a further four terms may be worthy of inclusion in a reduced model, although the ever increasing curvature of what is assumed to be a 'statistical noise line' gives much cause for concern. It becomes increasingly clear that no distinct cut-off point between significant and insignificant terms is likely to be forthcoming using this method, so that the plotting of coefficient magnitudes against rank order leads to little advantage over the parameter listing of Table 4B.1. This observation is further supported by the removal of the largest four terms of the above plot, to give the graph shown in Figure 4.10, below.

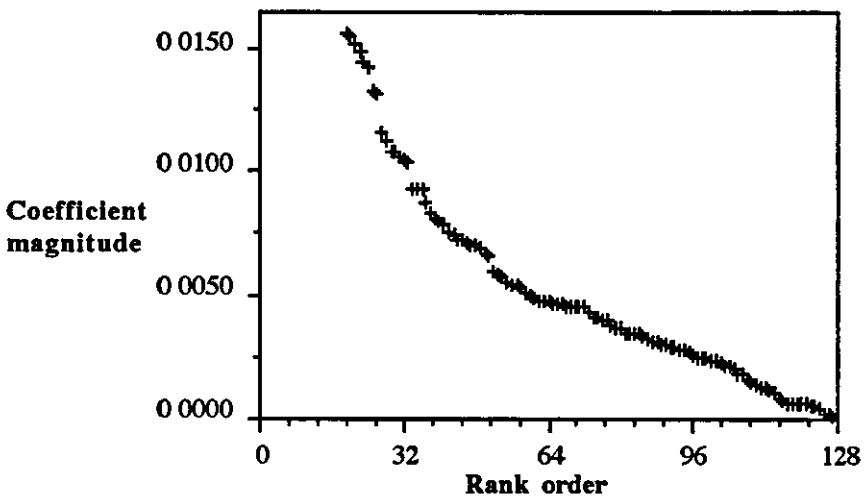

Figure 4.10 Model coefficients for full factorial test in seven variables
Function : noise
No. of coefficients = 109 of 127

It can be seen from comparison of this figure with the previous graphs that, as the scaling of the y-axis is varied, small groups of coefficients are gradually detaching themselves from the main body of parameters. This iterative procedure of removing the largest variables from the plot and re-scaling is thus likely to continue for some considerable time, and lead to no definite conclusions as to the significance of the model terms. The curvature of the line formed by the 109 plotted data points in Figure 4.10 is also in considerable contrast with the almost exact straight lines obtained with 113 points in Figures 4.3 and 4.6.

The conclusion to be drawn from the above example, therefore, is that simply plotting coefficient magnitude against rank order is insufficient to enable the identification of significant terms in the original model. It should also be noted, however, that even if the analysis had yielded a positive result, there is no sound fundamental basis on which to conclude that the

parameters identified are indeed statistically significant, or that those rejected are not. The procedure is reduced to an observation that some coefficients are bigger than others, with the only criterion for distinguishing 'significant' parameters being a slightly larger drop in the magnitude of adjacently ranked terms. This is clearly insufficient justification for making judgments as to the feasibility of a reduced predictive model.

4.3 Validation of a reduced model

Returning to the results of the probability plot analysis, it was found that the statistically significant terms are the fourteen which are ranked highest in terms of their (magnitude/√variance) parameter. Since the variance on the estimate of each parameter is identical, the division of the coefficient values by the square root of the variance will not affect the rank order of the terms, with the result that the terms found to contribute significantly to the model are the first fourteen (non-mean) terms of Table 4B.1. These, together with the mean, are shown in Table 4.4.

1.	MEAN	$8.8082\times10^{+1}$
2.	A	-2.2377
3.	C	-3.8596×10^{-1}
4.	G	-3.3608×10^{-1}
5.	D	-2.3021×10^{-1}
6.	F	-2.2872×10^{-1}
7.	B	-1.9920×10^{-1}
8.	E	-1.5888×10^{-1}
9.	AG	-7.4246×10^{-2}
10.	AD	5.2756×10^{-2}
11.	AF	-4.6907×10^{-2}
12.	CG	-4.4956×10^{-2}
13.	AE	-4.4724×10^{-2}
14.	FG	-4.4068×10^{-2}
15.	ACG	3.5056×10^{-2}

Table 4.4 Significant noise coefficients for full factorial test (n=7)

The results of the probability plot analysis indicate, therefore, that the only parameters of the 128 term model which can be regarded as being statistically significant are the mean and seven main effect terms, followed by six two-way interaction terms and just one three-way interaction. If it is the case that the variation in noise response can be adequately represented

using a much reduced model containing just these fifteen terms, then it is likely that an acceptable model could have been constructed using a much smaller quantity of function information. Further, if it is generally true that the noise response of engine blocks is dominated by main effect and two-way interaction terms, then there is clearly substantial scope for reducing the number of tests which need be performed when undertaking an investigation of a particular system.

Before such a conclusion can be reached, however, it is necessary to establish whether these fifteen terms are indeed sufficient to provide an acceptable representation of the variation of the noise function throughout the design variable space. This can be determined by consideration of the lack of fit between the fitted model and the actual response at the 2^n test points used to estimate the terms of the full 128 term model. If the fifteen terms of Table 4.4 are necessary and sufficient adequately to represent the noise function, then it is expected that any model lacking these terms will exhibit substantial lack of fit at the test points, whereas addition of further terms will lead to negligible improvement in model accuracy. Since the initial model of 128 terms contains as many parameters as there are tests in the experimental design, zero lack of fit (to machine precision) occurs between the actual test values and the representations of these generated by the approximating model, as shown in Table 4.3. It is thus necessary that the reduced 15 parameter model exhibit negligible lack of fit at these points.

The manner in which the lack of fit at the factorial test points varies with the number of coefficients in the model may be investigated by using the full set of $2^7 = 128$ test points to successively construct models containing between 1 and 128 terms. The first of these models will contain just the mean effect, with additional terms being introduced in decreasing order of magnitude. For each model, the difference between the true noise response and the prediction of the model is calculated at a number of the original test points. The lack of fit at each of these points is then plotted against the number of coefficients in the predictive model. The results for a randomly selected sample of five of these points, described by their locations in the normalised design variable space, are shown in Figures 4.11 to 4.15.

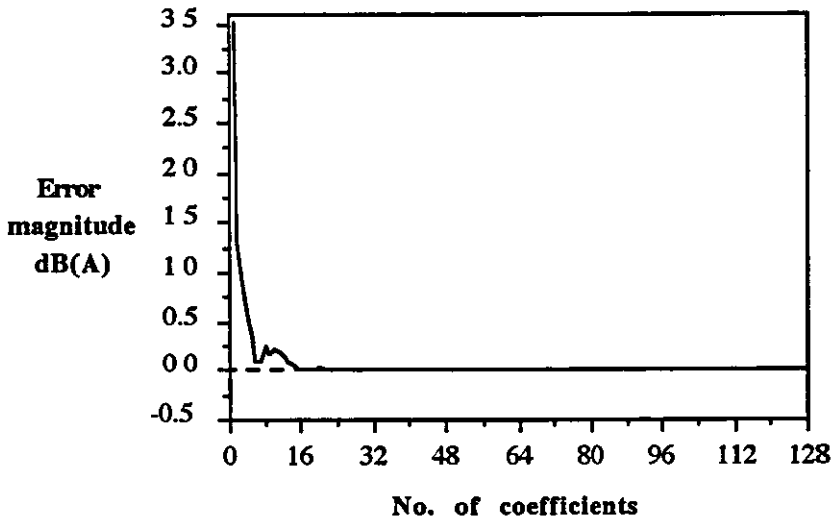

Figure 4.11 Lack of fit against number of coefficients in model
Test point = (-1, -1, -1, -1, -1, -1, -1)

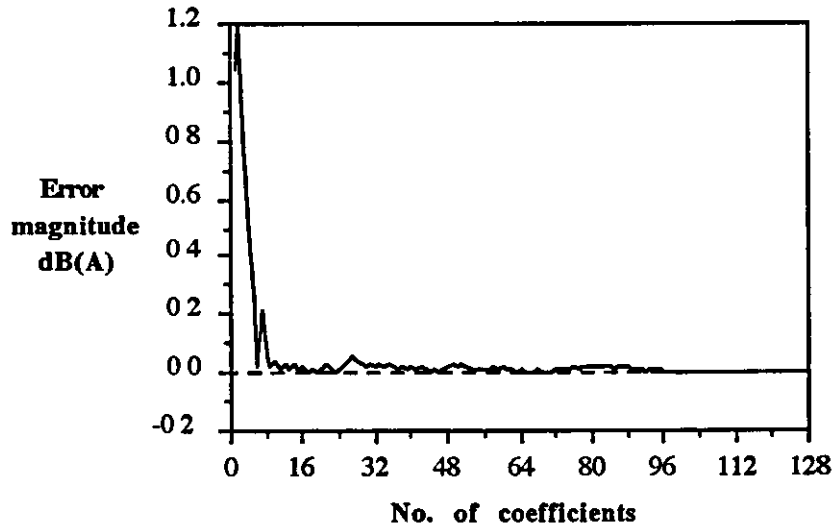

Figure 4.12 Lack of fit against number of coefficients in model
Test point = (1, 1, -1, -1, -1, -1, -1)

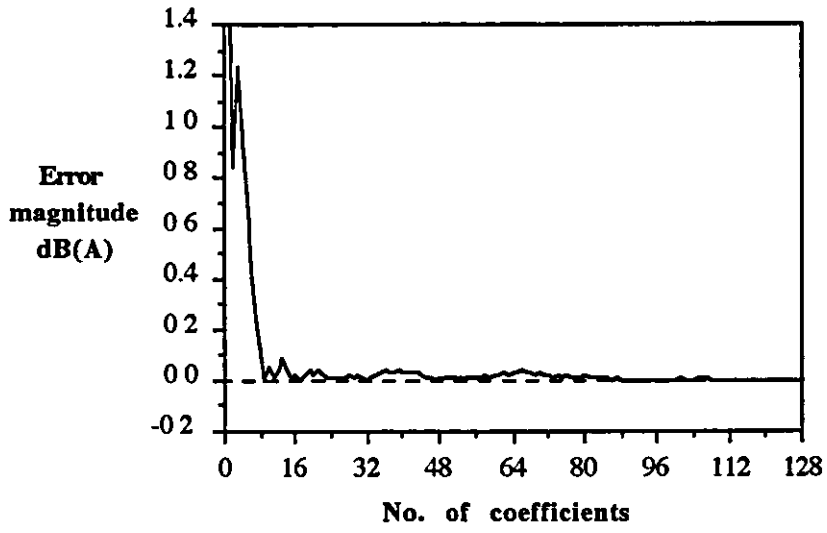

Figure 4.13 Lack of fit against number of coefficients in model
Test point = (1, -1, 1, -1, -1, -1, -1)

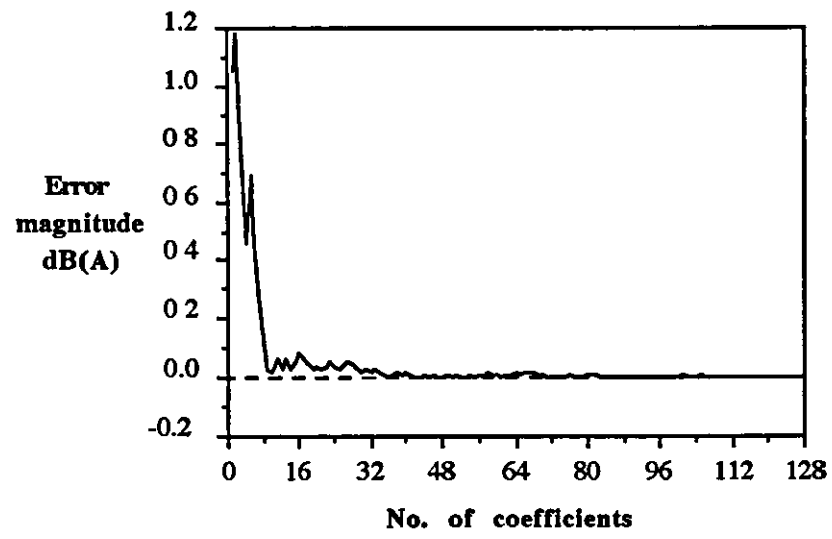

Figure 4.14 Lack of fit against number of coefficients in model
Test point = (1, -1, -1, 1, -1, -1, -1)

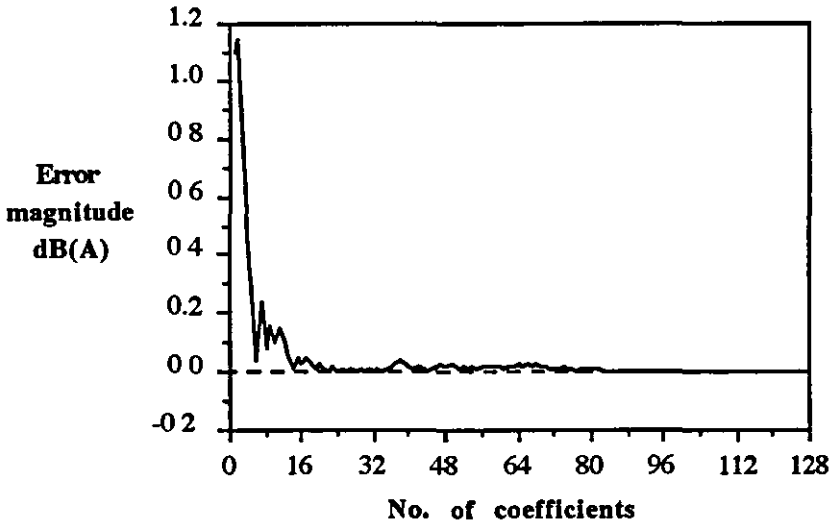

Figure 4.15 Lack of fit against number of coefficients in model
Test point = (1, -1, -1, -1, 1, -1, -1)

Figures 4.11 to 4.15 show that, in each case, the initially high lack of fit drops rapidly as the seven main effect terms are introduced. The high lack of fit value obtained when just one parameter is included in the model is of no particular significance, since this term is the mean value. Lack of fit using this model simply represents the difference between the average response at the 2^n points and the response at the particular point being examined. For three of the five points analysed (Figures 4.12, 4.13 and 4.14), the lack of fit obtained with the eight term (mean plus main effects) model has already fallen to within 0.1 dB(A) — equivalent to just 1.3% of the function range. The lack of fit values at all five points lie within a 3.5% error band (0.26 dB(A)), suggesting that the main effects model is providing a reasonably good approximation to the original surface at the points tested.

The effect of including the remaining seven terms of Table 4.4, to give a total of 15 terms in the predictive model, is to bring the lack of fit at all five points within 0.1 dB(A). All five graphs show that no significant improvement in accuracy is gained by the inclusion of further terms. Although the lack of fit necessarily reduces to zero as the number of terms approaches the full 128 parameters of the original model, the small lack of fit obtained using just 15 parameters provides little scope for any further reduction in the prediction error.

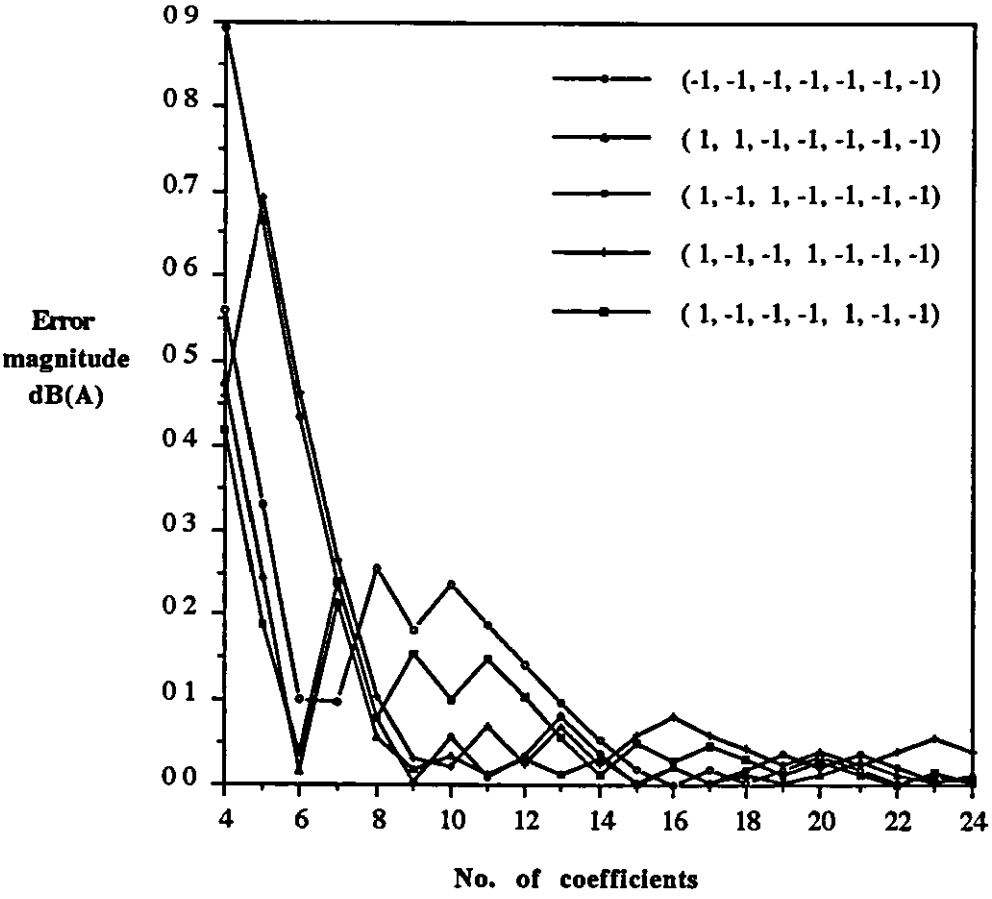

Figure 4.16 Lack of fit against number of coefficients in model Five factorial points

In order further to aid comparison of the above results, a magnified view of the lack of fit at each of the five points, resulting from models containing between 4 and 24 terms, is presented in Figure 4.16. This shows in more detail the convergence of the error at all points to within the 0.1 dB(A) band. This appears to occur when the number of terms in the model reaches 14, with the 15th term, which is the three way interaction ACG, having, on average, a substantially smaller effect on the degree of lack of fit in the model. The overall trend can be seen more clearly by plotting the average lack of fit value obtained from these five individual points, as shown in Figure 4.17. This confirms the presence of the characteristics already noted. A rapid decline in lack of fit with the inclusion of the main effect terms is followed by a more gradual reduction as the remaining significant terms are introduced. Addition of further terms leads to little further improvement.

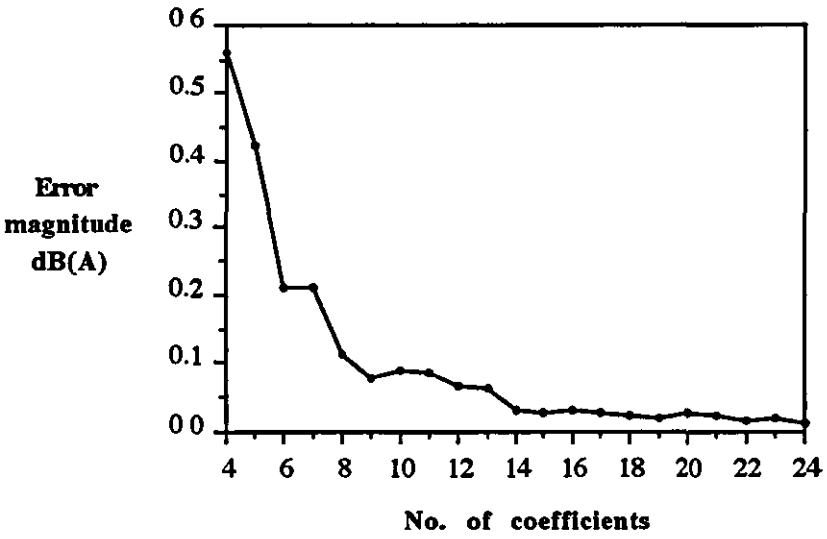

Figure 4.17 Lack of fit against number of coefficients in model Average of five factorial points

Since the object of the modelling process is to provide an acceptable representation of the original response surface across the whole of the design variable space, and not just at the factorial points, it is relevant to inquire whether it is possible to carry out an analysis of the above type at other locations within the region of interest. A suitable selection of tests points would appear to be the centre and star points introduced in Section 4.1, above, which are situated at locations distant from the vertices of the hypercube.

In attempting to carry out the 'lack of fit v. model size' analysis at these points, however, two main obstacles are encountered. The first of these is that, as shown in Table 4.3, substantial lack of fit, due to curvature of the original response surface, is already known to exist at these points. The magnitude of this error is such that it is likely to dominate any calculations of modelling inaccuracy carried out at these points. Of much more importance, however, is the fact that, at these locations, a maximum of just one variable has a non-zero (normalised) value. The result of this is that, at the centre point, for example, where all variables have a zero value, the only term which contributes to the model is the mean term, with all terms involving any of the variable values being set to zero. A graph of lack of fit against model size will thus simply yield a horizontal straight line, with the error value being equal to the difference between the measured value and the mean term of the linear model. At the star points, where one of the variables takes a value of ±1, the only non-zero terms will be the mean value and the main effect in that variable. All other main effects, as well as all interaction terms, which by definition include at least one of the other (zero valued) variables, will again have no influence on the model. This will produce a horizontal straight line with a single step in it, occurring when

the non-zero term is introduced. These effects are demonstrated by the three points whose results are plotted in Figure 4.18.

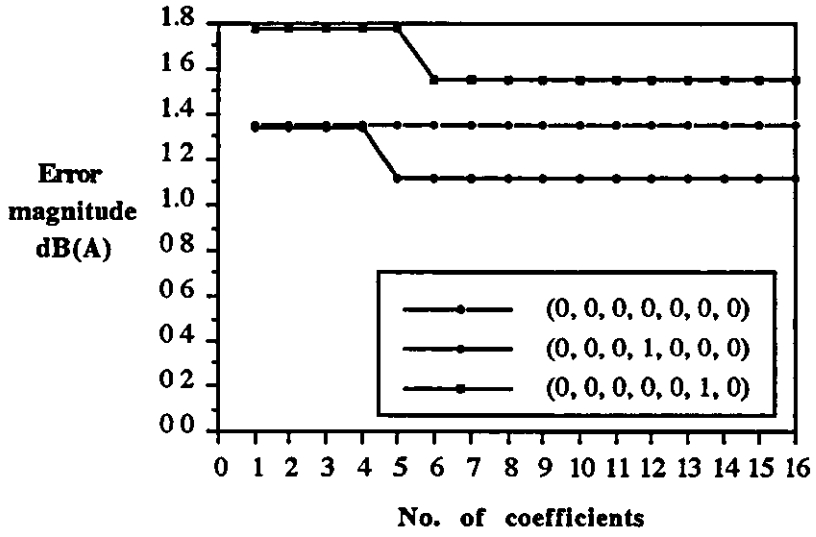

Figure 4.18 Lack of fit against number of coefficients in model centre point and two star points

It can be seen from Figure 4.18 that the step in the error value for the point (0, 0, 0, 1, 0, 0, 0), with variable D having a non-zero value, occurs on the introduction of the fifth model term. This corresponds with the result of Table 4.4, in which the main effect D is the fifth largest parameter. Similarly, the step for the point (0, 0, 0, 0, 0, 1, 0), with variable F non-zero, occurs as the sixth term is introduced, again agreeing with the parameter order of Table 4.4. This last test point is also the location at which the maximum lack of fit occurring at any of the centre or star points is found. Its final error value of 1.55 dB(A) corresponds with the maximum value given in Table 4.3.

Since these results have shown that test points lying on variable axes are of no use in determining the effect on lack of fit of reducing the number of terms in the linear model, it is necessary to choose some alternative test points at which to carry out the analysis. The requirements for such points are that they should lie at sufficient distance from each of the variable axes, and the centre point, and yet still be far enough away from the original test points to provide more general information concerning the performance of the predictive model throughout the design variable space as a whole. Suitable candidates for such tests would appear to be points which lie half way between the centre of the space and the factorial points, at a distance of $0.5\sqrt{n}$ from the centre. These points then lie on an n–dimensional hypercube of side n/2, having as its centre the centre of the design variable space, and with its sides aligned

with the sides of the original 2^n factorial hypercube. This new hypercube will of course contain 2^n points, although for reasons of economy a fraction of these points can be used, calculated as for the fractional factorial design. In the present case a 1/8 fraction containing 16 tests has been used. The exact specification of these points is given in Appendix 6A.

Calculation of the lack of fit at these points proceeds as for the original factorial points, with each of the parameters of the full model being successively introduced in descending order of magnitude. The results obtained for all 16 of these additional points are displayed in Figure 4.19.

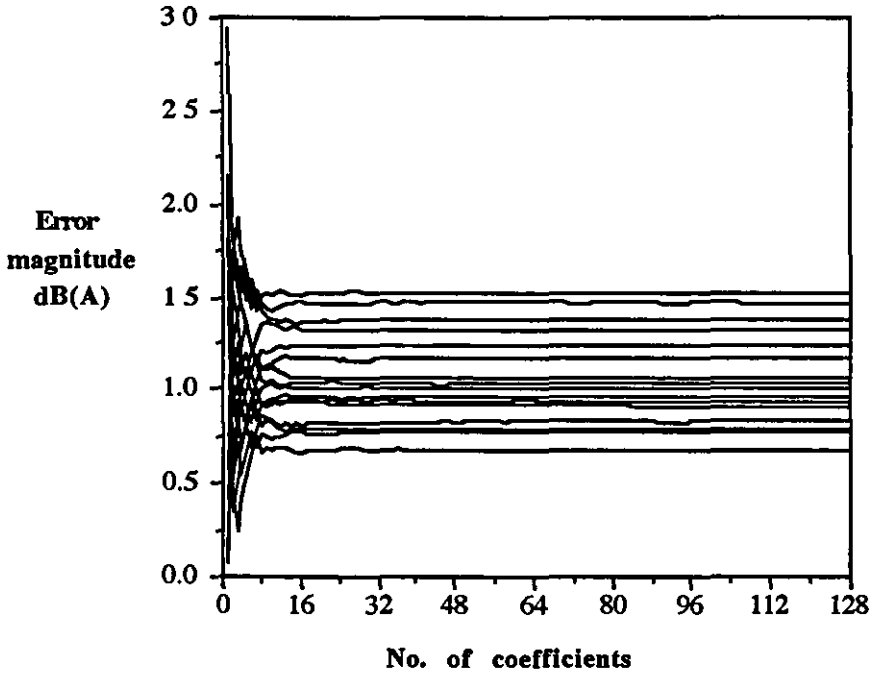

Figure 4.19 Lack of fit against number of coefficients in model 16 factorial points with parameter n/2

Two important features of these results may easily be identified. Firstly, in common with the results at the centre and star points, the lack of fit at these locations is dominated by the presence of higher order effects, which are not accounted for in the linear model. Thus, even with the full 128 terms, significant error in model prediction still occurs. The second observation, however, is that, in each case, the lack of fit appears to have stabilised at about this final value with the inclusion of just 16 or so terms. These results thus appear to confirm the conclusion drawn from the lack of fit calculations at the original test points, in that little improvement in the predictive ability of the linear model is obtained by including terms other than those which have been shown to be significant by the probability plot analysis.

The purpose of the present analysis is to investigate the predictive ability of reduced models in comparison with the performance of the full 128 term model, rather than to assess the absolute lack of fit occurring at particular points. A better approach might therefore be to plot the value of the error magnitude which occurs with respect to this full model, thus removing from the analysis that component of the lack of fit which is due to higher order terms, and cannot be represented by a linear + interactions model of any kind.

Such a plot is shown in Figure 4.20, and provides a much clearer picture of the effect of model reduction on prediction error. As was found for the analysis at the original test points, a rapid fall in lack of fit occurs as the seven main effect terms are brought into the model. This is followed by a more gradual reduction as the next seven or so terms are introduced, with virtually all lack of fit having been eliminated by this time.

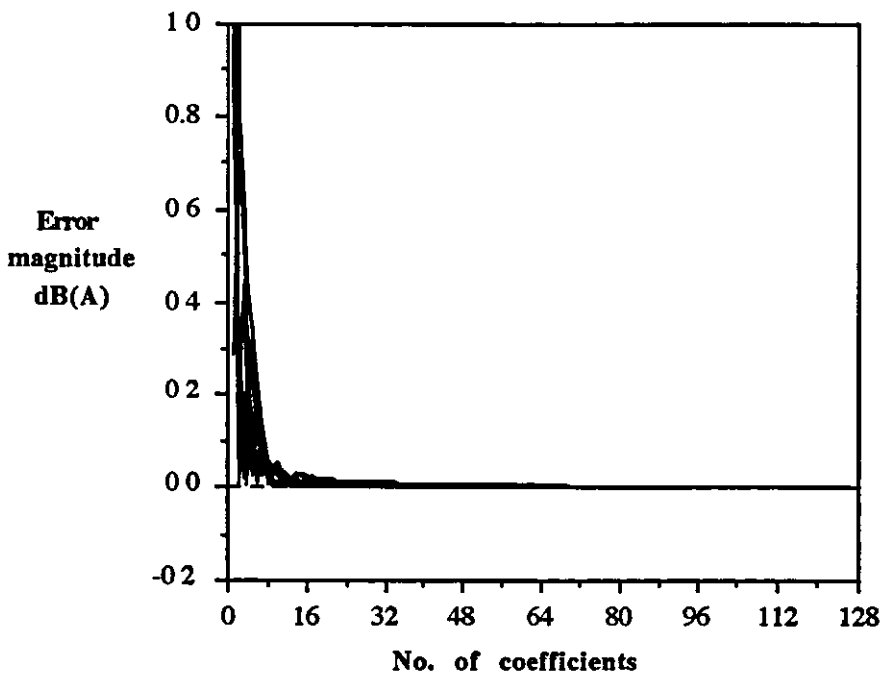

Figure 4.20 Standardised lack of fit against number of coefficients in model 16 factorial points with parameter n/2

The graph of Figure 4.21 shows a magnified view of this plot, from which it may be seen that the 8 term main effects model produces, in all cases, a lack of fit value of less than 0.06 dB(A) — representing 0.8% of the function range. The inclusion of 15 terms reduces this error to less than 0.03 dB(A) (0.4%).

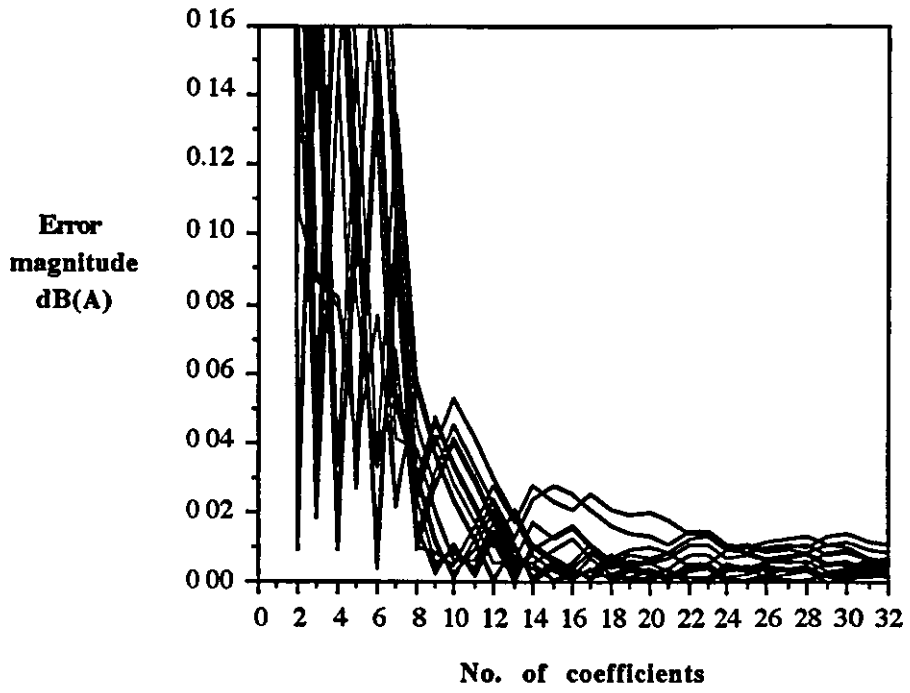

Figure 4.21 Standardised lack of fit against number of coefficients in model 16 factorial points with parameter n/2 - magnified view

This headline figure, however, masks the fact that many of the results are grouped at the lower end of the scale, as is shown by Figure 4.22, in which the average of the above 16 lack of fit values is plotted against model size. The average error resulting from use of the main effects model is now about 0.03 dB(A) (0.4% of range), with the 15 term model giving an average of just 0.01 dB(A), or 0.13% of the function range.

The results of the lack of fit analysis at these additional 16 test points thus confirms the conclusions drawn from the tests carried out at the original design points. The overall conclusion from this analysis is that the 15 terms, including the mean, identified by the probability plot technique account for virtually all of the variation in noise response which can be represented using a linear + interactions model. When using a 15 term model, substantial lack of fit still occurs at points away from the vertices of the design variable space, due to higher order terms. This lack of fit cannot be much reduced, however, by including the remaining terms of the full linear + interactions model.

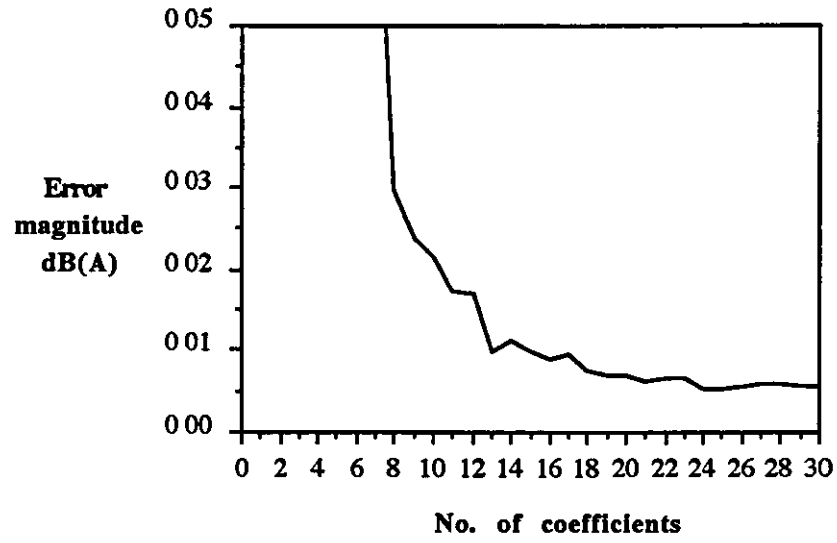

Figure 4.22 Average standardised lack of fit against number of coefficients in model 16 factorial points with parameter n/2 - magnified view

4.4 Conclusions based on full factorial test

Two independent methods have now been used to assess the coefficients of the 128 term linear + interactions model derived from the results of the seven variable full factorial experimental design. The first of these, the probability plot analysis, provides a test of the statistical significance of each parameter of the full model. This method showed that of the 128 parameters, only fifteen were clearly distinguishable from random data, and hence only these terms are reliable as indicators of the nature of the true response surface.

In the second method, the lack of fit between the predicted response of the model and the experimental value is calculated at a selection of test points, and plotted against the number of terms included in the model. This provides a test of what may be termed 'engineering significance', in that it reflects the usefulness of each term in improving the quality of the decisions which are taken, based on the approximating model. This method confirmed that those terms found to be insignificant by the probability plot analysis make a negligible contribution to the prediction accuracy of the model throughout the design variable space. Further, it was found that little additional lack of fit resulted from the use of a model containing just the mean and main effect terms.

4.5 Use of reduced models

The effect of a reduction in model size can be further demonstrated by calculating the maximum and average lack of fit which occurs in each of the categories of test points described in Section 4.1.1. The results for three reduced models are summarised in Table 4.5. Only the results at the original 128 design points are shown, since the lack of fit errors at the centre and star points are identical in each case. As discussed in Section 4.3, this is because the function prediction at these locations is only determined by the mean and main effect terms, which are present, and identical, in each model.

The first of these models contains just the mean and main effect terms, giving an average lack of fit at the design points of 1.6% of the variable range. This model still allows a maximum error of nearly 8% (0.6 dB(A)) at the design points, however, and this, following the discussion of Section 2.14, is likely to be of some concern in the identification of a favourable region of the design variable space.

The maximum error can be reduced to under 0.35 dB(A) (5%) by the use of the second model of Table 4.5, which includes all of those parameters found to be statistically significant by the probability plot analysis of Section 4.2. Additionally, the average lack of fit at the design points has been more than halved, to less than 0.06 dB(A) (0.8%).

Model	No. of terms	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
Main effects	8	0.59851	7.9236	0.12119	1.6045
Significant terms	15	0.34961	4.6718	0.05769	0.7710
≤ 2-way interactions	29	0.36697	4.8582	0.05618	0.7439

Table 4.5 Summary of lack of fit calculations for various noise models using full factorial design with 128 tests (n=7)

The main problem which occurs in using a reduced predictive model containing only significant terms is that it is not in general possible to know in advance which of the terms are to be included. It is only by carrying out a probability plot analysis, or similar procedure, that this information can be gained. Such an analysis can only be performed, however, if a sufficiently large number of tests are first carried out to enable all parameters to be estimated, thus precluding the use of a smaller experimental design.

To avoid this problem, a model may be used which includes all terms of each of the types thought likely to be significant. When investigating applications for which the results of Table 4.4 are representative, for example, a suitable model would be one which contains all main effect and two-way interaction terms. This procedure introduces two possible sources of error into the model specification. The first is due to the inclusion of terms which are in fact of no significance. As long as parameter estimates remain largely orthogonal, however, and the number of insignificant terms is fairly small, this is not expected to lead to substantial error. The second source of error is due to the exclusion from the model of significant parameters, such as, in the present example, the three-way interaction ACG of Table 4.4. If the excluded parameters are small in number and do not dominate the analytical model, then the attendant loss of accuracy would probably be considered a small price to pay for the large reduction in the number of design points, compared with those which would need to be tested were all terms of that type to be included in the model.

These observations are supported by the lack of fit values which result from the use of the third model of Table 4.5. This model contains the mean value and seven main effect terms, plus the 21 two-way interaction terms. The only statistically significant term excluded from this model is the three-way interaction term ACG. Table 4.5 shows that the performance of this model is little different to that of the previous model. Its advantage, however, is that it is likely to be more widely applicable to other problems of a similar type.

4.6 Use of fractional factorial designs

Since it has now been established that a model containing just mean, main effect and two-way interactions terms is sufficient to represent adequately the variation of noise response at the vertices of the design variable space, it is now possible to seek an experimental design which will allow for the estimation of the required parameters using a minimum number of analyser calls. It was shown in Section 3.5.4 that in order to estimate independently the mean value and all of the main effect and two-way interaction terms, a fractional factorial design must be of at least resolution V. Table 3.12 shows that the smallest design in seven variables which meets this criterion is the 2^{-1} fraction containing $2^{7-1} = 64$ tests, which is of resolution VII.

The 64 parameter estimates obtained using this resolution VII design are given in Table 4B.2 of Appendix 4B. Comparison of these with the values listed in Table 4.4 shows that the mean and main effect terms are all identical, to at least three significant figures, with the fifteen significant parameters identified by the probability plot analysis remaining the largest terms. The values of the two-way interaction parameters have been modified slightly due to the presence of confounding, although this effect is fairly small due to the high resolution of the design. Using a resolution VII design, the mean value is only confounded with the single

seven-way interaction term, each main effect is confounded with a six-way interaction, and two–way interactions are each confounded with an interaction of order five.

The similarity between the estimates of model parameters would lead one to expect that the performance of the model throughout the design variable space would be comparable with the ≤ 2-way interaction model of Table 4.5. A summary of the lack of fit data for the resolution VII design, calculated at each of the full $2^7 = 128$ factorial points, is shown in Table 4.6. The inclusion of all 128 points, rather than just the 64 of the resolution VII design, is more representative of the overall predictive ability of the 64 term model, since the lack of fit at the resolution VII design points will be effectively zero due to the saturated nature of the design.

Design Resolution	No. of tests/terms	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
Full	128	4.8317×10^{-13}	6.4575×10^{-12}	1.2512×10^{-13}	1.6722×10^{-12}
Resolution VII	**64**	**0.2080**	**2.7814**	**0.0348**	**0.4652**
Resolution IV	32	0.3646	4.8724	0.0521	0.6959
Resolution IV	16	0.5989	8.0047	0.0941	1.2571
Resolution III	8	0.6677	8.4699	0.1605	2.0362

Table 4.6 Summary of lack of fit calculations at 128 full factorial points using fractional factorial designs

Table 4.6 shows that the 64 term model derived from the resolution VII design performs slightly better than the 29 term model of Table 4.5. This is because the mild confounding introduced by halving the number of test points is more than outweighed by the inclusion of the 35 three-way interaction terms. Although most of these are of fairly small magnitude, their inclusion leads to an exact fit at 64 of the 128 points, thus reducing the overall error.

The lack of fit values resulting from the use of 2^{-2}, 2^{-3} and 2^{-4} fractions are also included in Table 4.6. These show that the use of the resolution IV design involving 32 tests results in only a modest increase in the prediction error at the original 128 points. This is despite the fact that confounding is now occurring between two-way interaction terms, several of which are known to be significant. The reason for this becomes clear when the pattern of confounding between parameters is examined. This pattern is given in Table 4B.3 of Appendix 4B, in which the terms listed in the left-hand column of the table are those which are included in the 32-term model. Each term of this column is confounded with the three terms which appear on the same row of the table. Comparison of this table with the list of significant parameters of Table 4.4

shows that each of the significant terms appears in the left-hand column of Table 4B.3, and is hence included in the model. This may be regarded as a fortunate coincidence, since the particular relationship between the pattern of confounding and the significant terms has resulted in a model in which no two significant terms are confounded, and additionally, in which the significant term of each confounding set is the one which has been included in the model. In general, this cannot be expected to occur, with the result that the lack of fit errors throughout the design variable space would necessarily be greater.

Inclusion of all significant terms does not occur, however, for the $2^{7-3} = 32$ test design, in which the size of each confounding set is further doubled. Although this design is also of resolution IV, the pattern of confounding, shown in Table 4B.4 of Appendix 4B, reveals that a number of the significant parameters are now absent from the fitted model, with the result that a larger increase in prediction error occurs, as shown in Table 4.6. The maximum lack of fit at any of the 128 points is now over 8% of the function range, and, although the average error is still less than 0.1 dB(A), this model may be expected to yield misleading results in the search for optimum designs.

The 8 term model which may be constructed using the resolution III design is identical to the main effects model of Table 4.5. The inferior prediction performance attained by the present model is due to the high level of confounding which occurs, with each parameter being confounded with 15 others. Additionally, only a $1/16^{th}$ fraction of the original points is being used to assess these terms. Bearing this in mind, it is a reflection of the overall dominance of the main effect terms, as shown in Section 4.3, that the average lack of fit using this model is as low as the 2% of function range which is in fact obtained.

The relatively low lack of fit values calculated at the vertices of the design variable space are, of course, in marked contrast with the prediction errors which occur at the centre and star points. These values are given, for each of the above models, in Table 4.7. As discussed previously, the accuracy of prediction at these points is determined entirely by the values of the mean and main effect terms. Since these are not confounded with any of the other significant parameters when the design is of resolution IV or greater, the error at the centre and star points is little changed from that obtain using the full 128 term model. When using a resolution III design, however, the main effects are each confounded with a number of two-way interactions, so that their parameter estimates are substantially modified, as shown in Table 4.8. Most noticeable are the changes to parameters C and E.

Design Resolution	No. of tests/terms	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
Full	128	1.5486	20.697	1.1890	15.891
Resolution VII	**64**	**1.5426**	**20.624**	**1.1847**	**15.839**
Resolution IV	32	1.5520	20.742	1.1900	15.904
Resolution IV	16	1.5891	21.238	1.1993	16.029
Resolution III	8	1.5505	19.668	1.1688	14.826

Table 4.7 Summary of lack of fit calculations at centre and star points using fractional factorial designs

Table 4.7 shows that the consequence of this error in main effect estimation is actually a reduction in the average lack of fit at the centre and star points. This, however, is entirely attributable to the modified mean value. Examination of the individual lack of fit figures shows that in each case the mathematical model is producing a prediction which is higher than the experimental result. For each pair of star points, the effect of a modification to that particular variable will thus have an equal and opposite effect on each member of the pair. The slightly reduced mean will, however, reduce the lack of fit at every point, and comparison of Tables 4.7 and 4.8 shows that the difference in average lack of fit between the full resolution and resolution III designs is equal to the change in the estimate of the mean parameter.

Term	Estimation based on full factorial design	Estimation based on resolution III design
MEAN	$8.8082\times10^{+1}$	$8.8062\times10^{+1}$
A	-2.2377	-2.2439
C	-3.8596×10^{-1}	-4.5327×10^{-1}
G	-3.3608×10^{-1}	-3.7115×10^{-1}
D	-2.3021×10^{-1}	-2.3684×10^{-1}
F	-2.2872×10^{-1}	-2.0659×10^{-1}
B	-1.9920×10^{-1}	-1.7140×10^{-1}
E	-1.5888×10^{-1}	-2.5857×10^{-1}

Table 4.8 Estimates of mean and main effect parameters using designs of full resolution and resolution III

In investigating the increase which occurs in the maximum prediction error, compared with the full factorial design, it is found that the greatest lack of fit using the full model occurred at the star point for which F is at its upper bound - (0, 0, 0, 0, 0, 1, 0). The increased error is due to the fact that the decrease in magnitude of 0.0221 dB(A) in the parameter estimate of F is just larger than the drop of 0.0202 in the value of the mean. When F is at its high bound, these two effects oppose each other, such that the lack of fit increases by (0.0221 - 0.0202) = 0.0019 dB(A), as shown in Table 4.7. Although a larger increase in lack of fit occurs at other star points involving high bounded variables, the increase is not sufficient to raise the absolute error higher than that found at (0, 0, 0, 0, 0, 1, 0).

These results show that the lack of fit throughout the design variable space is dominated by the error resulting from the inability of the linear + interactions model to represent the effect of higher order terms.

4.7 Suitability of linear models to the engine noise problem

The main conclusions which may be drawn from the numerical tests which have been carried out using the engine noise example are as follows.

- The mass function is, as expected, exactly representable using simply mean and main effect terms.
- The use of probability plots provides a convenient method of identifying those terms of a linear + interactions model which are statistically significant.
- For the example investigated, the significant parameters have been independently shown to be the only ones which substantially influence the accuracy of the predictive model.
- Virtually all of the variation in the noise function at the test points is, for the present example, attributable to main effect and two-way interaction terms. The degree of similarity which exists between related applications suggests that, in general, little increase in lack of fit will result from the exclusion of interaction terms higher than second order.
- Such a model can be constructed from a fractional factorial design of at least resolution V. This also allows for the inclusion in the model of a number of higher order interaction terms.
- The use of a model of this type, whilst accounting for all variation in the noise function at the vertices of the design variable space, results in substantial lack of fit errors at locations which are distant from the original test points. In order to enhance the ability of the model to represent the true response it is necessary to include higher order terms. It is not, however, possible to estimate these terms using the two-level designs discussed above.

Appendix 4A

First order models constructed from full and fractional factorial designs
Specification of test points used to estimate lack of fit
due to higher order terms

Variable						
A	B	C	D	E	F	G
.0090	.0260	.0260	.0090	.0175	.0260	.0090
.0060	.0260	.0260	.0090	.0175	.0260	.0090
.0120	.0260	.0260	.0090	.0175	.0260	.0090
.0090	.0200	.0260	.0090	.0175	.0260	.0090
.0090	.0320	.0260	.0090	.0175	.0260	.0090
.0090	.0260	.0200	.0090	.0175	.0260	.0090
.0090	.0260	.0320	.0090	.0175	.0260	.0090
.0090	.0260	.0260	.0040	.0175	.0260	.0090
.0090	.0260	.0260	.0140	.0175	.0260	.0090
.0090	.0260	.0260	.0090	.0100	.0260	.0090
.0090	.0260	.0260	.0090	.0250	.0260	.0090
.0090	.0260	.0260	.0090	.0175	.0200	.0090
.0090	.0260	.0260	.0090	.0175	.0320	.0090
.0090	.0260	.0260	.0090	.0175	.0260	.0060
.0090	.0260	.0260	.0090	.0175	.0260	.0120

Appendix 4B

Results of numerical tests using fractional factorial designs

1.	MEAN	8 8082×10^{+1}	44.	CDEG	7.4247×10^{-3}	87.	BDG	-3 2154×10^{-3}
2.	A	-2 2377	45.	DEF	7.2312×10^{-3}	88.	ACE	3.1474×10^{-3}
3.	C	-3 8596×10^{-1}	46.	CDF	-7 2296×10^{-3}	89.	CEFG	3.0834×10^{-3}
4.	G	-3.3608×10^{-1}	47.	ABC	7.1612×10^{-3}	90	ADF	3.0631×10^{-3}
5.	D	-2.3021×10^{-1}	48.	CDG	7 0216×10^{-3}	91.	BEFG	-3 0467×10^{-3}
6.	F	-2.2872×10^{-1}	49.	ACDEG	-6 9616×10^{-3}	92.	ABDG	2.9337×10^{-3}
7.	B	-1.9920×10^{-1}	50.	BF	-6 9547×10^{-3}	93	ABDEFG	-2.9106×10^{-3}
8.	E	-1.5888×10^{-1}	51.	CDFG	-6 6769×10^{-3}	94.	ACEFG	-2 8335×10^{-3}
9.	AG	-7.4246×10^{-2}	52	ACDFG	6 6182×10^{-3}	95.	BDEFG	2 8164×10^{-3}
10	AD	5 2756×10^{-2}	53.	ABCG	5 9315×10^{-3}	96.	BC	-2.7048×10^{-3}
11.	AF	-4 6907×10^{-2}	54	ABE	-5 8429×10^{-3}	97.	ABDF	2.5961×10^{-3}
12.	CG	-4 4956×10^{-2}	55.	ABEG	-5.7443×10^{-3}	98	BCDFG	2.5133×10^{-3}
13.	AE	-4 4724×10^{-2}	56.	ABEF	5 4687×10^{-3}	99	ADEFG	-2.4929×10^{-3}
14	FG	-4 4068×10^{-2}	57.	ACEF	5 4456×10^{-3}	100.	BCDEG	2 4517×10^{-3}
15.	ACG	3.5056×10^{-2}	58.	ABDEG	5 4233×10^{-3}	101.	ABCDE	-2 4162×10^{-3}
16	AEF	1.9958×10^{-2}	59.	BDEG	-5 2643×10^{-3}	102.	ACDEF	2.3762×10^{-3}
17.	EF	-1.9304×10^{-2}	60	BE	5 0225×10^{-3}	103.	BCDG	2 2716×10^{-3}
18.	AC	1.7965×10^{-2}	61.	BCDF	-4 9464×10^{-3}	104.	BEF	-2.1783×10^{-3}
19.	BCF	1.6793×10^{-2}	62.	BCEG	-4 8705×10^{-3}	105.	AB	-2.1579×10^{-3}
20.	ABCF	-1.5657×10^{-2}	63.	BEG	4 7925×10^{-3}	106.	CF	-2.0599×10^{-3}
21.	ACF	-1.5540×10^{-2}	64.	BDF	-4.7858×10^{-3}	107.	ABFG	1 8338×10^{-3}
22.	CE	1 5170×10^{-2}	65.	AEFG	-4.7292×10^{-3}	108.	ABDE	1.7930×10^{-3}
23.	ABDFG	-1 4890×10^{-2}	66.	ABCEF	-4.6608×10^{-3}	109.	ABCDEG	-1.6555×10^{-3}
24.	EG	-1 4491×10^{-2}	67.	BCDE	4 6442×10^{-3}	110.	BCEFG	-1.5398×10^{-3}
25	BDFG	1 4209×10^{-2}	68.	BCFG	4 5903×10^{-3}	111.	DEFG	1.4303×10^{-3}
26	DFG	1.3226×10^{-2}	69.	CEF	-4 5796×10^{-3}	112.	AEG	-1.3100×10^{-3}
27.	ACDF	1.3167×10^{-2}	70.	BCD	4 5612×10^{-3}	113.	ABDEF	-1.2484×10^{-3}
28.	ADFG	-1.1497×10^{-2}	71.	ABCE	-4 5580×10^{-3}	114.	ABCDEF	1.2094×10^{-3}
29.	CD	1.1175×10^{-2}	72.	BCEF	4.5231×10^{-3}	115.	BDEF	1.0884×10^{-3}
30.	ADEF	-1.0824×10^{-2}	73.	ABCDF	4.5133×10^{-3}	116.	BG	-8.7908×10^{-4}
31	DG	-1 0766×10^{-2}	74	ABCDEFG	-4.3667×10^{-3}	117.	ACEG	-7.7062×10^{-4}
32	ACDG	-1 0610×10^{-2}	75.	ABCEG	4.1502×10^{-3}	118	AFG	-6 8496×10^{-4}
33	BFG	-1.0457×10^{-2}	76	ACD	-4 0621×10^{-3}	119.	ADEG	-6.3499×10^{-4}
34.	ABCFG	-1 0397×10^{-2}	77.	ACDE	-4 0350×10^{-3}	120.	EFG	6.1510×10^{-4}
35.	BD	9.3193×10^{-3}	78.	ABCD	-3.9413×10^{-3}	121.	ACFG	6.1398×10^{-4}
36.	ADG	9.2839×10^{-3}	79.	CDE	3.7317×10^{-3}	122.	ABCEFG	5.9950×10^{-4}
37.	CFG	-9 2465×10^{-3}	80.	ABCDG	-3.6975×10^{-3}	123.	CDEF	-5.7055×10^{-4}
38.	CDEFG	8 6995×10^{-3}	81.	ADE	-3.6911×10^{-3}	124.	BCE	5 4370×10^{-4}
39.	DE	8 3466×10^{-3}	82.	ACDEFG	-3.4924×10^{-3}	125.	ABEFG	-4.7396×10^{-4}
40.	DF	-8.0816×10^{-3}	83.	BCDEFG	3.4912×10^{-3}	126.	ABCDFG	-1.7400×10^{-4}
41.	ABF	-7 9288×10^{-3}	84.	DEG	-3 4698×10^{-3}	127.	BCG	-1.7328×10^{-4}
42.	ABD	7.9016×10^{-3}	85	BDE	3 3976×10^{-3}	128	CEG	-1 5289×10^{-4}
43.	ABG	-7 5974×10^{-3}	86.	BCDEF	-3.3055×10^{-3}			

Table 4B.1. Noise coefficients for full factorial test (n=7)

1. MEAN	88.078	23. BCF	1.6158×10^{-2}	45. BD	6.4857×10^{-3}
2. A	-2.2342	24. DG	-1.5427×10^{-2}	46. AEG	-6.2564×10^{-3}
3. C	-3.8887×10^{-1}	25. BFG	-1.4492×10^{-2}	47. CDE	5.5655×10^{-3}
4. G	-3.3487×10^{-1}	26. BF	-1.3916×10^{-2}	48. BDF	-5.5565×10^{-3}
5. F	-2.3038×10^{-1}	27. ADG	1.3807×10^{-2}	49. BC	-5.1977×10^{-3}
6. D	-2.2961×10^{-1}	28. DEF	1.3163×10^{-2}	50. BDE	4.0116×10^{-3}
7. B	-2.0269×10^{-1}	29. CDF	-1.2974×10^{-2}	51. AFG	3.9592×10^{-3}
8. E	-1.5905×10^{-1}	30. BEF	-1.2788×10^{-2}	52. DF	-3.9314×10^{-3}
9. AG	-7.7552×10^{-2}	31. ABE	-1.2520×10^{-2}	53. CF	3.3634×10^{-3}
10. AD	5.1217×10^{-2}	32. CDG	1.2490×10^{-2}	54. EFG	-3.3262×10^{-3}
11. FG	-4.6484×10^{-2}	33. BE	1.1641×10^{-2}	55. CEG	2.4432×10^{-3}
12. CG	-4.6205×10^{-2}	34. BCG	-1.0997×10^{-2}	56. BDG	2.2302×10^{-3}
13. AF	-4.4455×10^{-2}	35. ABD	1.0985×10^{-2}	57. DE	-2.0508×10^{-3}
14. AE	-4.2211×10^{-2}	36. BCE	-1.0953×10^{-2}	58. ADF	-1.8074×10^{-3}
15. ACG	3.6145×10^{-2}	37. CD	1.0701×10^{-2}	59. CEF	-1.6460×10^{-3}
16. EF	-2.3002×10^{-2}	38. EG	-9.9781×10^{-3}	60. BG	1.4971×10^{-3}
17. AEF	2.2230×10^{-2}	39. DFG	8.6681×10^{-3}	61. ADE	8.9918×10^{-4}
18. ACF	-2.0805×10^{-2}	40. ABC	8.5915×10^{-3}	62. ABF	-5.0410×10^{-4}
19. AC	2.0781×10^{-2}	41. ABG	-8.1680×10^{-3}	63. CE	2.8026×10^{-4}
20. DEG	-1.9127×10^{-2}	42. CFG	-7.4534×10^{-3}	64. BCD	-1.6804×10^{-4}
21. BEG	1.7959×10^{-2}	43. ACD	-7.1088×10^{-3}		
22. ACE	1.7357×10^{-2}	44. AB	6.5416×10^{-3}		

Table 4B.2. Parameter estimates for 2^{7-1} fractional factorial design of resolution VII

MEAN	ABCD	CDEF	ABEF
A	BCD	ACDEF	BEF
B	ACD	BCDEF	AEF
C	ABD	DEF	ABCEF
D	ABC	CEF	ABDEF
E	ABCDE	CDF	ABF
F	ABCDF	CDE	ABE
G	ABCDG	CDEFG	ABEFG
AB	CD	ABCDEF	EF
AC	BD	ADEF	BCEF
AD	BC	ACEF	BDEF
AE	BCDE	ACDF	BF
AF	BCDF	ACDE	BE
AG	BCDG	ACDEFG	BEFG
BG	ACDG	BCDEFG	AEFG
CE	ABDE	DF	ABCF
CF	ABDF	DE	ABCE
CG	ABDG	DEFG	ABCEFG
DG	ABCG	CEFG	ABDEFG
EG	ABCDEG	CDFG	ABFG
FG	ABCDFG	CDEG	ABEG
ABG	CDG	ABCDEFG	EFG
ACE	BDE	ADF	BCF
ACF	BDF	ADE	BCE
ACG	BDG	ADEFG	BCEFG
ADG	BCG	ACEFG	BDEFG
AEG	BCDEG	ACDFG	BFG
AFG	BCDFG	ACDEG	BEG
CEG	ABDEG	DFG	ABCFG
CFG	ABDFG	DEG	ABCEG
ACEG	BDEG	ADFG	BCFG
ACFG	BDFG	ADEG	BCEG

Table 4B.3. Confounding between parameters for 2^{7-2} fractional factorial design of resolution IV

MEAN	ABCD	CDEF	ABEF	ACFG	BDFG	ADEG	BCEG
A	BCD	ACDEF	BEF	CFG	ABDFG	DEG	ABCEG
B	ACD	BCDEF	AEF	ABCFG	DFG	ABDEG	CEG
C	ABD	DEF	ABCEF	AFG	BCDFG	ACDEG	BEG
D	ABC	CEF	ABDEF	ACDFG	BFG	AEG	BCDEG
E	ABCDE	CDF	ABF	ACEFG	BDEFG	ADG	BCG
F	ABCDF	CDE	ABE	ACG	BDG	ADEFG	BCEFG
G	ABCDG	CDEFG	ABEFG	ACF	BDF	ADE	BCE
AB	CD	ABCDEF	EF	BCFG	ADFG	BDEG	ACEG
AC	BD	ADEF	BCEF	FG	ABCDFG	CDEG	ABEG
AD	BC	ACEF	BDEF	CDFG	ABFG	EG	ABCDEG
AE	BCDE	ACDF	BF	CEFG	ABDEFG	DG	ABCG
AF	BCDF	ACDE	BE	CG	ABDG	DEFG	ABCEFG
AG	BCDG	ACDEFG	BEFG	CF	ABDF	DE	ABCE
BG	ACDG	BCDEFG	AEFG	ABCF	DF	ABDE	CE
ABG	CDG	ABCDEFG	EFG	BCF	ADF	BDE	ACE

Table 4B.4. Confounding between parameters for 2^{7-3} fractional factorial design of resolution IV

Appendix 4C

4C.1 The use of normal probability plots

Normal probability plots are used within the present work in order to assess whether the coefficients of predictive models are statistically significant, and hence represent true characteristics of the response surface, or whether they are indistinguishable from a set of normally distributed random data. Details of the use of this technique is described by, for example, Box, Hunter and Hunter (1978), and Box and Draper (1987). The following is a summary of the underlying theory, with some discussion of its applicability to the engine noise optimization problem Details of the implementation of this technique within the optimization program are given by Hall (1992).

Normally distributed data x which have a mean of μ and a variance of σ^2 have a probability density function of the form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \qquad (4C.1)$$

This function is shown in Figure 4C.1. The cumulative density function, F(x), is the shaded area beneath the graph from $-\infty$ to x, and represents the probability that a given measurement point will have a value less than or equal to x

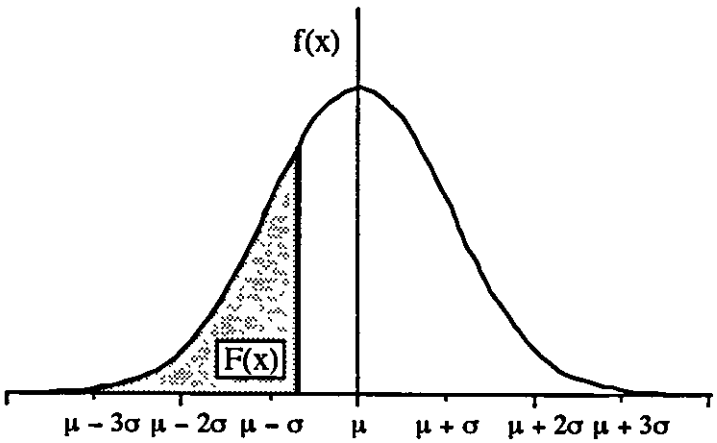

Figure 4C.1 Normal probability density function

The cumulative density function (c d.f) is as shown in equation (4C.2). When plotted, this function forms the 'S' shaped curve shown in Figure 4C 2.

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/(2\sigma^2)}\, dt \qquad (4C.2)$$

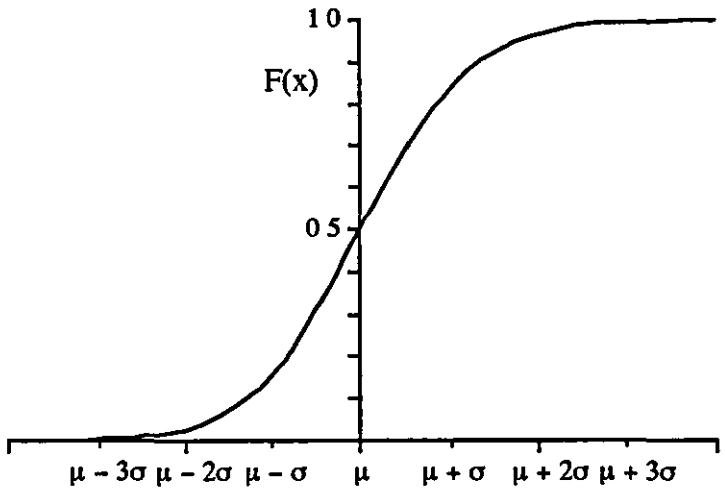

Figure 4C.2 Cumulative density function

It is this graph which forms the basis of the probability plot method. If a sample of normally distributed data is displayed as in Figure 4C.2, such that the rank order of each measurement, scaled to lie in the range 0 0 to 1 0, is plotted against its value, then all points will lie approximately on a c d f curve having the appropriate variance. When plotting a set of random data which is drawn from a normally distributed population, any deviation from this line is due to sampling error. If a sample of data to be investigated is plotted, then the degree to which the sample follows a normal distribution may be judged by the proximity of the points to the appropriate c d f. curve. Two main problems which arise when using the c d f. curve are that the variance of the sample data may not be known prior to plotting, and that comparison of the data with the 'S' shaped curve of Figure 4C 2 is difficult to carry out.

In order to facilitate this comparison, the probability plot method modifies the vertical scale of the c d f , either by plotting on special normal probability paper, or by carrying out a normalisation of the rank orders. For implementation within a computing environment, it is the second of these two approaches which has been followed within the present work. The effect of carrying out this modification of the vertical axis is that the 'S' shaped curve of Figure 4C 2 is stretched into a straight line. This aids interpretation of the c d f. graph, since a sample of random data can be judged to have a normal distribution if it lies on any straight line, with the slope of the line being determined by the variance of the distribution.

If the data is found to be normally distributed, then both the mean and variance of the sample may be deduced directly from the probability plot as follows.

i) The mean of the sample is the x value associated with a c.d.f. $F(x) = 0.5$.

ii) The value of the c.d.f. for the data point $x = \mu + \sigma$ is $F(x) = 0.8314$, and hence the value of σ^2 is equal to $(x_{[F(x)=0\ 8314]} - x_{[F(x)=0\ 5]})^2$.

In using the method to identify which coefficients of a predictive model represent real effects, it has been convenient to invert the probability plot such that the coefficient magnitudes are plotted against their normalised rank. The normalisation of rank order has been carried out using a rational approximation to the inverse c d f., to give a 'Z score'. This is performed in such a way that the normalised rank for the mean value is 0 0, rather than 0 5. Additionally, in comparing coefficient estimates of predictive models, it is essential to compare values which have been estimated with equal precision. In order to achieve this, the coefficient values have each been divided by their standard error prior to plotting.

Figures 4C 3 – 4C 5 show examples of normal probability plots of the form used within the present investigation Each of the three data sets shown is drawn from a normally distributed population with zero mean,

and in each case it can be seen that all of the data points lie on an approximately straight line. Note that the effect of sampling error increases as the size of the sample is reduced, so that the straightened c d f. line is less well defined, making it harder to identify values which do not conform to the normal distribution.

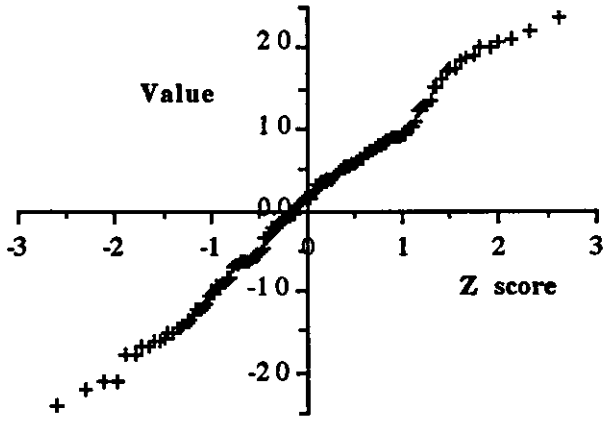

Figure 4C.3 Normal probability plot of a random sample of 160 data points drawn from a normally distributed population

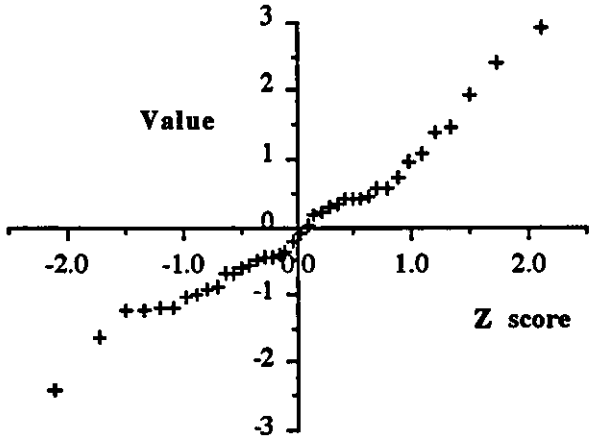

Figure 4C.4 Normal probability plot of a random sample of 40 data points drawn from a normally distributed population

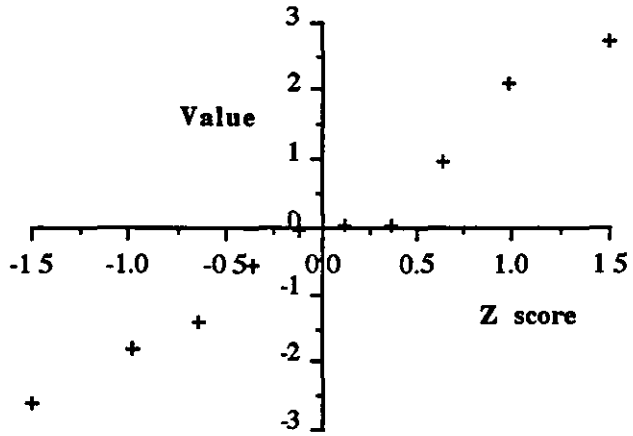

Figure 4C.5 Normal probability plot of a random sample of 10 data points drawn from a normally distributed population

When investigating the statistical significance of the coefficient value of a response surface model, these are plotted as shown above. Any which lies significantly away from the straight line formed by the majority of the points is considered inconsistent with an assumption of normal distribution, and is taken to represent a true characteristic of the data set from which it is derived. Such values must be removed from the graph, and the remaining values replotted using recalculated Z scores in order to establish whether all the remaining coefficients conform to an assumption of normal distribution Figure 4C.6 shows an example of a data set in which a number of the values deviate substantially from the normal distribution line; five of them negative and two positive.

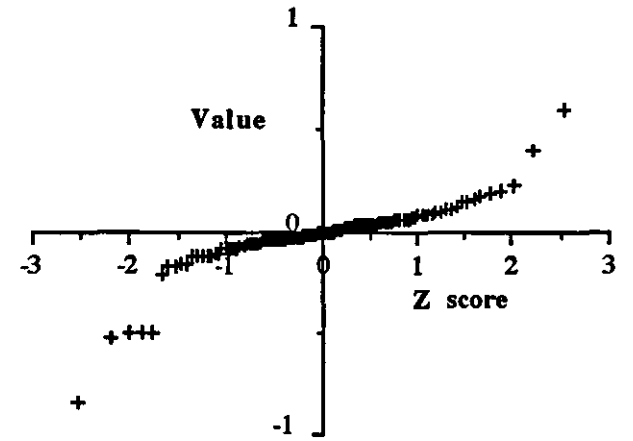

Figure 4C.6 Normal probability plot of a normally distributed sample with seven outliers

For analyses in which it is only the magnitudes of the data points which are of interest, a modified version of the normal probability plot can be used. This variation was introduced by Daniel (1959), and is known as the half-normal probability plot, or Daniel plot Using this technique, the magnitude of each data point is again plotted against its normalised rank, but with the normalisation carried out on the assumption that the sample has a single-sided normal distribution with zero mean. Figure 4C.7 shows the data of Figure 4C 6, displayed using a Daniel plot.

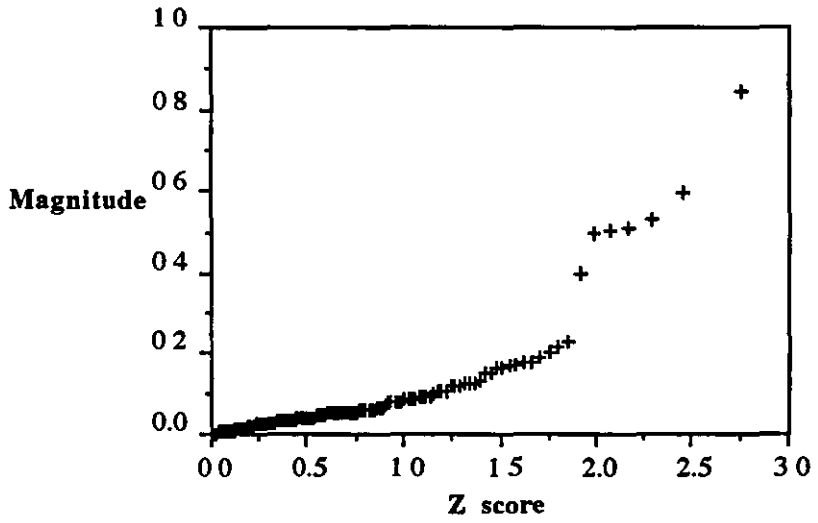

Figure 4C.7 Half-normal probability plot of a normally distributed sample with seven outliers

4C.2 Treatment of random experimental error in deterministic applications

The use of the probability plot is intended to provide a tool to aid the investigator in differentiating between those effects which are due to random experimental error, and those which are due to the underlying characteristics of the system being investigated When using computer simulation methods to provide the function information from which the model is to be constructed, however, no random experimental error can occur, since a given combination of input variables will always produce identical results The only possible cause of lack of fit between the values of the response functions generated by the simulation program and the mathematical model which is derived from the original sample points is thus underspecification of the fitted model. The validity of the probability plot technique when applied to such deterministic applications clearly rests upon the assumption that the effect of model underspecification introduces errors in the parameter estimates which may, in some sense, be regarded as being randomly distributed

Although a detailed investigation into the validity of this assumption is beyond the scope of the present work, a useful test of its applicability is to examine the nature of the discrepancy between the FE analysis result and the response surface prediction at each test point of the experimental design. Two tests which may be performed on these residuals are:

i) Plot residual values against the value of each of the design variables to assess correlation with variable value

ii) Construct a probability plot of the residuals to assess whether these error terms are normally distributed.

In order to derive meaningful information from these tests, the design used must be of a non-saturated nature, since otherwise all residuals will be due purely to rounding error The saturated factorial designs presented in Chapter 4 are thus unsuited to this type of investigation. Results are presented below for the 7 variable Central Composite Design of Section 6 5.3, comprising 79 tests, to which a 36 term strict second-order model is fitted. In Figures 4C 8 — 4C 14 the residuals are plotted against each of the variable values, and these show that there is no noticeable correlation between the lack-of-fit values and the levels of any of the design variables.

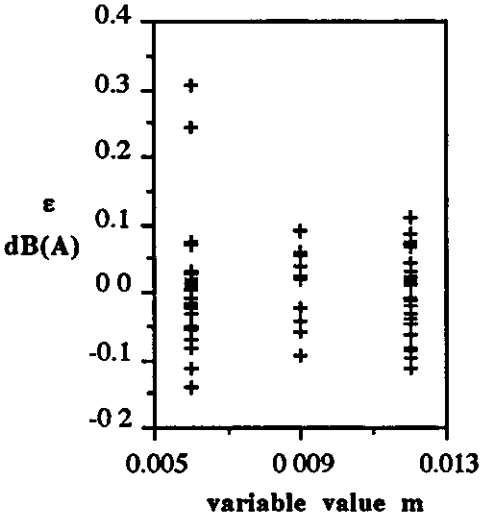

Fig. 4C.8 Residuals against variable 1

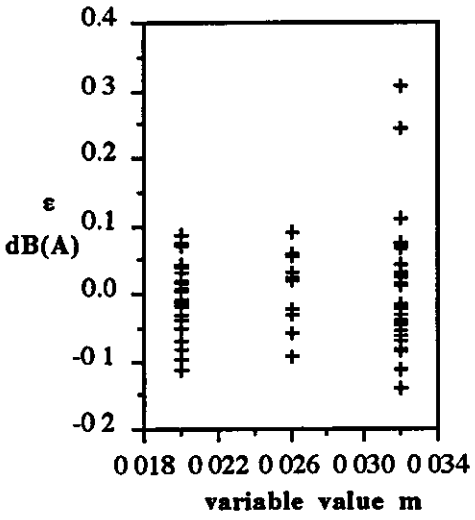

Fig. 4C.9 Residuals against variable 2

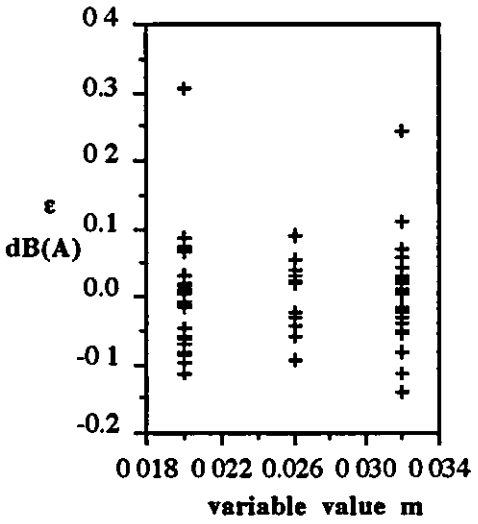

Fig. 4C.10 Residuals against variable 3

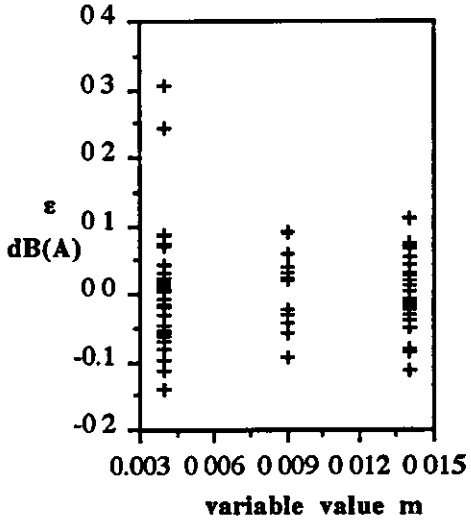

Fig. 4C.11 Residuals against variable 4

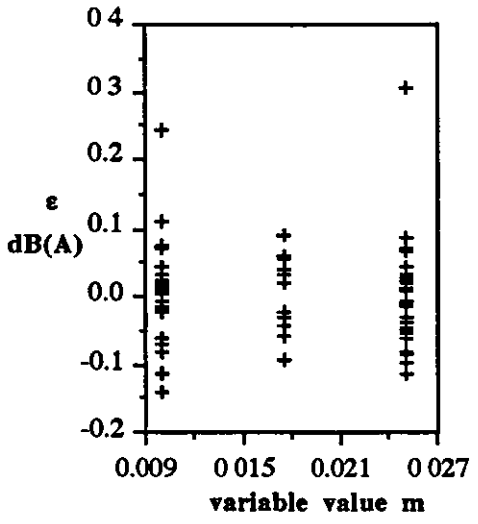

Fig. 4C.12 Residuals against variable 5

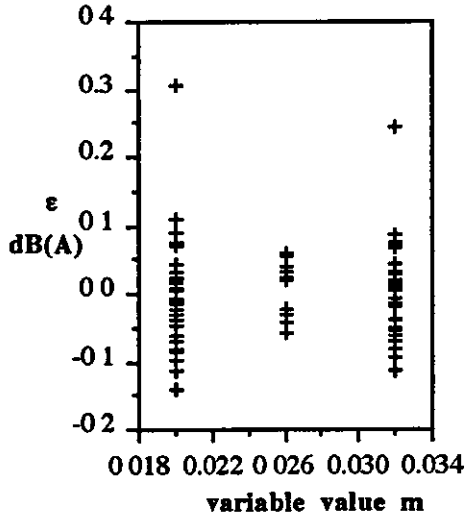

Fig. 4C.13 Residuals against variable 6

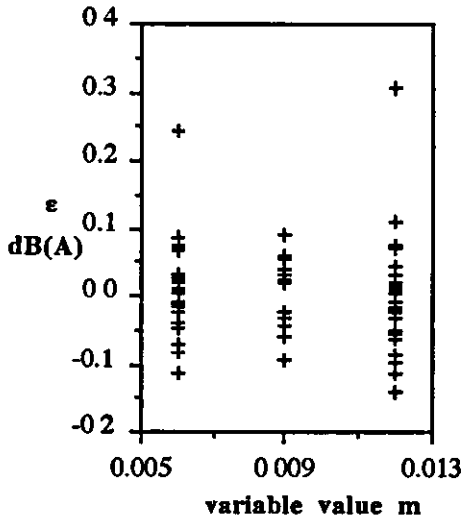

Fig. 4C.14 Residuals against variable 7

Figure 4C.15 shows a normal probability plot of the residuals, which may be used to assess whether the residual values are drawn from a normally distributed population. This plot shows that, apart from two clear outliers, the residuals lie on an approximately straight line, indicating that the errors are indeed normally distributed Additionally, this straight line passes approximately through the origin of the plot axis, suggesting that the mean of the error values is approximately zero.

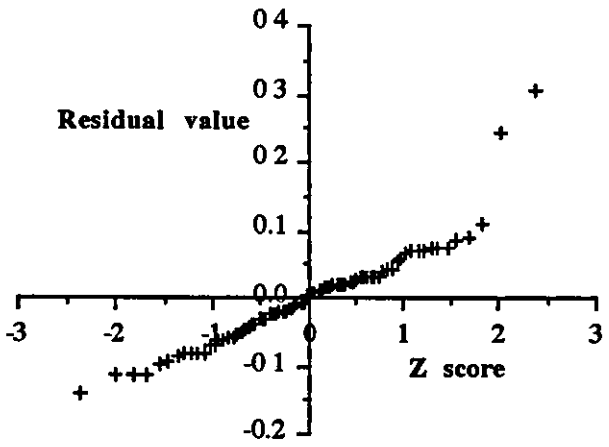

Figure 4C.15 Normal probability plot of residuals

In conclusion, investigation of the residuals at the test points of a non-saturated experimental design shows that these values are uncorrelated with any of the variable levels, and that they are randomly distributed with a zero mean These results suggest that there is some validity in assuming that the effect of model underspecification introduces errors which may be treated similarly to random experimental error. Bearing in mind the restrictions imposed upon the use of a probability plot analysis by the deterministic nature of the noise simulation process, experience has shown that an informed use of this technique provides an extremely useful tool in assessing the parameters of a multi-dimensional mathematical model.

5. Designs for quadratic models

5.1 Introduction

When an analysis of the type carried out in the previous chapter, using just a linear + interactions model, has shown that there is substantial lack of fit at points away from the corners of the design variable space, due to curvature of the actual response surface, then the next step in an investigation is to include terms in the prediction model which will account for these effects.

Such terms include those which involve higher powers of each variable, and the simplest extension to the original linear model of equation (4.1) is the quadratic model

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n} \sum_{j=i}^{n} \beta_{ij} X_i X_j \qquad (5.1)$$

Analysis of the simple one-dimensional case shows that there are now three free parameters in the equation;

$$Y = \beta_0 + \beta_1 X_1 + \beta_{11} X_1^2 \qquad (5.2)$$

and hence at least three test points are required in order to fit the model. This feature generalises to multiple dimensions as in Chapter 4, such that each variable must appear in the experimental design matrix at at least three different levels.

As for the first order designs of Chapter 4, the values of the independent variables are normalised using the convention of Section 2.12. The second-order strict quadratic model can be expressed in these normalised coordinates as

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j=i}^{n} \beta_{ij} x_i x_j \qquad (5.3)$$

Before considering specific designs for fitting second order models, the following two sections discuss the characteristics which a second order design must possess in order to meet the requirements of orthogonality and rotatability.

5.2 Orthogonal second-order designs

As stated in Appendix 2A, the condition which must be met for a design to be orthogonal is that the moment matrix $N^{-1}\mathbf{X'X}$ must be diagonal. Considering a general second-order design for the simple case of n=2, the moment matrix, using the notation of Section 2.12, is of the form

$$\frac{1}{N}\mathbf{X'X} = \begin{array}{c} \\ 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_2^2 \\ x_1x_2 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \end{array} \\ \left[\begin{array}{cccccc} 1 & [1] & [2] & [11] & [22] & [12] \\ & [11] & [12] & [111] & [122] & [112] \\ & & [22] & [112] & [222] & [122] \\ & & & [1111] & [1122] & [1112] \\ & \text{symmetric} & & & [2222] & [1222] \\ & & & & & [1122] \end{array}\right] \end{array} \qquad (5.4)$$

for which all of the off-diagonal terms must be equal to zero if the design is to be orthogonal. It can be seen from equation (5.4) that these off-diagonal elements include terms of the form [ii] and [iijj], which are the sums of the products $1.x_i^2$ and $x_i^2.x_j^2$ respectively. Since each of the columns x_i and x_j must necessarily contain non-zero elements, and since each element of x_i^2 and x_j^2 must be greater than or equal to zero, each of the [ii] and [iijj] must be non-zero, so that a diagonal moment matrix is impossible to obtain.

An alternative approach to second-order orthogonality was derived by Box and Hunter (1957), who first defined a set of orthogonal polynomials of degree m in each of the variables x_i (i=1, ..., n), such that

$$x_1^{(m)} = x_1^m + \alpha_{m-1,m}x_i^{m-1} + \alpha_{m-2,m}x_i^{m-2} + \ldots + \alpha_{1,m}x_i + \alpha_{0,m} \qquad (5.5)$$

with each value of α chosen such that

$$\sum_{u=1}^{N} x_{iu}^{(m)}x_{iu}^{(m-p)} = 0, \qquad p = 1, 2, \ldots, m \qquad (5.6)$$

The original equations $\mathbf{Y} = \mathbf{X\beta}$ can then be expressed in terms of these polynomials as

$$\mathbf{Y} = (\mathbf{XP})\,(\mathbf{P}^{-1}\boldsymbol{\beta}) = \dot{\mathbf{X}}\dot{\boldsymbol{\beta}} \qquad (5.7)$$

where $\mathbf{P}$ is the transformation matrix which maps the original independent variables to the new set.

A design is now orthogonal if

$$N^{-1}\dot{\mathbf{X}}'\dot{\mathbf{X}}$$

is diagonal. The requirements for this are that, for $i < j < p < q$,

$$[ij] = [ijj] = [iij] = 0$$
$$[iijj] = 1$$
$$[iijp] = [ijjp] = [ijpp] = 0$$
$$[iij] = [ijj] = 0$$
$$[ijpq] = [ijp] = 0$$

Appendix 5A demonstrates the orthogonality of a three level factorial design of the type described in Section 5.4 below.

5.3 Rotatability in second-order designs

Referring to the conditions for rotatability of Section 2.12, a second-order design must satisfy

$$[ii] = \lambda_2$$
$$[iijj] = \lambda_4 \qquad \text{for } i, j = 1, 2, \ldots, n \text{ and } i < j$$
$$[iiii] = 3\lambda_4$$

where the required value of $\lambda_2 = 1$ is fixed as a consequence of the scaling convention. This leaves only λ_4 to be determined, which may be selected in order to achieve other design criteria. If the design is also to be orthogonal, for example, then Section 5.2 shows that a value of $\lambda_4 = [iijj] = 1$ is required.

However, an analysis by Box and Hunter (1957), showed that, for a rotatable design, the prediction variance (see Appendix 2A) at a point $\mathbf{x}$ at radius ρ from the design centre, such that $\rho^2 = \mathbf{x}'\mathbf{x}$, is

$$\text{Var}[y(\mathbf{x})] = A\{ 2(n+2)\lambda_4^2 + 2\lambda_4(\lambda_4-1)(n+2)\rho^2 + [(n+1)\lambda_4 - (n-1)]\rho^4\} \qquad (5.8)$$

where $$A = \sigma^2\{2N\lambda_4[(n+2)\lambda_4 - n]\}^{-1}$$

For the orthogonal case of $\lambda_4 = 1$, the change of variance with radius ρ is very marked, with a much higher value being obtained at $\rho = 1$ than at the design centre $\rho = 0$. Since the

investigator would ideally like to be able to predict the response at all points within the design variable space with equal precision, this is a very worrying characteristic. Plotting the variance against ρ for a number of different values of λ_4, it is found that at a certain value of λ_4 the variance at $\rho = 1$ is equal to that at $\rho = 0$, and Box and Hunter suggest that the orthogonality requirement be relaxed slightly so as to allow the value of λ_4 to be chosen to meet this criterion. A design which satisfies this requirement is known as a *uniform precision design.*

The requirement, then, is that the value of variance evaluated from equation (5.8) should be the same for $\rho = 0$ and $\rho = 1$. From (5.8), it can be seen that the value of both A and the term $2(n+2)\lambda_4^2$ are invariant with ρ, so that the criterion is met if the expression

$$2\lambda_4(\lambda_4-1)(n+2)\rho^2 + [(n+1)\lambda_4 - (n-1)]\rho^4 \qquad (5.9)$$

is equal at values of $\rho = 0$ and $\rho = 1$.

Since the above expression is equal to zero when $\rho = 0$, uniform precision is achieved if, at $\rho = 1$,

$$2\lambda_4(\lambda_4-1)(n+2) + (n+1)\lambda_4 - (n-1) = 0 \qquad (5.10)$$

and hence

$$\lambda_4 = \frac{(n+3) \pm \sqrt{9n^2 + 14n - 7)}}{4n + 8} \qquad (5.11)$$

Evaluating this expression for various values of n gives

n	λ_4
1	0.6666
2	0.7844
3	0.8385
4	0.8704
5	0.8918
6	0.9070
7	0.9184
8	0.9274
9	0.9346
10	0.9404
11	0.9453
12	0.9495

showing that the uniform precision design is almost orthogonal for all but small n, and grows more so as the number of variables increases. It is interesting to note that if equation (5.10) is

evaluated for $\lambda_4 = 1$, then

$$(n+1) - (n-1) = 0$$

which is only approximated as $n \rightarrow \infty$, demonstrating that a second order design cannot be both exactly orthogonal and rotatable with uniform precision.

5.4 The three level factorial design

A simple design which allows the estimation of the required terms in equation (5.1) is the three level factorial experiment. This is constructed on the model of the 2^n full factorial design of Section 3.3, with the design matrix consisting of every combination of the three levels of each variable, giving a total of 3^n points. Figure 5.1 shows the test points which are required in the three-dimensional case.

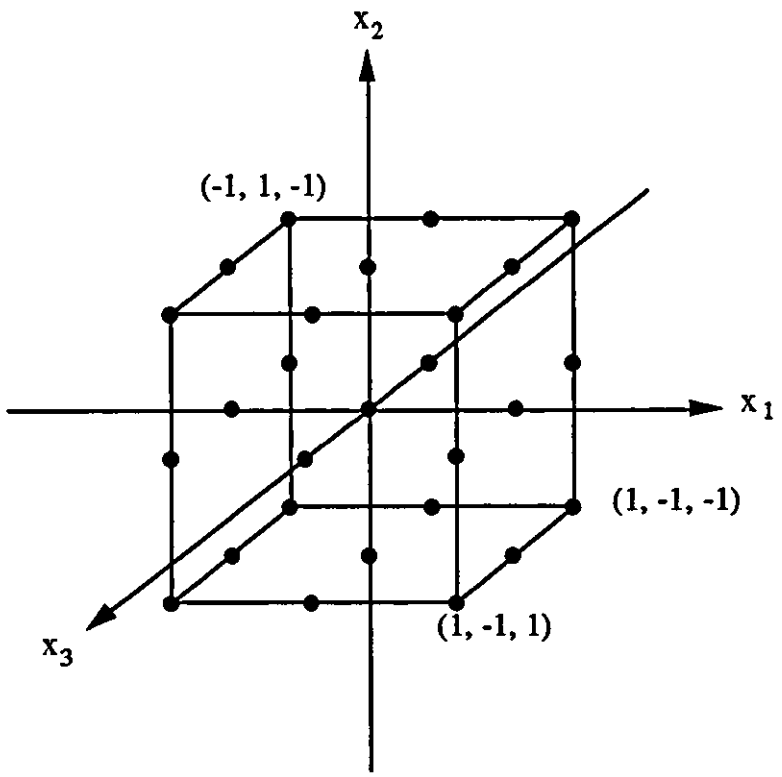

Figure 5.1. A three level factorial design in three dimensions

for which the associated design matrix is

$$D = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \\ \begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & 0 \\ -1 & -1 & 1 \\ -1 & 0 & -1 \\ -1 & 0 & 0 \\ -1 & 0 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & 0 \\ -1 & 1 & 1 \\ 0 & -1 & -1 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 0 \\ 1 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \end{array}$$

Since there are 3^n tests, it is in fact possible to fit a total of 3^n terms, allowing the construction of a second order plus interactions model of the form

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n-1} \sum_{j=i}^{n} \sum_{\substack{k=j \\ k>i}}^{n} \beta_{ijk} x_i x_j x_k + \sum_{i=1}^{n-1} \sum_{j=i}^{n-1} \sum_{\substack{k=j \\ k>i}}^{n} \sum_{\substack{l=k \\ l>j}}^{n} \beta_{ijkl} x_i x_j x_k x_l + \cdots\cdots \beta_{112233\ldots nn} x_1^2 x_2^2 x_3^2 x_4^2 \cdots x_n^2 \qquad (5.12)$$

in which no more than two of the subscripts in any term may take the same value. An example of the complete set of terms generated by this equation is given in Appendix 5B for the four variable case. The example of Appendix 5A shows that a 3^n design is orthogonal. It also

shows, however, that the rotatability requirement of [iiii] = 3[iijj] is not met. This is because, in this case, [iijj] = 1 (a necessary condition for orthogonality) whilst [iii] = 3/2.

The principal disadvantage of the three level factorial design, however, is that the number of test points rises rapidly with increasing n, such that for 5 variables, $3^5 = 243$ tests are required; for 7 variables, 2187; and for 12 variables over 530,000 tests are needed. Although it is possible to construct fractional factorial designs containing 3^{n-m} terms, by a procedure analogous to that followed for two level testing, even quite small fractions are still prohibitively large when the original full factorial array is of this size. As an example, the smallest fraction of the 7 variable factorial test which will allow all of the two-way linear × linear interactions to be estimated without mutual confounding is the 3^{7-2} design, a 1/9 fraction, which still contains 243 tests (National Bureau of Standards, 1959). The simple three level factorial test is thus impractical unless either the number of factors involved is very small, or the cost of performing each test is extremely low.

Additionally, analysis of equation (5.1) shows that the strict quadratic model contains just (k+1)(k+2)/2 terms. For a seven variable example just 36 terms are involved, whilst the model of equation (5.12) contains 2187 terms, with the additional 2151 terms representing interaction effects between the linear and quadratic components. It has been shown in the previous chapter, however, that even interactions of order greater than two between linear terms have been found to be insignificant for the engine noise application, and it is therefore unlikely that many of the additional interactions in the present model will have a measurable effect. In cases where the coefficients of the strict quadratic model are the only significant ones, the ratio of significant terms to the number of tests carried out in a seven variable three-level factorial design is therefore likely to be around 1.67%, improving only to around 14.8% for the 1/9 fraction. This *significance ratio* is reduced if a larger number of variables is to be investigated, such that for a 12 variable example the full model contains fewer than 0.02% of significant terms.

A similar view of the efficiency of a design for fitting a particular model was taken by Box and Behnken (1960), who defined a *redundancy factor* as the ratio of the number of tests carried out to the number of coefficients to be fitted. This factor is the inverse of the *saturation ratio* defined in Section 2.6, and, when only significant terms are fitted, is also the inverse of the *significance ratio* described above. They commented that "in situations in which the experimental error variance is not so large as to require large numbers of observations to obtain necessary precision, designs having small redundancy factors are desirable", suggesting further that a value of two or less (more that 50% saturation) would be appropriate. When it is borne in mind that in the present application no random experimental error is experienced, it is clear that the unmodified three level full and fractional factorial designs represent an extreme over-testing of the region of interest, when compared with the model to be fitted. For this reason the simple three level factorial design will not be considered further within the present work.

Appendix 5A

Orthogonality and rotatability characteristics of the three level factorial design

The design matrix of a three level factorial design in two variables, with variable levels coded as -1, 0 and +1, is

$$\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

Normalising using the convention of Section 2 12,

$$x_{ui} = \frac{X_{ui}}{\left[\frac{1}{N}\sum_{u=1}^{N} X_{ui}^2\right]^{1/2}}$$

The denominator of this expression is calculated as

$$\left[\frac{1}{N}\sum_{u=1}^{N} X_{ui}^2\right]^{1/2} = \left[\frac{1}{9}\{3.(-1)^2 + 3.(0)^2 + 3.(1)^2\}\right]^{1/2} = \sqrt{\frac{6}{9}}$$

$$\therefore x_{ui} = g\,X_{ui} \qquad \text{where} \quad g = \frac{3}{\sqrt{6}}$$

The normalised regressor matrix is thus

$$\mathbf{X} = \begin{array}{c} \begin{matrix} 1 & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \end{matrix} \\ \begin{bmatrix} 1 & -g & -g & g^2 & g^2 & g^2 \\ 1 & -g & 0 & g^2 & 0 & 0 \\ 1 & -g & g & g^2 & g^2 & g^2 \\ 1 & 0 & -g & 0 & g^2 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & g & 0 & g^2 & 0 \\ 1 & g & -g & g^2 & g^2 & g^2 \\ 1 & g & 0 & g^2 & 0 & 0 \\ 1 & g & g & g^2 & g^2 & g^2 \end{bmatrix} \end{array}$$

The moment matrix for this design is

$$\frac{1}{N}\mathbf{X'X} = \begin{array}{c} \\ 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_2^2 \\ x_1x_2 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & \frac{2}{3}g^2 & \frac{2}{3}g^2 & 0 \\ & \frac{2}{3}g^2 & 0 & 0 & 0 & 0 \\ & & \frac{2}{3}g^2 & 0 & 0 & 0 \\ \text{symmetric} & & & \frac{2}{3}g^4 & \frac{4}{9}g^4 & 0 \\ & & & & \frac{2}{3}g^4 & 0 \\ & & & & & \frac{4}{9}g^4 \end{bmatrix} \end{array}$$

from which it can be seen that

$$[ii] = \frac{2}{3}g^2 = \frac{2}{3}\left[\frac{3}{\sqrt{6}}\right]^2 = \frac{2}{3}\cdot\frac{9}{6} = 1$$

$$[iiii] = \frac{6}{9}g^4 = \frac{6}{9}\left[\frac{3}{\sqrt{6}}\right]^4 = \frac{6}{9}\cdot\frac{81}{36} = \frac{3}{2}$$

$$[iijj] = \frac{4}{9}g^4 = \frac{4}{9}\left[\frac{3}{\sqrt{6}}\right]^4 = \frac{4}{9}\cdot\frac{81}{36} = 1$$

Thus this design meets the orthogonality criterion of [iijj] = 1, but not the rotatability criterion of [iiii] = 3[iijj]

Appendix 5B

Complete set of terms for a three level full factorial test in 4 variables

1. A
2. B
3. C
4. D
5. A^2
6. AB
7. AC
8. AD
9. B^2
10. BC
11. BD
12. C^2
13. CD
14. D^2
15. A^2B
16. A^2C
17. A^2D
18. AB^2
19. ABC
20. ABD
21. AC^2
22. ACD
23. AD^2
24. B^2C
25. B^2D
26. BC^2
27. BCD
28. BD^2
29. C^2D
30. CD^2
31. A^2B^2
32. A^2BC
33. A^2BD
34. A^2C^2
35. A^2CD
36. A^2D^2
37. AB^2C
38. AB^2D
39. ABC^2
40. $ABCD$
41. ABD^2
42. AC^2D
43. ACD^2
44. B^2C^2
45. B^2CD
46. B^2D^2
47. BC^2D
48. BCD^2
49. C^2D^2
50. A^2B^2C
51. A^2B^2D
52. A^2BC^2
53. A^2BCD
54. A^2BD^2
55. A^2C^2D
56. A^2CD^2
57. AB^2C^2
58. AB^2CD
59. AB^2D^2
60. ABC^2D
61. $ABCD^2$
62. AC^2D^2
63. B^2C^2D
64. B^2CD^2
65. BC^2D^2
66. $A^2B^2C^2$
67. A^2B^2CD
68. $A^2B^2D^2$
69. A^2BC^2D
70. A^2BCD^2
71. $A^2C^2D^2$
72. AB^2C^2D
73. AB^2CD^2
74. ABC^2D^2
75. $B^2C^2D^2$
76. $A^2B^2C^2D$
77. $A^2B^2CD^2$
78. $A^2BC^2D^2$
79. $AB^2C^2D^2$
80. $A^2B^2C^2D^2$

plus process mean

The first fourteen terms plus the mean are those which appear in the strict quadratic model of equation 5.1.

6. The Central Composite Design (CCD)

Perhaps one of the most widely established of the second-order designs is the Central Composite Design, or CCD, introduced by Box and Wilson (1951) as a less costly alternative to the 3^n factorial design. This design is formed by adding to a two-level factorial or fractional factorial design a small number of tests at a third level of each variable. The CCD lends itself to sequential investigation of a design variable space, since the results of a previous two-level experiment can be incorporated into the second-order design.

The main purpose of the present chapter is to establish whether the CCD is suitable as a design for computer experiments, and in particular whether it can successfully be used to estimate the coefficients of radiated noise response surfaces. Before carrying out numerical trials using the CCD, however, the standard orthogonality, rotatability and uniform precision criteria which apply when investigating a general experimental region are presented in Section 6.3. An understanding of these criteria is extremely important to the successful use of the CCD, since it is these requirements which determine the values of the various parameters of the CCD which are chosen for a particular experiment. Consideration of these parameters motivates the discussion, in Section 6.4, of the constraints which the computer simulation environment imposes on the use of the CCD. A number of ways of addressing these limitations is considered. In Section 6.5 the results of numerical tests using appropriately selected designs are presented, from which the suitability of the CCD to the present application is assessed (Section 6.6).

Chapter 7 describes a new method of extending the CCD, which has been developed as part of the present work in order to address the problem of executing replicate tests when investigating deterministic systems.

6.1 Introduction to the CCD

The standard CCD, as used for general experimental applications, is made up of three separate portions, as follows;

1. *A factorial portion*, consisting of a full or fractional two level factorial design. Selection of the appropriate fraction proceeds in the same way as for the first-order designs of Chapter 3, with the requirement often being that this portion should be of resolution V or higher. The variable values are again encoded to ±1, as described in Section 3.3. In the analyses which follow, the number of tests in this factorial portion will be denoted by F.

2. *An axial portion*, consisting of a total of 2n points, each of which lies on the axis of one of the design variables, at a distance α from the design centre. The points have the specification $(\pm\alpha, 0, 0, \ldots, 0)$, $(0, \pm\alpha, 0, \ldots, 0)$,...., $(0, 0, \ldots, 0, \pm\alpha, 0)$, $(0, 0, \ldots, 0, \pm\alpha)$, and form the vertices of a cross-polytope or "star". The choice of the parameter α is discussed in Section 6.3, below.

3. *Centre points*. A certain number n_0 of replicate tests performed at the centre of the design space (0, 0, ... , 0). The choice of n_0 is discussed in Section 6.3, below.

An example of a Central Composite Design in three dimensions with a full 2^3 factorial portion is shown in Figure 6.1.

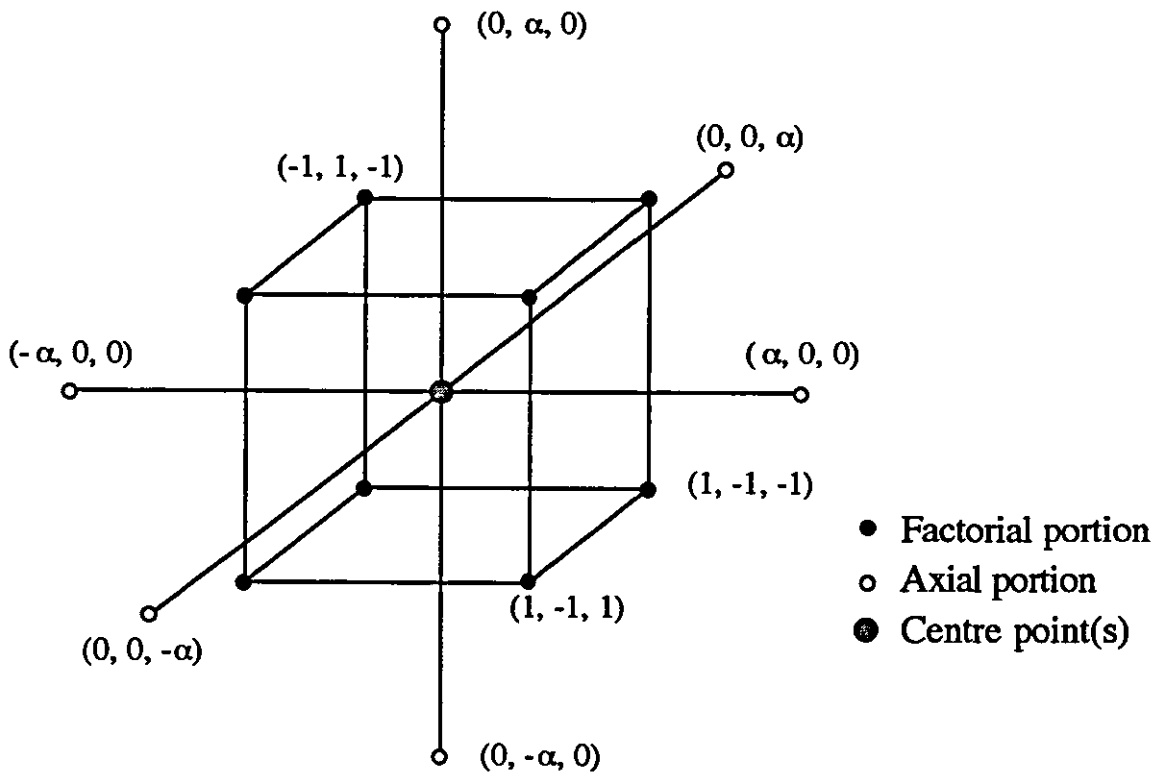

Figure 6.1 A CCD in three dimensions

The design matrix in this case is of the form

$$D = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & +1 \\ -1 & +1 & -1 \\ -1 & +1 & +1 \\ +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \\ -\alpha & 0 & 0 \\ +\alpha & 0 & 0 \\ 0 & -\alpha & 0 \\ 0 & +\alpha & 0 \\ 0 & 0 & -\alpha \\ 0 & 0 & +\alpha \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\1\\1\\1\\1\\1\\1\\1\end{matrix}}\right\} F \\ \left.\vphantom{\begin{matrix}1\\1\\1\\1\\1\\1\end{matrix}}\right\} 2n \\ \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} n_0 \end{matrix}$$

6.2 Scaling of variable values

Referring to the scaling convention of equation (3.23), both the scale factor κ_i (which is the same for each variable, since the CCD is permutation invariant) and the scaled variable value x_{ui} may be derived explicitly in terms of the parameters of the CCD. The derivation of this scaling is seldom found in standard works on the subject, but is presented here in some detail, since an understanding of it aids the discussion of the moment matrix, and orthogonality and rotatability criteria, which follows.

Considering first the average value over the N tests of a variable i, whose upper and lower bounds are b_{ui} and b_{li} respectively, the mean of the bounded levels is given by $\mu_i = (b_{ui} + b_{li}) / 2$, and the distance from this mean to either bound is $\delta_i = (b_{ui} - b_{li}) / 2$. Thus $b_{ui} = \mu_i + \delta_i$, and $b_{li} = \mu_i - \delta_i$. The average value over all of the tests of the CCD may then be expressed as follows

$$\overline{X}_i = \frac{1}{N}\sum_{u=1}^{N} X_{ui} = \frac{1}{N}\left[\frac{F}{2}(\mu_i+\delta_i) + \frac{F}{2}(\mu_i-\delta_i) + 1(\mu_i+\alpha\delta_i) + 1(\mu_i-\alpha\delta_i) + (2n-2)(\mu_i) + n_0(\mu_i)\right]$$

$$= \frac{1}{N}\left[(F + 2n + n_0)\,\mu_i\right]$$

$$\therefore \overline{X_i} = \mu_1, \qquad \text{since } N = F + 2n + n_0$$

This confirms that the average value of each variable over all tests is equal to the mean of the bounded levels. This is due to the fact that the factorial portion of the design contains an equal number of points for which variable i is at its high and low bounds, and that the star points occur in equal and opposite pairs. The scale factor κ of (2.19) may now be expressed as

$$\kappa = \left[\sum_{u=1}^{N} (X_{ui} - \overline{X_i})^2 / N\right]^{-1/2}$$

$$= \left[\frac{1}{N}\left[\frac{F}{2}((\mu_1+\delta_1)-\mu_1)^2 + \frac{F}{2}((\mu_i-\delta_1)-\mu_1)^2 + ((\mu_1+\alpha\delta_1)-\mu_1)^2 + ((\mu_1-\alpha\delta_1)-\mu_1)^2 \right.\right.$$

$$\left.\left. + (2n-2+n_0)(\mu_1-\mu_1)^2\right]\right]^{-1/2}$$

$$= \frac{1}{\delta_1}\left[\frac{1}{N}(F + 2\alpha^2)\right]^{-1/2}$$

The standardised variable value may now be explicitly expressed as

$$x_{ui} = \frac{X_{ui} - \mu_1}{\delta_1}\left[N / (F + 2\alpha^2)\right]^{1/2}$$

or

$$x_{ui} = \frac{X_{ui} - \mu_1}{\delta_1}\, g, \qquad \text{where } g = \left[N / (F + 2\alpha^2)\right]^{1/2} \tag{6.1}$$

The term $(X_{ui} - \mu_1)/\delta_i$ has the effect of standardising the variable range to ± 1, so that, for example, when X_{ui} is the upper bound $\mu_1 + \delta_1$,

$$x_{ui} = \frac{(\mu_i+\delta_1) - \mu_1}{\delta_1}\, g = g$$

and for the star point $X_{ui} = \mu_1 + \alpha\,\delta_1$,

$$x_{ui} = \frac{(\mu_1+\alpha\delta_1) - \mu_1}{\delta_1}\, g = \alpha g$$

The design matrix for the CCD may now be written in the following form

$$
\mathbf{D} = \begin{bmatrix}
-g & -g & -g \\
-g & -g & +g \\
-g & +g & -g \\
-g & +g & +g \\
+g & -g & -g \\
+g & -g & +g \\
+g & +g & -g \\
+g & +g & +g \\
-\alpha g & 0 & 0 \\
+\alpha g & 0 & 0 \\
0 & -\alpha g & 0 \\
0 & +\alpha g & 0 \\
0 & 0 & -\alpha g \\
0 & 0 & +\alpha g \\
0 & 0 & 0 \\
\vdots & \vdots & \vdots \\
0 & 0 & 0
\end{bmatrix}
\begin{matrix} \left.\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}\right\} F \\ \left.\begin{matrix} \\ \\ \\ \\ \\ \\ \end{matrix}\right\} 2n \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_0 \end{matrix}
\qquad (6.2)
$$

If the bounds on variable i are already normalised to ±1, then, since $\mu_i = 0$ and $\delta_i = 1$, equation (6.1) reduces to

$$x_{ui} = g\, X_{ui}\,, \qquad \text{where } g = \left[N / (F + 2\alpha^2)\right]^{1/2} \qquad (6.3)$$

6.3 Orthogonality and rotatability

6.3.1 The moment matrix for a CCD

For a CCD having a design matrix scaled as in equation (6.1), the regressor matrix is as follows, here shown for two variables.

$$\mathbf{X} = \begin{array}{c} \begin{matrix} 1 & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \end{matrix} \\ \left[\begin{array}{cccccc} 1 & -g & -g & g^2 & g^2 & g^2 \\ 1 & -g & +g & g^2 & g^2 & -g^2 \\ 1 & +g & -g & g^2 & g^2 & -g^2 \\ 1 & +g & +g & g^2 & g^2 & g^2 \\ 1 & -\alpha g & 0 & \alpha^2 g^2 & 0 & 0 \\ 1 & +\alpha g & 0 & \alpha^2 g^2 & 0 & 0 \\ 1 & 0 & -\alpha g & 0 & \alpha^2 g^2 & 0 \\ 1 & 0 & +\alpha g & 0 & \alpha^2 g^2 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array} \tag{6.4}$$

so that the moment matrix is of the form

$$N^{-1}\mathbf{X}'\mathbf{X} = \begin{array}{c} \begin{matrix} \quad & 1 & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \end{matrix} \\ \left[\begin{array}{cccccc} 1 & 1 & 0 & 0 & \frac{1}{N}\left[Fg^2+2\alpha^2g^2\right] & \frac{1}{N}\left[Fg^2+2\alpha^2g^2\right] & 0 \\ x_1 & 0 & \frac{1}{N}\left[Fg^2+2\alpha^2g^2\right] & 0 & 0 & 0 & 0 \\ x_2 & 0 & 0 & \frac{1}{N}\left[Fg^2+2\alpha^2g^2\right] & 0 & 0 & 0 \\ x_1^2 & \frac{1}{N}\left[Fg^2+2\alpha^2g^2\right] & 0 & 0 & \frac{1}{N}\left[Fg^4+2\alpha^4g^4\right] & \frac{1}{N}Fg^4 & 0 \\ x_2^2 & \frac{1}{N}\left[Fg^2+2\alpha^2g^2\right] & 0 & 0 & \frac{1}{N}Fg^4 & \frac{1}{N}\left[Fg^4+2\alpha^4g^4\right] & 0 \\ x_1x_2 & 0 & 0 & 0 & 0 & 0 & \frac{1}{N}Fg^4 \end{array}\right] \end{array}$$

Using the notation of Section 2.12, it may be verified that the CCD meets the following requirements, for all $i < j < p < q$, as outlined in Section 5.2.

$$[ij] = [ijj] = [iij] = 0$$
$$[iijp] = [ijjp] = [ijpp] = 0$$
$$[iiij] = [ijjj] = 0$$
$$[ijpq] = [ijp] = 0$$

also for any variable i,

$$[ii] = N^{-1}[Fg^2 + 2\alpha^2 g^2] = g^2 N^{-1}[F + 2\alpha^2]$$

The adoption of the scaling convention of equation (6.1), in which $g = [N / (F + 2\alpha^2)]^{1/2}$, results in a value of $[ii] = 1$, and also yields

$$[iiii] = N^{-1}[Fg^4 + 2\alpha^4 g^4]$$

and $$[iijj] = N^{-1} Fg^4$$

6.3.2 Orthogonality

For a second order design to be orthogonal, Section 5.2 requires that it must also meet the condition $[iijj] = 1$. From the above,

$$[iijj] = \frac{1}{N} Fg^4$$

Substituting for g yields

$$[iijj] = \frac{1}{N} F\left[N / (F+2\alpha^2)\right]^2$$

$$[iijj] = \frac{FN}{(F+2\alpha^2)^2} \qquad (6.5)$$

so that for an orthogonal design

$$FN = (F+2\alpha^2)^2 \qquad (6.6)$$

This may be re-arranged for α to give

$$\alpha = \left[\frac{(FN)^{1/2} - F}{2}\right]^{\frac{1}{2}} \qquad (6.7)$$

Thus for a seven variable example with a half fraction of the 2^7 factorial containing 64 test points, and with n_0 centre point replications, the required value of α is

$$\alpha = \left[\frac{[64\ (64+14+n_0)]^{1/2} - 64}{2}\right]^{\frac{1}{2}} \qquad (6.8)$$

which for various n_0 gives

n_0	α
0	1.824
1	1.885
2	1.943
3	2.000
4	2.055
5	2.108

6.3.3 Rotatability

The conditions for rotatability of second order designs, as outlined in Section 5.3, include the requirements that [ii] = 1 and [iiii] = 3[iijj]. The first of these is automatically met when the scaling convention of equation (6.1) is followed. Referring to the moment matrix above, the second is that

$$Fg^4 + 2\alpha^4 g^4 = 3\,Fg^4$$

$$\alpha^4 g^4 = Fg^4$$

$$\alpha^4 = F$$

$$\text{or} \qquad \alpha = F^{1/4} \tag{6.9}$$

Thus for the seven variable example above, with F = 64, the required value of α is

$$\alpha = 64^{1/4} = 2.8284$$

Comparing this value with those required for orthogonality in the previous section, it is apparent that a large number of centre point replications would be necessary before the two requirements coincided, producing a design which is both orthogonal and rotatable. This is formalised in the following section.

6.3.4 Choice of parameters for both orthogonality and rotatability

If both orthogonality and rotatability are desired, then the value of α is necessarily set by the rotatability requirement of $\alpha = F^{1/4}$. Thus only the number of centre points can be modified in order to attain orthogonality. Substituting $\alpha = F^{1/4}$ into equation (6.6),

$$FN = (F+2F^{1/2})^2$$

but $N = F + 2n + n_0$, so that

$$F(F + 2n + n_0) = (F+2F^{1/2})^2$$

rearranging for n_0,

$$F^2 + 2nF + Fn_0 = F^2 + 4F + 4F(F)^{1/2}$$

$$F + 2n + n_0 = F + 4 + 4(F)^{1/2}$$

$$n_0 = 4(F)^{1/2} - 2n + 4 \qquad (6.10)$$

Thus for the seven variable example with $F = 64$,

$$n_0 = 4(64)^{1/2} - 14 + 4 = 22$$

which is clearly a substantial requirement for centre point runs, representing 22% of the total test burden for this design.

6.3.5 Uniform precision designs

The uniform precision criterion, described in Section 5.3, is an alternative requirement to that of orthogonality. Since a uniform precision design must also be rotatable, the value of α is again set to $\alpha = F^{1/4}$. The required value of n_0 may be assessed by substituting the identity $[iijj] = \lambda_4$ into equation (6.5),

$$[iijj] = \lambda_4 = \frac{FN}{(F+2\alpha^2)^2} \qquad (6.11)$$

Introducing the rotatability condition $\alpha = F^{1/4}$,

$$\frac{FN}{(F+2F^{1/2})^2} = \lambda_4$$

$$N = \left[\lambda_4 (F+2F^{1/2})^2\right] / F$$

$$F + 2n + n_0 = \left[\lambda_4 (F^2 + 4F + 4F(F)^{1/2})\right] / F$$

$$n_0 = \lambda_4 (F + 4 + 4(F)^{1/2}) - F - 2n \qquad (6.12)$$

The value of n_0 for a particular design is obtained by substituting the appropriate value of λ_4, calculated from equation (5.11). Thus, for the seven variable case (n = 7, F = 64), which requires $\lambda_4 = 0.9184$, this yields

$$n_0 = 0.9184\,(64 + 4 + 4(64)^{1/2}) - 64 - 14 = 13.84 \approx 14$$

Equations (6.12) and (6.10) show that a design which is both orthogonal and rotatable will have an additional Δn_0 centre point runs, compared with the uniform precision design, where:

$$\Delta n_0 = (1-\lambda_4)\,(F + 4 + 4(F)^{1/2})$$

Hence, for n = 7, F = 64, and $\lambda_4 = 0.9184$,

$$\Delta n_0 = (1 - 0.9184)\,(64 + 4 + 4(64)^{1/2}) = 8.16 \approx 8$$

6.3.6 Selection of parameters for the CCD

The influence of the above criteria on the selection of parameters for the CCD may be summarised as follows. The first step in the specification of a CCD in n variables is to determine the appropriate factorial portion to use in the design, following the procedures of Chapter 3. With the values of n and F known, the remaining parameters α and n_0 may be chosen to meet one or more of the following criteria.

(i) For orthogonality

$$\alpha = \left[\frac{(FN)^{1/2} - F}{2}\right]^{\frac{1}{2}}$$

any convenient combination of α and n_0 may be selected which meets these requirements, as described in Section 6.3.2.

(ii) For rotatability $\alpha = F^{1/4}$; the value of n_0 does not affect the attainment of rotatability.

(iii) For both orthogonality and rotatability, $\alpha = F^{1/4}$ and $n_0 = 4(F)^{1/2} - 2n + 4$.

(iv) For uniform precision, $\alpha = F^{1/4}$ and $n_0 = \lambda_4 (F + 4 + 4(F)^{1/2}) - F - 2n$.

6.4 Special requirements for engine noise simulation using the CCD

As described in Section 2.4, normalisation of the bounds set on each of the variables gives rise to a design variable region which is hypercuboidal in nature. If the variable values chosen for the factorial portion of the CCD are the variable bounds, then the normalised factorial points (±1, ±1, ... , ±1) lie at the extreme edge of the range of variables. The requirement for rotatability, however, is that $\alpha = F^{1/4}$, which must be greater than 1, since $F > 1$. Thus the star points of the CCD would lie well beyond the limits of each of the variables, which is likely to give rise to engine designs which are physically implausible and/or mathematically impossible.

As an example, consider the variables of the engine model described in Appendix 1C, in which the thickness of the crankcase skirt varies between physically sensible limits of 6mm and 12mm. The mean value of the variable is thus 9mm, and if the bounds are scaled to ±1, the star points at $\alpha = 2.8284$ would lie at 0.51mm and 17.49mm. In this case the upper star point, although higher than would generally be considered in practice, is not a particular problem. The lower value of 0.51mm, however, is clearly impractical as an engine block skirt thickness, both from the point of view of manufacturing capability and of stress levels and structural integrity under load.

This, however, is not the main reason for the lower star point being unacceptable, since the lower bound of 6mm would prevent such a design from being selected at the optimization stage of an analysis. Of much more importance is the fact that the value of the noise function is likely to become highly non-linear as the thickness of this component approaches zero, especially as the skirt is one of the major noise radiating surfaces. This unrepresentative value is therefore likely to be extremely misleading when a model is constructed over the normal variable range.

The problem is even greater in the case of the longitudinal stiffener, whose bounds of 4mm and 14mm would lead to star points of -5.1mm and 23.1mm if $\alpha = 2.8284$. It is clearly impossible to obtain meaningful results from the FE analysis when negative thicknesses are specified, and a value of zero thickness is only possible if the variable is removed from the model.

In order to avoid this problem, it is necessary to modify the experimental design so that the star points do not lie at such inconvenient values. Since, in general, it is not clear where such values become 'inconvenient', a sensible choice of design would seem to be one in which the star points lie at the extreme values of each variable, and not beyond. This condition may be met in one of two ways. The first of these is to modify the variable values in the factorial portion, such that the new range multiplied by α results in star points which lie on the original variable bounds. As an example, if the values of skirt thickness in the factorial portion of the design were modified from 6mm/12mm to 7.94mm/10.06mm, then, for a value of $\alpha = 2.8284$, the star points would lie at 9±[2.8284 × (9-7.94)]mm, which yields 6mm and 12mm. This

approach is clearly unsatisfactory, since the interval between the high and low values of the factorial portion of the design is now reduced from 6mm to 2.12mm, and will thus provide a much inferior estimate of the linear and interaction terms of the model. For a single variable (of the seven introduced in Appendix 1C) only 35% of the range is enclosed within the new bounds. If the same scaling is applied to $k \leq 7$ of the variables, then the total volume enclosed is just $(0.35^k)\times 100\%$, so that for all seven variables only 0.069% of the original design variable space is covered.

The alternative approach is to relax the requirement for exact rotatability, which, as discussed in Section 2.8, is not necessarily a beneficial property when investigating a non-spherical design variable space. The star points may then be set at the edges of the design region, with the factorial points unchanged. For the cuboidal region of the present example, this leads to a value of $\alpha = 1$, which, from equation (6.9) will only give rotatability when $F = 1$. Since a factorial portion containing a single test is clearly of little value, (6.9) shows that no practical CCD with $\alpha = 1$ is rotatable.

If $\alpha = 1$, then the requirement for orthogonality, from equation (6.6), is that

$$FN = (F+2\alpha^2)^2$$

$$FN = (F+2)^2$$

$$FN = (F^2+ 4F + 4)$$

so that

$$N = F + 2n + n_0 = (F^2+ 4F + 4) / F$$

$$n_0 = (F+ 4 + 4/F) - F - 2n$$

$$n_0 = 4 + 4/F - 2n \qquad (6.13)$$

Since $F \geq 1$, n_0 can only possibly have a positive value for $n \leq 4$, and is likely to be negative for all practical designs in $n > 2$. E.g. for $n = 7$, $F = 64$,

$$n_0 = 4 + 4/64 - 14 = -9.9375$$

demonstrating that an orthogonal design is generally unattainable for a CCD with $\alpha = 1$. This is in agreement with the analysis of Section 6.3.2, where it was seen that for orthogonal designs α decreases with decreasing n_0, but that even at $n_0 = 0$ a value of $\alpha = 1.824$ is required when $n = 7$ and $F = 64$.

Since for a design to be of uniform precision it must first be rotatable, it is clear that uniform precision designs cannot be achieved with $\alpha = 1$. This may be confirmed by substitution of $\alpha = 1$ into equation (6.11):

$$FN = \lambda_4(F^2 + 4F + 4)$$

$$n_0 = \lambda_4(F + 4 + 4/F) - F - 2n$$

so that for n=7, F=64, λ_4=0.9184,

$$n_0 = 0.9184\ (64 + 4 + 4/64) - 64 - 14 = -15.49$$

Thus for all practical Central Composite Designs, the choice of parameter $\alpha = 1$ prevents selection of designs which are either rotatable, orthogonal or of uniform precision.

As outlined above, an intrinsic characteristic of the CCD is the use of replicate tests at the centre of the design variable space. When carrying out a programme of physical experimentation, this replication of tests serves the purpose of reducing the effect of random experimental error, allowing improved estimation of the amount of curvature exhibited by each response surface. When using computer simulation techniques, however, the value of a response function which is obtained for a particular combination of input variable levels will always be identical (to machine precision). Thus, replication of centre points yields no additional information in a deterministic environment, imposing an important and fundamental constraint upon the choice of parameters for the CCD.

There would appear to be three possible ways of dealing with this restriction:

(i) Include the replicate centre points within the design, as for normal experimental practice, and simply return identical results for each.

(ii) Reduce the number of centre points in the design to $n_0 = 1$, so that no replication is required.

(iii) Simulate multiple centre points in some way, so as to provide the required information concerning response surface curvature, whilst avoiding the problem of duplicate results which do not contribute to the investigator's knowledge of the system.

Of these three approaches, the first would seem to be the least satisfactory, as it effectively involves misleading the surface-fit algorithm by suggesting that quite independent tests at this point in the design space have led to exactly identical answers. The result of this will

be that the model will be constructed with artificially high confidence levels in the values of the calculated coefficients, with the model itself being unduly influenced by the multiple identical test data. This effect is briefly discussed in Section 6.5.4.

The second option provides the simplest way of avoiding the problem. The main disadvantage of this approach is that only a small amount of data is collected for estimation of the pure quadratic terms of the model. This option is considered in Section 6.5, below, where the extent to which the reduced amount of centre point information affects parameter estimation is investigated in detail. An additional disadvantage of this method is that it is not possible to select the number of centre point replications in order to fulfil the orthogonality or uniform precision criteria. In cases where testing beyond the variable bounds are infeasible, however, this second consideration is of less importance, since the above discussion has shown that, with the necessary value of $\alpha = 1$, it is not generally possible to construct a standard CCD which is either orthogonal or of uniform precision.

The third option requires making fundamental modifications to the standard Central Composite Design. The aim is to allow a greater quantity of information to be gathered close to the centre of the design space, both for deterministic systems generally, and for the case of noise optimization in particular. This important topic is the subject of Chapter 7.

6.5 Numerical tests using the CCD

In order to assess the suitability of the CCD to the optimization of computer simulated radiated noise, a series of numerical studies has been carried out using the seven variable example described in Appendix 1C, based on the FE model of a four cylinder engine block.

The tests outlined below are presented in four sections. Firstly, in Section 6.5.1, a 79 test CCD is used to construct a model containing all 64 of the available linear + interaction terms and the seven pure quadratic coefficients. The performance of this model is discussed, following which the probability plot analysis, previously employed with the linear models of Chapter 4, is used to determine which of the 71 parameters are statistically significant.

The second section, 6.5.2, employs the same experimental design to fit a model containing just those parameters identified to be statistically significant. The characteristics of this model are discussed, and its predictive ability compared with that of the previous model.

In order to widen the scope of the procedure to address cases in which the investigator has no prior knowledge concerning the identity of the significant parameters, Section 6.5.3 investigates the suitability of the strict quadratic model of equation (5.1), again using the 79 test CCD.

Finally, in Section 6.5.4, the use of simple replicate centre points is briefly discussed, with sample results for the strict quadratic model of Section 6.5.3. Further details of tests carried out using simple centre point replication are included in Appendix 6B.

6.5.1 Use of the CCD with $\alpha = 1$, $n_0 = 1$ and 71 coefficients

A standard CCD was first employed, with an axial parameter of $\alpha = 1$ and a single centre point. As recommended in the discussion of Chapter 4, the factorial portion was chosen to be of at least resolution V. In order to obtain this resolution it is necessary to use a 2^{7-1} fraction containing 64 test points, which is of resolution VII. The total number of tests is thus $64 + (2 \times 7) + 1 = 79$. This design allows the construction of a model containing the mean value and up to 63 linear plus interaction terms, to which are added the seven pure quadratic terms, giving a total of 71 coefficients.

The complete list of coefficient estimates obtained from the analysis is given in Table 6.1. The coefficients are arranged in rank order and are normalised for variable values scaled to lie in the range ±1.

1.	MEAN	$8.6711\times10^{+1}$	26.	AC	2.0781×10^{-2}	51.	AB	6.5416×10^{-3}
2.	A	-2.2352	27.	DEG	-1.9127×10^{-2}	52.	BD	6.4857×10^{-3}
3.	A^2	9.8271×10^{-1}	28.	BEG	1.7959×10^{-2}	53.	AEG	-6.2564×10^{-3}
4.	C	-3.8703×10^{-1}	29.	ACE	1.7357×10^{-2}	54.	CDE	5.5655×10^{-3}
5.	G	-3.3550×10^{-1}	30.	BCF	1.6158×10^{-2}	55.	BDF	-5.5565×10^{-3}
6	F	-2.3327×10^{-1}	31.	DG	-1.5427×10^{-2}	56.	BC	-5.1977×10^{-3}
7.	D	-2.2783×10^{-1}	32.	BFG	-1.4492×10^{-2}	57.	BDE	4.0116×10^{-3}
8.	B	-2.0402×10^{-1}	33.	BF	-1.3916×10^{-2}	58.	AFG	3.9592×10^{-3}
9.	D^2	1.9729×10^{-1}	34.	ADG	1.3807×10^{-2}	59.	DF	-3.9314×10^{-3}
10	E	-1.5832×10^{-1}	35.	DEF	1.3163×10^{-2}	60.	CF	3.3634×10^{-3}
11.	G^2	1.1884×10^{-1}	36.	CDF	-1.2974×10^{-2}	61.	EFG	-3.3262×10^{-3}
12.	F^2	-7.8670×10^{-2}	37.	BEF	-1.2788×10^{-2}	62.	CEG	2.4432×10^{-3}
13.	AG	-7.7552×10^{-2}	38.	ABE	-1.2520×10^{-2}	63.	BDG	2.2302×10^{-3}
14.	C^2	6.5926×10^{-2}	39.	CDG	1.2490×10^{-2}	64.	DE	-2.0508×10^{-3}
15.	E^2	5.8004×10^{-2}	40.	BE	1.1641×10^{-2}	65.	ADF	-1.8074×10^{-3}
16.	AD	5.1217×10^{-2}	41.	BCG	-1.0997×10^{-2}	66.	CEF	-1.6460×10^{-3}
17.	FG	-4.6484×10^{-2}	42.	ABD	1.0985×10^{-2}	67.	BG	1.4971×10^{-3}
18	CG	-4.6205×10^{-2}	43.	BCE	-1.0953×10^{-2}	68.	ADE	8.9918×10^{-4}
19.	AF	-4.4455×10^{-2}	44.	CD	1.0701×10^{-2}	69.	ABF	-5.0410×10^{-4}
20	AE	-4.2211×10^{-2}	45.	EG	-9.9781×10^{-3}	70	CE	2.8026×10^{-4}
21.	ACG	3.6145×10^{-2}	46.	DFG	8.6681×10^{-3}	71.	BCD	-1.6804×10^{-4}
22.	EF	-2.3002×10^{-2}	47.	ABC	8.5915×10^{-3}			
23.	B^2	2.2509×10^{-2}	48.	ABG	-8.1680×10^{-3}			
24.	AEF	2.2230×10^{-2}	49.	CFG	-7.4534×10^{-3}			
25.	ACF	-2.0805×10^{-2}	50.	ACD	-7.1088×10^{-3}			

Table 6.1 Noise coefficients for CCD with 79 tests and 71 coefficients (n=7)

Examination of this list reveals that all of the pure quadratic terms appear within the first 23 places, together with the mean effect, the seven main effects, seven of the two-way interactions and just one three–way interaction. The pure quadratic terms are thus making a substantial contribution to the overall model, confirming the conclusion of Section 4.1.2 that significant curvature exists within the design variable space. Indeed, four of the pure quadratic terms (A^2, D^2, G^2 and F^2), are of greater magnitude than any of the interaction terms, and the effect of A^2 is second only to the main effect of that variable.

In evaluating the performance of this design, three factors should be considered;

(i) The relative precision of the parameter estimates for different coefficients

(ii) The statistical independence of each of the estimates

(iii) The accuracy of the analytical model in representing the original response surface.

The first two criteria can be assessed by examining the elements of the covariance matrix for the fitted model. Because the CCD is a permutation invariant design, each coefficient of a given type (main effect, linear interaction, etc.) has the same variance as all others of the same type, and the values of these may be summarised as follows:

Class of Term	Variance
Mean	0.0158
Main effects	0.0152
Linear interactions	0.0156
Pure quadratic	0.1915

Table 6.2 Variance values for CCD with 79 tests and 71 coefficients (n=7)

These results show that the variance on the pure quadratic terms is much greater than those on other terms in the model, indicating that the quadratic parameters have been estimated with much less precision than the linear and interaction terms of the model. This is a source of considerable concern since the quadratic terms are of large magnitude, and hence have a significant influence on the overall shape of the fitted surface.

This difference in variance values is clearly a reflection of the differing amount of data available to evaluate the terms. The estimates of the linear interaction terms rely only upon data being available at two levels of each variable, as supplied by the factorial portion of the design (64 points). The estimation of main effect parameters is also affected by the data collected at star points, since at each of these one of the variables is at either its high or low bound. The star points cannot contribute to estimation of interaction effects, however, since only one variable has a non-zero value at these points, and this accounts for the slightly lower variance on the main effect terms than on the interaction terms. This factor also explains why the parameter estimates for the interaction terms are the same as the values gained from the linear plus interactions model of Section 4.6, (Table 4B.2 of Appendix 4B), based on the resolution VII two-level factorial design. The main effects, however, are slightly modified by the additional star-point data, although comparison of Table 6.1 with Table 4B.2 shows that these modifications are fairly small, and that all linear plus interaction parameters occur in the same order in each table.

In order to estimate the pure quadratic terms of the model, tests must also be carried out at a third value of each of the variables, and this information is provided only by the star and centre portions of the design (15 points). This relative scarcity of third level data leads to a less precise estimate of the pure quadratic terms, compared with the linear and interaction coefficients, and is reflected in the higher variance values for the quadratic parameters.

Estimation of the mean effect is based on the full set of 79 analyser tests, and might thus be expected to have been carried out with greater certainty, and hence lower variance, than that for any other parameter. Table 6.2, however, shows this not to be the case, as the variance value is marginally higher than those for either the main effect or linear interaction terms. The reason for this becomes clear when the magnitude of the covariance effects between different variables are examined. The only non-zero off-diagonal elements of the variance-covariance matrix are the following.

Class of Term	Covariance
between mean and each pure quadratic term	-0.0018
between each pair of pure quadratic terms	-0.0308

Table 6.3 Covariance values for CCD with 79 tests and 71 coefficients (n=7)

Because the factorial portion of the design is an orthogonal array, no covariance exists between main effect or linear interaction terms. These terms are also orthogonal to the pure quadratic terms, with the only dependence being between pairs of pure quadratic terms, and between each of these and the mean. This then is the reason for the reduced precision of the estimate of the mean parameter, since, although a large number of tests is available on which to base the estimate, the possibility exists that the estimate is being distorted by the presence of the poorly defined quadratic terms. Comparison of the magnitudes of these covariance values with the variance on quadratic terms given in Table 6.2 shows that the covariances are significantly smaller than the variance values, so that the quadratic parameter estimates are still largely independent.

The above analysis suggests that there is a significant mismatch between the characteristics of the fitted model and the characteristics of the experimental design which has been carried out in order to estimate the coefficients of the model. A model has been constructed which is considerably influenced by quadratic terms of large magnitude, but these have been estimated with poor precision. Sixty four tests are used to evaluate the linear plus interaction coefficients, many of which are of small magnitude, whilst only fifteen tests have been added in order to assess the important effect of the pure quadratic terms.

The adverse effect of the relatively low precision of estimation of pure quadratic terms is further demonstrated by the use of the probability plot technique to assess the statistical significance of the model terms. An initial plot of all 70 parameters (the mean is not included in this analysis) for the noise function is shown in Figure 6.2.

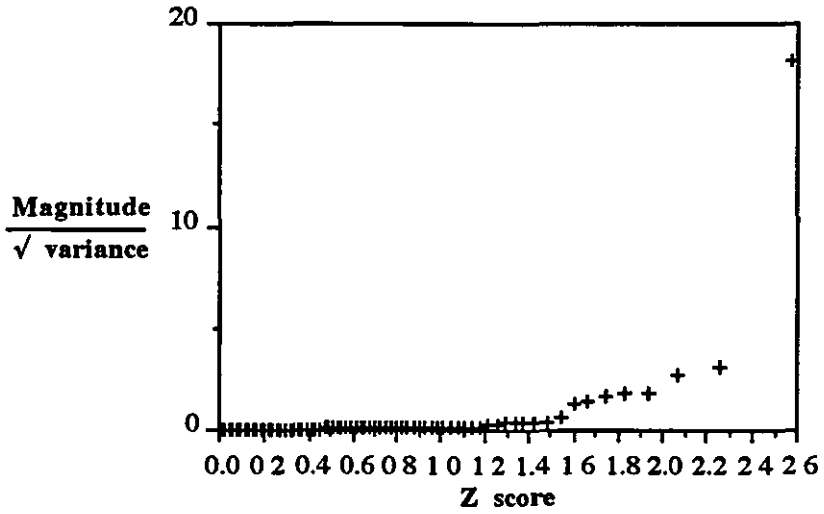

Figure 6.2 CCD in seven variables with 71 terms
Function : noise — No. of coefficients = 70 of 70

From this it is clear that at least the eight largest parameters deviate above the line formed by the smaller terms. In order to assess the remaining terms, the obviously significant terms are first removed from the plot, and the Z score then recalculated for the reduced number of parameters. The resulting plot is shown in Figure 6.3.

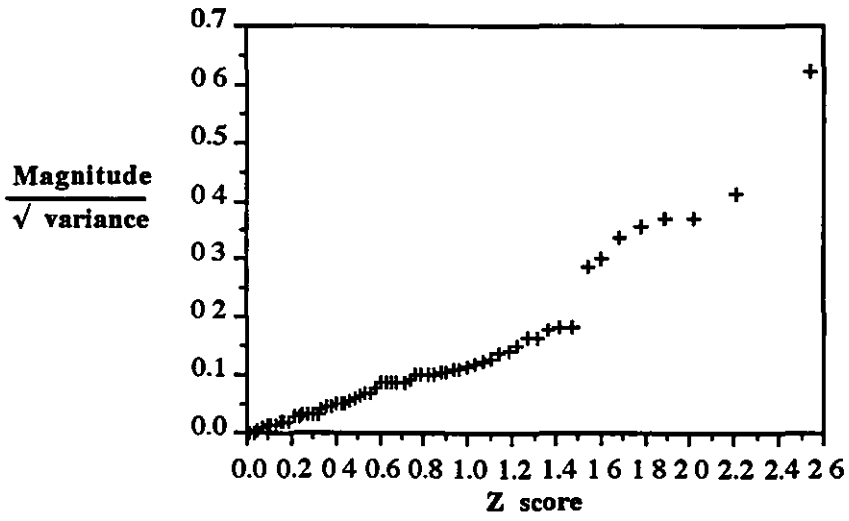

Figure 6.3 CCD in seven variables with 71 terms
Function : noise — No. of coefficients = 62 of 70

This plot shows a further eight terms to be significant, with the remaining terms all appearing to lie on an approximate straight line. This may be confirmed by again removing the significant terms and replotting. Figure 6.4. shows that no further terms deviate above the 'statistical noise' line.

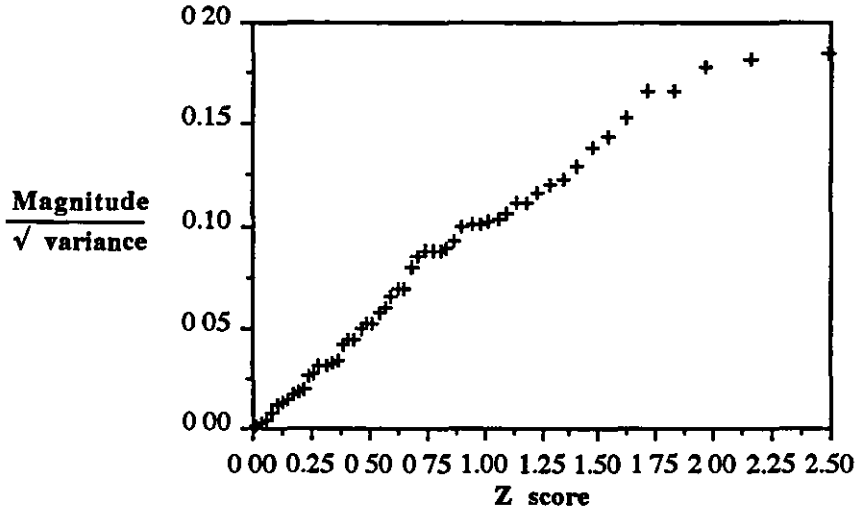

Figure 6.4 CCD in seven variables with 71 terms
Function : noise — No. of coefficients = 54 of 70

The probability plot analysis indicates that, including the mean, only 17 of the initial 71 terms are making a significant contribution to the accuracy of the model in representing the variation of the original response. The terms which contribute significantly to the model, listed in order of significance, are the following.

Term	Coefficient value	Variance	Coeff / √(Var)
MEAN	$8.6711\times10^{+1}$	1.5773×10^{-2}	$6.9769\times10^{+2}$
A	-2.2352	1.5152×10^{-2}	$-1\,8159\times10^{+1}$
C	$-3\,8703\times10^{-1}$	$1\,5152\times10^{-2}$	-3 1443
G	-3.3550×10^{-1}	1.5152×10^{-2}	-2.7256
F	-2.3327×10^{-1}	1.5152×10^{-2}	-1 8951
D	$-2\,2783\times10^{-1}$	1.5152×10^{-2}	-1.8509
B	$-2\,0402\times10^{-1}$	$1\,5152\times10^{-2}$	-1.6574
A^2	$9\,8271\times10^{-1}$	$1\,9147\times10^{-1}$	1 4972
E	-1.5832×10^{-1}	1.5152×10^{-2}	-1.2862
AG	-7.7552×10^{-2}	1.5625×10^{-2}	-6.2041×10^{-1}
AD	5.1217×10^{-2}	1.5625×10^{-2}	$4\,0973\times10^{-1}$
FG	$-4\,6484\times10^{-2}$	1.5625×10^{-2}	-3.7188×10^{-1}
CG	$-4\,6205\times10^{-2}$	$1\,5625\times10^{-2}$	$-3\,6964\times10^{-1}$
AF	$-4\,4455\times10^{-2}$	$1\,5625\times10^{-2}$	-3.5564×10^{-1}
AE	$-4\,2211\times10^{-2}$	1.5625×10^{-2}	-3.3769×10^{-1}
D^2	1.9729×10^{-1}	1.9147×10^{-1}	3.0058×10^{-1}
ACG	$3\,6145\times10^{-2}$	$1\,5625\times10^{-2}$	$2\,8916\times10^{-1}$

Table 6.4 Significant noise coefficients for CCD with 79 tests and 71 coefficients (n=7), listed in order of significance.

Comparison of this coefficient list with Table 6.1 shows that a number of coefficients appear in a different position within the list. This is due to the fact that, in order to assess statistical significance, the values used in constructing the probability plot are first divided by the square root of their variance. These values are given in the 'Coeff / $\sqrt{(Var)}$' column of Table 6.4. Examination of the two lists shows that those main effect and linear interaction terms which have been shown to be significant appear in the same order with respect to each other in both tables, due to their similar variance values. The change in the order of the coefficients is thus due to the fact that the pure quadratic terms have moved further down the table. Of the six pure quadratic terms which appear in the first 17 places of Table 6.1, only two have been found to be statistically significant. The A^2 term is now lower in the list than all but one of the main effects, despite the fact that its magnitude is second only to the main effect of A. This is entirely attributable to the higher variance values which apply to the quadratic terms, and further illustrates the fact that the constructed model is heavily influenced by parameters which are of large magnitude, but for which there is insufficient test data to enable them to be distinguished from random noise. Comparison of Table 6.4 with Table 4.4 shows that those linear and interaction parameters which are identified as being significant are the same as were identified using the 128-term first-order model.

The accuracy of the analytical model in representing the original response surface may be assessed by calculating the lack of fit between the model prediction and the measured value at a number of points within the design variable space. The points used to test the lack of fit in the present model fall into three categories, chosen to investigate three different sources of model inaccuracy. These are as follows.

i) Design points. The lack of fit is calculated at those test points which were used to construct the analytical model. This indicates the ability of the chosen parameters to account for the variations in the function value which occur at the points specified by the CCD. If a design contains the same number of tests as there are coefficients in the model (a saturated design), then the model will reproduce the original data exactly, to machine precision, with no lack of fit at any of these points. In the present case there are 79 tests and 71 coefficients, so that a very close approximation to the original points can be expected.

ii) Factorial points. These are the remaining 64 tests of the complete 2^7 factorial portion which are not included in the CCD. Sizeable lack of fit at these points would indicate that a number of important higher order interaction terms have been omitted from the model. The above probability analysis, showing that only low order interactions contribute significantly, together with the results of Section 4.6, suggests that substantial lack of fit at these points is unlikely. Although useful for the purpose of demonstration, it should be noted that the FE results at these points would not normally be available during an analysis, since their calculation would nearly double the total number of tests required.

iii) Model under-specification due to omission of higher order terms can best be investigated by calculating the lack of fit which occurs at points in the design space which are distant from those used to construct the model. Suitable candidates for such tests would appear to be those points which lie at a radius of $0.5\sqrt{n}$ from the centre of the space and are half way between the centre of the space and the factorial tests at radius $\sqrt{n}$. These points then lie on an n–dimensional hypercube of side n/2, having as its centre the centre of the design variable space, and with its sides aligned with the sides of the 2^n factorial hypercube. This hypercube will of course contain 2^n points, although for reasons of economy a fraction of these points can be used, calculated as for the fractional factorial design. In the present case a 1/8 fraction containing 16 tests has been used. The exact specification of these points is given in Appendix 6A.

The lack of fit results for each of the above categories is summarised in Table 6.5.

Category	No. of tests	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
i) Design points	79	0.094	1.25	0.011	0.15
ii) Factorial points	64	0.208	2.75	0.070	0.93
iii) Higher order	16	0.473	6.25	0.207	2.73
Average over all	159 tests			0.055	0.72

Table 6.5 Summary of lack of fit calculations for CCD with 79 tests and 71 coefficients (n=7)

As expected, the lack of fit at each of the design points is extremely small, due to the near-saturated nature of the design. The maximum value is less than 0.1 dB(A), which is only just over 1% of the function range, with the average value being only 0.15% of the range. Calculation of lack of fit at the remaining factorial points also shows a very small average error of less than 1% of the range, with a maximum of less than 3% (0.2 dB(A)), confirming that the confounding of the 64 higher order terms has little effect on model accuracy.

Of most interest are the lack of fit values at the 16 additional tests, which show that the average prediction error at points lying midway between the centre point and factorial points is 0.2 dB(A), or 2.73% of the variable range, with a maximum of 6.25% ; less than 0.5 dB(A). If the test points chosen are indicative of the design variable region as a whole, then these results suggest that the analytic response surface is providing a very close approximation to the results of the finite element analysis for this particular problem. Although there may well be some locations in the design variable region where the prediction error exceeds the 6.25%-of-range maximum found in these trials, an average error of less than 3% is likely to be sufficiently

accurate to enable the location of a design which is at least in a highly favourable region of the design variable space, even if it is not the absolute optimum. If the average over all of the above tests is computed, then a mean error throughout the region of 0.055 dB(A), or less than 0.75% is obtained. These results show that the second-order model is clearly a substantial improvement on the linear plus-interactions models used in Chapter 4

6.5.2 Use of the CCD with $\alpha = 1$, $n_0 = 1$ and 17 coefficients

The probability plot analysis carried out above indicated that, including the mean value, only 17 of the initial 71 terms contribute significantly to the accuracy of the model in representing the variation of the original response throughout the design variable space. It is reasonable to suppose that a better fit to the original data might be achieved if the insignificant terms were removed from the model and the available data used to obtain a better estimate of those parameters which are known to be significant. If, using the same CCD consisting of 79 test points, the model is respecified to contain only those terms listed in Table 6.4, then the revised parameter estimates, again arranged in rank order and normalised for variable values scaled to lie in the range ±1, are as follows.

Term	Coefficient value
MEAN	$8.6727\times10^{+1}$
A	-2.2352
A^2	1.0666
C	-3.8703×10^{-1}
G	-3.3550×10^{-1}
D^2	2.8113×10^{-1}
F	-2.3327×10^{-1}
D	-2.2783×10^{-1}
B	-2.0402×10^{-1}
E	-1.5832×10^{-1}
AG	-7.7552×10^{-2}
AD	5.1217×10^{-2}
FG	-4.6484×10^{-2}
CG	-4.6205×10^{-2}
AF	-4.4455×10^{-2}
AE	-4.2211×10^{-2}
ACG	3.6145×10^{-2}

Table 6.6 Noise coefficients for CCD with 79 tests and 17 coefficients (n=7)

Estimation of the main effect and linear interaction parameters is unaffected by this reduction in model specification, as these terms are orthogonal to all others within the model. The removal of five of the seven pure quadratic terms, however, has resulted in modified estimates of both the mean and the two remaining pure quadratic terms. Although the coefficients of the mean and A^2 terms have been raised slightly, the most noticeable change is in the value of the D^2 parameter, which has increased in magnitude by over 40% and is now larger than the linear 'main' effect of that variable. The variance values for each type of coefficient are now as follows.

Class of Term	Variance
Mean	0.0155
Main effects	0.0152
Linear interactions	0.0156
Pure quadratic	0.1224

Table 6.7 Variance values for CCD with 79 tests and 17 coefficients (n=7)

The variance on the remaining pure quadratic terms has been substantially reduced (36%) by the removal of five of these terms, since far fewer parameters are now being estimated with the same amount of data. The potential for distortion of the mean value estimate by pure quadratic terms is also now diminished, leading to a slightly smaller variance on this parameter (reduced by 2%). Only three covariance effects now occur, as follows.

Class of Term	Covariance
between A^2 and mean	-0.0057
between D^2 and mean	-0.0057
between A^2 and D^2	-0.0998

Table 6.8 Covariance values for CCD with 79 tests and 17 coefficients (n=7)

The covariance between the two quadratic terms is now more than three times its previous value, but is still small with respect to the variance of the pure quadratic terms. Although the covariance between each quadratic term and the mean has increased by a

factor of approximately 3.2, this is just outweighed by the reduction in the number of quadratic terms by a factor of 3.5, leading to the slight reduction in variance on the mean value.

Recalling the three categories of test point used to assess the accuracy of the model throughout the design region, as described in Section 6.5.2, the lack of fit results for the reduced model may be summarised as follows.

Category	No. of tests	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
i) Design points	79	0.354	4.64	0.059	0.77
ii) Factorial points	64	0.237	3.11	0.059	0.78
iii) Higher order	16	0.490	6.43	0.199	2.62
Average over all 159 tests				0.073	0.96

Table 6.9 Summary of lack of fit calculations for CCD with 79 tests and 17 coefficients (n=7)

The minimal lack of fit at design points when using the 71 parameter model of Section 6.5.1 is largely due to the near-saturated nature of the design, and Table 6.9 shows that the reduction in model size to just 17 terms results in reduced accuracy at these points. The average lack of fit at the design points, however, is still less than 0.06 dB(A), or approximately 0.78% of the variable range, with a maximum error of about 0.35 dB(A).

At the factorial points and at the additional 16 test points the average lack of fit has improved slightly compared with the 71 coefficient model. This indicates that, although the accuracy of fit at design points has diminished, the predictive ability of the model throughout the complete design region has been moderately enhanced. Maximum errors in each of these categories are marginally increased. Calculation of the average lack of fit over all 159 tests shows only a small increase from 0.055 dB(A) to 0.073 dB(A), and this is still less than 1% of the variable range.

The above results show that the predictive ability of the reduced size model is approximately equal to that of the original model, despite the fact that the reduced model contains less than a quarter of the terms used by the original. This confirms the findings of the probability plot analysis that many of the terms contribute little to the overall accuracy of the model. If this is generally true for noise analysis applications of a similar nature, then this provides further evidence that the CCD is less than ideally suited to the characteristics of the response surface being investigated. The low value of the ratio of the number of significant

terms to the number of test points in the design indicates that the CCD involves substantial over-testing for the model which is constructed. At the same time, the selection of these design points is inadequate to allow a precise enough estimation of those quadratic terms which are significant. A possible solution to this problem would be to use a different experimental design from which to construct the model; one which involved a smaller total number of test points, but in which the design points were selected in such a way that the estimates of each of the parameters of interest were attained with roughly equal precision. The use of designs which address these requirements is the subject of Chapter 8.

6.5.3 Use of the CCD with $\alpha = 1$, $n_0 = 1$ and 36 coefficients

As was the case with the linear models of Chapter 4, the main problem which occurs in using a smaller experimental design to fit a reduced model containing only significant terms is that it is not in general possible to know in advance which of the terms are to be included. It is only by carrying out a probability plot analysis, or similar procedure, that this information can be gained. Such an analysis can only be performed, however, if a larger number of tests are first carried out to enable all parameters to be estimated.

To avoid this problem, a model may be used which includes all terms of each of the types thought likely to be significant. When investigating applications for which the results of Table 6.6 are representative, for example, a suitable model would be one which contains all main effect, two-way interaction and quadratic terms. The strict quadratic model of equation (5.1) is such a model. This procedure introduces two possible sources of error into the model specification. The first is due to the inclusion of terms which are in fact of no significance. As long as parameter estimates remain largely orthogonal, however, and the number of insignificant terms is fairly small, this is not expected to lead to substantial error. The second source of error is due to the exclusion from the model of significant parameters, such as, in the present example, the three-way interaction ACG of Table 6.6. If the excluded parameters are small in number and do not dominate the analytical model, then the attendant loss of accuracy would often be considered a small price to pay for the large reduction in the number of design points which need to be tested.

In order to test the suitability of the strict quadratic model in representing the original noise response, this model was constructed using the same CCD consisting of 79 test points, to enable comparison with the results of Sections 6.5.1 and 6.5.2, above. The model contains a total of 36 coefficients, and the parameter estimates obtained from this analysis, arranged in rank order and normalised for variable values scaled to lie in the range ±1, are as follows.

	Term	Coefficient value		Term	Coefficient value
1.	MEAN	$8.6711\times10^{+1}$	19.	AF	-4.4455×10^{-2}
2.	A	-2.2352	20.	AE	-4.2211×10^{-2}
3.	A^2	9.8271×10^{-1}	21.	EF	-2.3002×10^{-2}
4.	C	-3.8703×10^{-1}	22.	B^2	2.2509×10^{-2}
5.	G	-3.3550×10^{-1}	23.	AC	2.0781×10^{-2}
6.	F	-2.3327×10^{-1}	24.	DG	-1.5427×10^{-2}
7.	D	-2.2783×10^{-1}	25.	BF	-1.3916×10^{-2}
8.	B	-2.0402×10^{-1}	26.	BE	1.1641×10^{-2}
9.	D^2	1.9729×10^{-1}	27.	CD	1.0701×10^{-2}
10.	E	-1.5832×10^{-1}	28.	EG	-9.9781×10^{-3}
11.	G^2	1.1884×10^{-1}	29.	AB	6.5416×10^{-3}
12.	F^2	-7.8670×10^{-2}	30.	BD	6.4857×10^{-3}
13.	AG	-7.7552×10^{-2}	31.	BC	-5.1977×10^{-3}
14.	C^2	6.5926×10^{-2}	32.	DF	-3.9314×10^{-3}
15.	E^2	5.8004×10^{-2}	33	CF	3.3634×10^{-3}
16.	AD	5.1217×10^{-2}	34	DE	-2.0508×10^{-3}
17.	FG	-4.6484×10^{-2}	35.	BG	1.4971×10^{-3}
18	CG	-4.6205×10^{-2}	36.	CE	2.8026×10^{-4}

Table 6.10 Noise coefficients for CCD with 79 tests and 36 coefficients (n=7)

The only parameters which have been removed from the original 71 term model of Section 6.5.1 are the linear interactions of order three and higher, the estimates of which are orthogonal to all others in the model. In consequence, the parameter estimates for all remaining terms are identical to those obtained using the full model. All terms of Table 6.10 appear in the same order with respect to each other as in Table 6.1, with the only difference being the removal of the higher order interaction terms. The variance and covariance values for each type of coefficient are also identical to those for the complete model, given in Tables 6.2 and 6.3. A summary of the lack of fit results for the strict quadratic model, calculated at each of the three categories of test point, as described in Section 6.5.2, above, is given in Table 6.11.

Category	No. of tests	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
i) Design points	79	0.306	4.02	0.052	0.68
ii) Factorial points	64	0.202	2.65	0.065	0.86
iii) Higher order	16	0.476	6.24	0.199	2.61
Average over all 159 tests				0.072	0.94

Table 6.11 Summary of lack of fit calculations for CCD with 79 tests and 36 coefficients (n=7)

As may be expected, the above results show that the effect of including 36 terms in the model rather than 17 has been to improve the fit at the design points. The average lack of fit at the factorial points, however, has worsened slightly (but by less than 0.01 dB(A)), perhaps due to the exclusion from the model of the three-way interaction factor ACG.

More important are the results of the lack of fit calculations carried out at the additional sixteen test points. These show that the error in function estimation is virtually identical to that occurring when the 17 coefficient model is used. At these test points the omission of the three-way interaction term is of far less significance than the effect of the various pure quadratic parameters. The probability plot analysis carried out earlier indicated that there is insufficient test information to be able to distinguish from random data the five quadratic terms not appearing in the 17 coefficient model. The accuracy at these additional points suggests, however, that the actual error in their estimation is well within the probable range indicated by the variance values. The uncertainty over the validity of this observation is a further reflection of the inadequacy of the 79 test CCD for fitting a second-order model, due to the relatively small number of tests in which each variable appears at a third level.

6.5.4 Use of the CCD with $\alpha = 1$ and replicate centre points

The results of Sections 6.5.1 - 6.5.3 have shown that one of the major factors which impede the successful application of the CCD to the investigation of deterministic systems is the relatively low precision of estimation which is obtained for the pure quadratic coefficients of the fitted model. When the CCD is used in general experimental practice for the investigation of non-deterministic systems, this problem is alleviated to some extent by the inclusion of replicate tests at the centre of the design variable space, in order to provide additional data at the mean level of the variables. As described in Section 6.4, one simple way of increasing the amount of third level information over that available to the designs of 6.5.1 - 6.5.3 would be to include multiple tests at the centre of the design variable space. This does not require further tests to be carried out, since with a deterministic system the function values calculated at this point will always be identical, and the only action which need be taken is for the appropriate result set to be replicated in the analyser output file.

As discussed in Section 6.4, the result of the inclusion of these extra data sets is likely to be an artificial increase in the confidence levels with which pure quadratic components are estimated, without any real gain in the accuracy of the model. This is because the addition of replicate tests effectively involves misleading the surface-fit algorithm by suggesting that quite independent tests at this point in the design space have led to exactly identical answers. The variance of all parameters in the model could quite easily be decreased, for example, by replicating the entire design. Any apparent gain in accuracy of parameter estimates would be entirely illusory, however, as it would be derived from identical test data, and could not be

expected to result in increased accuracy of prediction at other locations within the design variable space.

In order to assess the effect of using a standard CCD with simple centre point replicates when investigating a deterministic system, the above examples of Sections 6.5.1 - 6.5.3 were repeated using designs containing between 5 and 20 centre points. More detailed results of these tests are presented in Appendix 6B. In each case, the net result of the addition of replicate points is to significantly reduce the variance value for the mean parameter, whilst not substantially improving those for other parameters of the model, or modifying the parameter estimates themselves. Consequently, no significant improvement in the predictive ability of the model throughout the design space is achieved. As an example of the negligible improvement in predictive ability which is obtained, Table 6.12 shows a summary of the lack of fit data for the 36 parameter strict quadratic model introduced in Section 6.5.3. The table lists the prediction error for each of the three categories of test point described in Section 6.5.1, using designs containing a differing number of centre points n_0.

	Maximum lack of fit — E_R			Average lack of fit — E_R		
n_0	Design	Factorial	Higher	Design	Factorial	Higher
1	4.0151	2.6508	6.2354	0.67564	0.85803	2.6048
5	4.0169	2.6527	6.2851	0.65136	0.85811	2.6025
10	4.0182	2.6540	6.3196	0.61978	0.85816	2.6009
15	4.0190	2.6547	6.3404	0.58955	0.85819	2.6000
20	4.0195	2.6552	6.3544	0.56144	0.85821	2.5933

Table 6.12 Variation of lack of fit with number of centre points n_0 for CCD with 78 + n_0 tests and 36 coefficients (n=7)

These results, together with those of Appendix 6B, confirm that simple replication of centre points yields no advantage within the present deterministic environment.

6.6 Observations on the use of the standard CCD for engine noise simulation

- Although the CCD is widely used in general experimental practice, where the value of the parameters α and n_0 can be chosen to achieve certain desirable design properties, its usefulness in investigating the results of an engine noise simulation exercise is adversely affected by the constraints which are placed upon these two parameters by the nature of the problem. Firstly, the value of the axial parameter is limited to $\alpha \leq 1$ by the strictly cuboidal design variable space, with the result that it is not possible to specify designs which are either orthogonal or rotatable. Secondly, the deterministic nature of the engine

noise calculation procedure precludes the use of replicate centre point tests, which would provide additional information concerning the characteristics of the response functions at the centre of the design variable space.

- Numerical tests have shown that a seven variable CCD with a resolution VII factorial portion, an axial parameter of $\alpha = 1$, and a single centre point can be used to construct a second order model which gives a close approximation to the original noise response throughout the region of interest. The predictive ability of the model can be further enhanced by removing those coefficients which, using a probability plot technique, have been found to be statistically insignificant. Further results suggest that when the identity of the significant terms is not known, little lack of performance results from the substitution of a strict quadratic model, in which no linear interaction terms are included of order higher than two.
- The main concern in the use of this design is that the small number of tests which involve a third level of each of the design variables leads to a low level of precision in the estimation of the pure quadratic terms of the model. This is of particular importance, since many of the quadratic coefficients have a dominant effect on the performance of the model, and yet, due to the relative imprecision of their estimated values, it is not possible to say with certainty whether their inclusion is justified by the available data.
- When investigating non-deterministic systems, this problem is alleviated to some extent by the use of replicate centre point tests, in order to improve the quality of the estimate of the actual response at this point, and to counter the bias in the design towards testing at the outer limits of the design variable space. In a deterministic environment, however, no gain in information results from such a strategy, and numerical tests have confirmed that negligible gains in predictive performance of the fitted model are achieved by including even a fairly large number of replicate tests.
- The results of the numerical tests have thus verified that the main issue which needs to be addressed in using the CCD to investigate the engine noise simulation problem, and deterministic systems in general, is the improvement of the precision with which each of the quadratic terms are estimated, with the aim of enhancing the overall performance of the fitted model. It appears likely that some benefit may be obtained by inclusion of modifications to the CCD which allow a greater quantity of information to be gathered in the vicinity of the design space centre. The aim of such modifications is to achieve a high level of prediction accuracy throughout the region of interest, and to aid in distinguishing between the effects of the various quadratic terms in the model. Possible strategies which address these shortcomings of the standard CCD are the subject of Chapter 7.

Appendix 6A

Second order model constructed from CCD
Specification of test points used to estimate lack of fit
due to higher order terms

Variable						
A	B	C	D	E	F	G
.0075	.0230	.0230	.0064	.01375	.0230	.0075
.0105	.0290	.0230	.0064	.01375	.0230	.0105
.0105	.0230	.0290	.0064	.02125	.0230	.0075
.0105	.0230	.0230	.0115	.01375	.0290	.0075
.0075	.0290	.0290	.0064	.01375	.0290	.0075
.0075	.0290	.0230	.0115	.02125	.0230	.0075
.0075	.0230	.0290	.0115	.01375	.0230	.0105
.0075	.0230	.0230	.0064	.02125	.0290	.0105
.0105	.0290	.0290	.0115	.01375	.0230	.0075
.0105	.0290	.0230	.0064	.02125	.0290	.0075
.0105	.0230	.0290	.0064	.01375	.0290	.0105
.0105	.0230	.0230	.0115	.02125	.0230	.0105
.0075	.0290	.0290	.0064	.02125	.0230	.0105
.0075	.0290	.0230	.0115	.01375	.0290	.0105
.0075	.0230	.0290	.0115	.02125	.0290	.0075
.0105	.0290	.0290	.0115	.02125	.0290	.0105

Appendix 6B

Use of the standard CCD with $\alpha = 1$ and replicate centre points

CCD with $\alpha = 1$ and 71 coefficients

The designs used are based on that of Sections 6 5 1, with a number of centre point replications added. The model to be constructed is identical to the 71 coefficient model of Section 6 5.1.

Estimates of the main effect and interaction parameters are unaffected by the inclusion of additional centre point data, since at the centre point each variable is at its mean value (normalised as zero), and no information is provided on the linear effects of these variables.

The variation of mean and pure quadratic coefficients with n_0 is given in Table 6B.1. The terms are listed, from left to right, in order of magnitude. The variance and covariance values are given in Table 6B 2.

n_0	MEAN	A^2	D^2	G^2	F^2	C^2	E^2	B^2
1	86.711	0 9827	0 1973	0 1188	-0 0787	0 0659	0.0580	0.0225
5	86 716	0 9820	0.1965	0 1181	-0 0794	0 0652	0.0573	0.0218
10	86 720	0.9815	0.1960	0 1176	-0 0799	0 0647	0.0567	0 0212
15	86 722	0 9811	0.1957	0.1173	-0 0802	0 0643	0.0564	0 0209
20	86.724	0.9809	0 1955	0.1171	-0 0805	0.0641	0.0562	0 0207

Table 6B.1 Variation of noise coefficients with number of centre points n_0 for CCD with 78 + n_0 tests and 71 coefficients (n=7)

	Variances		Covariances	
n_0	MEAN	Quadratic	MEAN / Quad	Quad / Quad
1	0 0158	0.1915	-0 00176	-0 0308
5	0 0134	0.1912	-0 00104	-0 0310
10	0.0118	0 1911	-0 00054	-0.0311
15	0.0109	0.1910	-0 00024	-0 0312
20	0.0102	0 1909	-0 00003	-0.0313

Table 6B.2 Variation of variance and covariance values with number of centre points n_0 for CCD with 78 + n_0 tests and 71 coefficients (n=7)

The variance value for the mean term has been greatly reduced, whilst the magnitude of the coefficient remains virtually unchanged. This small effect on the mean coefficient is as may be expected, since the additional results used to fit the model are all calculated at the centre of the design space. When using linear or quadratic models the model prediction at the centre point is by definition equal to the value of the mean coefficient of the model. Since the tests of 6.5 2 have shown that the lack of fit between the model and the actual result at this point is small, this means that the value of the mean term in the model is already almost identical to the actual value at the design centre, and addition of further results can have little effect.

The reduction in variance value reflects the illusory gain in precision which results from using an increased number of test points to estimate the parameter value.

The variance values for the pure quadratic terms remain virtually unchanged by the increase in centre point information, whilst the covariance effects between them increase. The reason for this is that the additional test points provide function information at a point at which all of the variables are at their mean level, so that the individual effect of each term cannot be separately identified. Indeed, the proportion of third level tests which represent the effect of all of the quadratic components is now larger, and leads to greater interdependency of the parameter estimates, and hence higher covariance values.

The small change in parameter estimates for the quadratic terms is also due to the small lack of fit occurring in the original model. The additional tests have the effect of reducing the lack of fit at the centre point, equivalent to weighting a single test result at this point, and the scope for such a reduction is extremely limited.

The maximum and average percentage lack of fit for each value of n_0 at each category of test point, as defined in Section 6 5 1, are shown in Table 6B.3.

	Maximum lack of fit - E_R			Average lack of fit - E_R		
n_0	Design	Factorial	Higher	Design	Factorial	Higher
1	1.2483	2.7467	6.2519	0 15039	0.93156	2.7334
5	1 3058	2.7486	6.3022	0 15145	0.93143	2.7312
10	1.3457	2.7499	6.3371	0 14830	0.93134	2.7297
15	1.3698	2.7507	6.3582	0.14346	0 93129	2.7287
20	1 3859	2.7512	6 3723	0 13812	0 93125	2.7281

Table 6B.3 Variation of lack of fit with number of centre points n_0 for CCD with 78 + n_0 tests and 71 coefficients (n=7)

Although the number of tests points in these designs has increased by between 5% and 24%, virtually no gain in accuracy has been obtained.

The maximum reduction of 0 005% in average lack of fit at the additional 16 test points from 2.7334 to 2 7281 does not compare favourably with the reduction to 2.6160 which was gained in Section 6.5 2 by using the original design and just 17 coefficients.

The slightly larger reduction at the design points is of no practical significance as the fit at these points is already close to zero.

The addition of replicate centre points has thus resulted in a significant reduction in the variance value for the mean parameter, with no substantial improvement in those for other parameters.

There is no significant modification of parameter values, and no significant improvement in the predictive ability of the model throughout the design space has been achieved.

CCD with α = 1 and 17 coefficients

The same selection of designs is used as in the above analysis.

The reduced model of Section 6 5.2 is employed, which contains the 17 terms listed in Table 6.6.

The estimates of the main effect and interaction parameters are again unaffected by the inclusion of additional centre point data.

Only two of the seven pure quadratic terms appear in this reduced model, so that only the mean value and the two quadratic terms A^2 and D^2 may be affected by the additional data.

The variation of these coefficients with n_0 is given in Table 6B.4, and their variance and covariance values are given in Table 6B.5.

n_0	MEAN	A^2	D^2
1	86.727	1.0666	0.2811
5	86.728	1 0661	0.2807
10	86.729	1 0658	0 2804
15	86 729	1 0656	0 2802
20	86.729	1.0654	0.2800

Table 6B.4 Variation of noise coefficients with number of centre points n_0 for CCD with 78 + n_0 tests and 17 coefficients (n=7)

Again, all parameter magnitudes are virtually unchanged, as are the variance values of the pure quadratic terms. The variance of the mean value is significantly reduced, whilst the covariance effects between the two pure quadratic terms gradually increases with n_0.

	Variances		Covariances	
n_0	MEAN	A^2 and D^2	MEAN / A^2 or D^2	A^2 / D^2
1	0 0155	0.1224	-0 00571	-0 0998
5	0 0134	0.1200	-0 00342	-0.1022
10	0 0118	0.1183	-0 00179	-0 1039
15	0 0109	0.1172	-0 00079	-0 1050
20	0 0102	0.1165	-0 00011	-0 1057

Table 6B.5 Variation of variance and covariance values with number of centre points n_0 for CCD with 78 + n_0 tests and 17 coefficients (n=7)

The maximum and average percentage lack of fit are as follows.

	Maximum lack of fit - E_R			Average lack of fit - E_R		
n_0	Design	Factonal	Higher	Design	Factorial	Higher
1	4.6395	3.1124	6 4331	0 77241	0.77967	2.6160
5	4.6399	3.1124	6 4423	0.73677	0.77973	2 6157
10	4 6403	3.1124	6.4488	0 69597	0.77976	2 6154
15	4 6404	3.1124	6.4528	0 65918	0.77979	2 6153
20	4.6406	3 1124	6 4555	0 62595	0 77980	2.6152

Table 6B.6 Variation of lack of fit with number of centre points n_0 for CCD with 78 + n_0 tests and 17 coefficients (n=7)

There is virtually no change in the lack of fit calculations The reduction in lack of fit at the design points is made possible by the greater initial error resulting from the small number of coefficients in the model

CCD with α = 1 and 36 coefficients

The model used is the 36 parameter strict quadratic, which contains all seven pure quadratic terms.

The variation of these coefficients with n_0 is given in Table 6B.7, and their variance and covariance values are given in Table 6B 8. The values of all parameter estimates, variance and covariance values are identical to those of the 71 coefficient model.

The maximum and average percentage lack of fit are shown in Table 6B.9.

n_0	MEAN	A^2	D^2	G^2	F^2	C^2	E^2	B^2
1	86.711	0.9827	0.1973	0.1188	-0 0787	0.0659	0.0580	0.0225
5	86.716	0.9820	0 1965	0 1181	-0 0794	0 0652	0 0573	0.0218
10	86 720	0 9815	0 1960	0.1176	-0 0799	0 0647	0 0567	0 0212
15	86 722	0 9811	0.1957	0 1173	-0 0802	0 0643	0.0564	0 0209
20	86 724	0 9809	0 1955	0 1171	-0 0805	0 0641	0.0562	0 0207

Table 6B.7 Variation of noise coefficients with number of centre points n_0 for CCD with 78 + n_0 tests and 36 coefficients (n=7)

	Variances		Covariances	
n_0	MEAN	Quadratic	MEAN / Quad	Quad / Quad
1	0 0158	0 1915	-0 00176	-0 0308
5	0 0134	0.1912	-0.00104	-0 0310
10	0 0118	0.1911	-0 00054	-0 0311
15	0 0109	0 1910	-0 00024	-0.0312
20	0 0102	0.1909	-0 00003	-0.0313

Table 6B.8 Variation of variance and covariance values with number of centre points n_0 for CCD with 78 + n_0 tests and 36 coefficients (n=7)

	Maximum lack of fit - E_R			Average lack of fit - E_R		
n_0	Design	Factorial	Higher	Design	Factorial	Higher
1	4.0151	2.6508	6 2354	0 67564	0.85803	2.6048
5	4 0169	2.6527	6 2851	0 65136	0 85811	2.6025
10	4.0182	2 6540	6.3196	0 61978	0 85816	2.6009
15	4.0190	2 6547	6.3404	0.58955	0 85819	2.6000
20	4.0195	2.6552	6 3544	0.56144	0.85821	2.5933

Table 6B.9 Variation of lack of fit with number of centre points n_0 for CCD with 78 + n_0 tests and 36 coefficients (n=7)

The average lack of fit at both the factorial points and the additional test points are essentially unchanged for each value of n_0. The reduction in lack of fit at the design points is again made possible by the initial error caused by the relatively high ratio of tests to coefficients, compared with the full 71 term model.

7. Centre point replication

The results of the previous chapter have shown that one of the major disadvantages of the standard CCD is the poor precision with which the pure quadratic coefficients of a response surface are estimated. This is a particular problem when investigating deterministic systems, since it is not possible to gain additional information concerning the effect of pure quadratic terms by performing replicate tests at the centre of the design space.

The purpose of the present chapter is to describe a number of strategies which have been developed within the current work in order to simulate centre point replication when investigating deterministic systems. Investigation of proposed designs falls into two categories. Theoretical considerations, based on the proposed design matrix, establish the effect on the orthogonality and rotatability characteristics of the standard CCD. These are followed by numerical trials to assess improvements in parameter estimation and quality of fit of the predictive model.

In Section 7.1 the general requirements of a strategy for simulated centre-point replication are first described. In Section 7.2 a modification to the standard CCD is proposed which incorporates a second axial portion of tests close to the centre of the design variable space, and the design matrix, moment matrix and scaling factor for this design are derived. The orthogonality and rotatability criteria for the design are developed in Sections 7.2.1—7.2.3. It is shown that, although the criteria differ slightly from those of the standard CCD, the orthogonality and rotatability characteristics of this design are little altered from those of the standard CCD. In Section 7.3 a variant of this design is proposed, in which only one pair of the additional axial points is included. It is shown that this results in a need for multiple scale factors, which lead to multiple criteria for rotatability and orthogonality which cannot be fulfilled simultaneously. The small difference in scaling factors is shown to be of no practical importance, however, so that the standard CCD scaling can be employed with little error. In Section 7.4 the results of this analysis are shown to be valid for designs which involve more than one pair of additional axial points. The application of these designs, and possible extensions to the modifications, are discussed in Section 7.5.

The results of numerical tests using the modified designs are presented in Sections 7.6.1 to 7.6.3, and these are introduced in more detail in Section 7.6. Section 7.7 summarises the results of these trials, and assesses the use of simulated centre point replications.

7.1 Requirements of a strategy for simulation of centre point replication

The aim of a centre point replication simulation strategy is to provide a means of collecting additional response function information at points close to the centre of the region of interest. The purpose of this is both to improve the distribution of test points throughout the design variable space, and to enhance the ability of the design to estimate the effect of pure quadratic terms in a fitted model. In order to be of use, any strategy which attempts to achieve such an aim must fulfil the following requirements:

i) The data collected must be calculated at test points which are sufficiently separate from each other that the values returned do genuinely give additional information on the process under investigation. Quite what the lower bound on such separation should be is unclear, with the limiting case being that in which zero separation between the multiple points occurs, as employed in the investigation of non-deterministic system.

ii) A contrasting constraint on this separation is that the multiple points must each still lie in the required section of the design region. Centre point 'approximations' which lie close to the perimeter of the design region are clearly unacceptable, but again, the specification of an upper bound on the allowable separation is far from straightforward.

ii) A further consideration when selecting a suitable strategy for replication simulation is that the points chosen to be added to the original CCD should ideally not compromise other beneficial characteristics of the design, such as rotatability or orthogonality.

7.2 Addition of the ε-star portion

One possible strategy which appears to have the potential to meet the requirements described above is the addition of a further 'star' portion to the existing Central Composite Design, but with parameter ε replacing the value of α defined in Section 6.1. Within this general class of design, the value of ε can be selected to provide the required separation between the test points, as well as maintaining other desirable design properties. A typical design matrix for such an augmented CCD, here shown for three variables, would be of the form

$$\mathbf{D} = \begin{bmatrix} -g & -g & -g \\ -g & -g & +g \\ -g & +g & -g \\ -g & +g & +g \\ -g & -g & -g \\ -g & -g & +g \\ -g & +g & -g \\ -g & +g & +g \\ -\alpha g & 0 & 0 \\ +\alpha g & 0 & 0 \\ 0 & -\alpha g & 0 \\ 0 & +\alpha g & 0 \\ 0 & 0 & -\alpha g \\ 0 & 0 & +\alpha g \\ -\varepsilon g & 0 & 0 \\ +\varepsilon g & 0 & 0 \\ 0 & -\varepsilon g & 0 \\ 0 & +\varepsilon g & 0 \\ 0 & 0 & -\varepsilon g \\ 0 & 0 & +\varepsilon g \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} \left.\right\} F \\ \left.\right\} 2n \\ \left.\right\} 2n \\ \left.\right\} n_0 \end{matrix} \qquad (7.1)$$

where the factorial and axial portions are the same as for the standard CCD, and the ε-star portion is of the same form and consists of the same number of tests as the axial portion, but with parameter ε. Since the object of adding this portion is to avoid replicate centre points, n_0 should clearly only take a value of 0 or 1.

In a similar analysis to that for the standard CCD, the value of the scaling factor g may be calculated according to the scaling convention of Section 2.12, so that

$$\overline{X_i} = \frac{1}{N}\sum_{u=1}^{N} X_{ui} = \frac{1}{N}[(F + 2n + 2n + n_0)\ \mu_i] = \mu_i$$

and hence

$$\kappa = \left[\sum_{u=1}^{N} (X_{ui} - \overline{X}_i)^2 / N\right]^{-1/2} = \delta_i \left[\frac{1}{N}(F + 2\alpha^2 + 2\varepsilon^2)\right]^{-1/2}$$

The gives the value of the scaled variable as

$$x_i = \frac{X_i - \mu_i}{\delta_i}\, g$$

where

$$g = \left[N / (F + 2\alpha^2 + 2\varepsilon^2) \right]^{\frac{1}{2}} \tag{7.2}$$

Note that the only difference between this scaling and that used for the standard CCD is in the expression for the scaling factor g. The regressor matrix for this design is as follows

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccccccccc} 1 & x_1 & x_2 & x_3 & x_1^2 & x_2^2 & x_3^2 & x_1x_2 & x_1x_3 & x_2x_3 \end{array} \\ \left[\begin{array}{cccccccccc} 1 & -g & -g & -g & g^2 & g^2 & g^2 & g^2 & g^2 & g^2 \\ 1 & -g & -g & +g & g^2 & g^2 & g^2 & g^2 & -g^2 & -g^2 \\ 1 & -g & +g & -g & g^2 & g^2 & g^2 & -g^2 & g^2 & -g^2 \\ 1 & -g & +g & +g & g^2 & g^2 & g^2 & -g^2 & -g^2 & g^2 \\ 1 & +g & -g & -g & g^2 & g^2 & g^2 & -g^2 & -g^2 & g^2 \\ 1 & +g & -g & +g & g^2 & g^2 & g^2 & -g^2 & g^2 & -g^2 \\ 1 & +g & +g & -g & g^2 & g^2 & g^2 & g^2 & -g^2 & -g^2 \\ 1 & +g & +g & +g & g^2 & g^2 & g^2 & g^2 & g^2 & g^2 \\ 1 & -\alpha g & 0 & 0 & \alpha^2 g^2 & 0 & 0 & 0 & 0 & 0 \\ 1 & +\alpha g & 0 & 0 & \alpha^2 g^2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -\alpha g & 0 & 0 & \alpha^2 g^2 & 0 & 0 & 0 & 0 \\ 1 & 0 & +\alpha g & 0 & 0 & \alpha^2 g^2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -\alpha g & 0 & 0 & \alpha^2 g^2 & 0 & 0 & 0 \\ 1 & 0 & 0 & +\alpha g & 0 & 0 & \alpha^2 g^2 & 0 & 0 & 0 \\ 1 & -\varepsilon g & 0 & 0 & \varepsilon^2 g^2 & 0 & 0 & 0 & 0 & 0 \\ 1 & +\varepsilon g & 0 & 0 & \varepsilon^2 g^2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -\varepsilon g & 0 & 0 & \varepsilon^2 g^2 & 0 & 0 & 0 & 0 \\ 1 & 0 & +\varepsilon g & 0 & 0 & \varepsilon^2 g^2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -\varepsilon g & 0 & 0 & \varepsilon^2 g^2 & 0 & 0 & 0 \\ 1 & 0 & 0 & +\varepsilon g & 0 & 0 & \varepsilon^2 g^2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \end{array}$$

Calculation of the elements of the moment matrix $N^{-1}\mathbf{X}'\mathbf{X}$ shows that, as for the standard CCD, this design meets the following requirements, for all $i < j < p < q$, as outlined in Section 5.2.

$$[ij] = [\text{ı}jj] = [iij] = 0$$
$$[iijp] = [\text{ı}jjp] = [ijpp] = 0$$
$$[\text{ı}ij] = [ijjj] = 0$$
$$[ijpq] = [\text{ı}jp] = 0$$

and that for any variable i,

$$[ii] = N^{-1}[Fg^2 + 2\alpha^2 g^2 + 2\varepsilon^2 g^2] = g^2 N^{-1}[F + 2\alpha^2 + 2\varepsilon^2]$$

The adoption of the scaling convention of equation (7.2), in which $g = [N / (F + 2\alpha^2 + 2\varepsilon^2)]^{1/2}$, results in a value of $[ii] = 1$, and also yields

$$[\text{ı}ii] = N^{-1}[Fg^4 + 2\alpha^4 g^4 + 2\varepsilon^4 g^4]$$

and $$[\text{ı}jj] = N^{-1} Fg^4$$

7.2.1 Orthogonality

The requirement for orthogonality, as described in Section 5.2, is that $[iijj] = 1$, so that, from the above

$$\frac{1}{N} F\left[N / (F + 2\alpha^2 + 2\varepsilon^2)\right]^2 = 1$$

$$FN = (F + 2\alpha^2 + 2\varepsilon^2)^2 \qquad (7.3)$$

Since both F and N are necessarily positive, and $N \geq F$, this equation can be rearranged for α and ε as

$$\alpha^2 + \varepsilon^2 = \left[\frac{(FN)^{1/2} - F}{2}\right]^{\frac{1}{2}} \qquad (7.4)$$

Thus, with F and N set, it is possible to choose the value of either α or ε and solve for the other. With a value of $\varepsilon = 0$, this equation is clearly identical to equation (6.7) for the standard CCD, and returns a value for the axial parameter of $\alpha = 2.5641$ for the seven variable case with $F = 64$ and $N = F + 2n + 2n + 1 = 93$. Specification of values of ε within the range to $\varepsilon = 0.1$ give the following values for α.

ε	α
0.000	2.5641
0.025	2.5640
0.050	2.5636
0.075	2.5630
0.100	2.5622

Selection of even a fairly large ε parameter thus has a virtually negligible effect on the choice of axial parameter α. It is therefore reasonable to employ the same value of α as would be used in a standard CCD, which will result in little loss of orthogonality. This analysis indicates that the design may be treated as though it were a standard CCD with $2n + n_0$ centre points, but with replicate centre points replaced with the ε-star points. Substitution of appropriate values into equation (7.4) show that this is generally true for all practical designs, with either $n_0 = 1$ or $n_0 = 0$.

In cases where the design region is strictly cuboidal, such as the noise radiation problem currently under investigation, the value of α is already fixed at $\alpha = 1.0$, as discussed in Section 6.4. There is then only one possible value of ε which will give an orthogonal design, and, for the seven variable example, (7.4) gives values of $\varepsilon = 2.361$ for $n_0 = 1$, and $\varepsilon = 2.317$ for $n_0 = 0$. These points are clearly unacceptable as centre point approximations, and it is generally true that orthogonal designs of this type are unattainable when $\alpha = 1.0$, as indeed is the case for the standard CCD.

7.2.2 Rotatability

The rotatability requirement, as discussed in Section 5.3, is that [iiii] = 3[iijj], where for the CCD plus ε-star design

$$[iiii] = N^{-1}[Fg^4 + 2\alpha^4 g^4 + 2\varepsilon^4 g^4]$$

and $$[iijj] = N^{-1} Fg^4$$

so that for rotatability

$$\frac{g^4}{N}(F + 2\alpha^4 + 2\varepsilon^4) = 3\frac{g^4}{N}F$$

giving

$$\alpha^4 + \varepsilon^4 = F \qquad (7.5)$$

For any ε which is small enough to give tests which approximate the centre point replications, the value of ε^4 will clearly be small in relation to the magnitude of F, so that for all designs of this type the value of α from equation (7.5) is in practice identical to that of (6.9) for the standard CCD. This value is independent of both the number of centre points n_0 and the total number of test points. For the seven variable example with F = 64, the required value of α, if $\varepsilon = 0.05$, is 2.8284, only differing at the ninth significant figure from that required for the standard CCD. There is therefore little error in treating this design as a standard CCD with replicate centre points, and choosing the value of α according to (6.9).

As is the case for the standard CCD, the imposition of a value of $\alpha = 1.0$ prevents any design from being rotatable. From (7.5), the requirement would be that $\varepsilon = (F - 1)^{1/4}$, which requires $\varepsilon = 0$ for $F = 1$, and $\varepsilon \geq 1$ for $F \geq 2$, and so will not yield sufficiently small values of ε for any practical F.

7.2.3 Choice of parameters for both orthogonality and rotatability

If both of these properties are desired, then values of α and ε must be chosen such that they simultaneously satisfy both (7.4) and (7.5). Unlike the standard CCD, the number of centre points cannot be freely selected to aid the attainment of these requirements, and must be set to either 0 or 1. The two criteria which must be met are then

$$FN = (F + 2\alpha^2 + 2\varepsilon^2)^2$$

and

$$\alpha^4 + \varepsilon^4 = F$$

with the restriction that the only solution of these equations which is of practical use is one for which both α and ε are both real and positive. Since all of the terms of the above equations must then be positive, it is possible to form

$$(FN)^{1/2} = (F + 2\alpha^2 + 2\varepsilon^2)$$

without loss of sign information. Also, since $\varepsilon = (F - \alpha^4)^{1/4} > 0$,

$$(FN)^{1/2} = F + 2\alpha^2 + 2(F - \alpha^4)^{1/2}$$

$$(F - \alpha^4) = \left[\frac{(FN)^{1/2} - F - 2\alpha^2}{2}\right]^2$$

so that

$$8\alpha^4 + \alpha^2(4F - 4(FN)^{1/2}) + (F^2 + FN - 2F(FN)^{1/2} - 4F) = 0$$

giving

$$\alpha^2 = \left[-(4F - 4(FN)^{1/2}) \pm \sqrt{(4F - 4(FN)^{1/2})^2 - 32(F^2 + FN - 2F(FN)^{1/2} - 4F)}\right] / 16 \qquad (7.6)$$

The positive value of α^2 which this analysis produces, together with the value of ε^2 which is necessary to satisfy both of the initial equations, is tabulated in Appendix 7A for all designs in $3 < n < 12$ which have factorial portions of resolution V or greater. This clearly shows that in general it is not possible to produce designs of this type which are both orthogonal and rotatable, since for a positive value of α^2 the value of ε^2 which results is almost always negative. Of the designs considered in Appendix 7A, the only one which is feasible is

the resolution V design in 5 variables, which yields $\alpha \approx 2.00$ and $\varepsilon \approx 0.411$. Even this design is not acceptable, however, as the value of ε obtained places the resulting points at a considerable distance from the design centre. For any problem which has a strictly cuboidal design region, the value of $\alpha \approx 2.00$ is also unacceptable.

Sections 7.2.1 and 7.2.2 have shown that the orthogonality and rotatability criteria of the standard CCD can be used with little resulting error. If the standard CCD criterion is used in the present case, however, then Section 6.3.4 shows that, for the seven variable example, a value of $n_0 = 22$ is required. Since the maximum number of centre point and pseudoreplicate tests available when using an additional ε star is $n_0 = 2n + 1$, giving 15 points for $n = 7$, the rotatability and orthogonality criteria cannot simultaneously be met with the present design. An extension to the design which allows for additional pseudoreplicate tests is discussed in Section 7.5.

7.2.4 Observations on the use of the CCD with additional ε-star portion

In summary, the analyses of the previous sections have shown that the addition of an axial portion of 2n tests with parameter $\varepsilon << \alpha$ causes no significant change in rotatability and orthogonality characteristics compared with the standard CCD. This design shares with the standard CCD the feature that an imposed value of $\alpha = 1$ leads to designs which are neither orthogonal nor rotatable.

7.3 The CCD with one ε pair

If the additional number of test points introduced by the incorporation into the CCD of a complete ε star is considered excessive, then a possible modification to this design is to include only a fraction of these points. In general, the augmentation might consist of k ($\leq$ n) pairs of points with parameter $\pm\varepsilon$ lying on each of k variable axes. Unless specific knowledge of the application suggests particular choices of variable, this selection is entirely arbitrary, although the choice would not normally be expected to be critical, since the points are closely grouped around the centre of the design variable space in comparison with the dimension of the complete design region. In the following sections a design incorporating just one ε pair is analysed in detail. In Section 7.4 it is shown that these results, together with the results of the standard CCD and the CCD with full ε star, are sufficient to allow conclusions to be drawn regarding all designs with $1 \leq k \leq n$ ε pairs.

For the case in which the ε pair for just one variable is included, the design matrix for a three variable example is as follows, with the value of n_0 being either 0 or 1, as above.

$$\mathbf{D} = \begin{bmatrix} -g & -g & -h \\ -g & -g & +h \\ -g & +g & -h \\ -g & +g & +h \\ -g & -g & -h \\ -g & -g & +h \\ -g & +g & -h \\ -g & +g & +h \\ -\alpha g & 0 & 0 \\ +\alpha g & 0 & 0 \\ 0 & -\alpha g & 0 \\ 0 & +\alpha g & 0 \\ 0 & 0 & -\alpha h \\ 0 & 0 & +\alpha h \\ 0 & 0 & +\varepsilon h \\ 0 & 0 & -\varepsilon h \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\\a\\a\\a\\a\\a\end{matrix}}\right\} F \\ \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} 2n \\ \left.\vphantom{\begin{matrix}a\\a\end{matrix}}\right\} 2 \\ \} \; n_0 \end{matrix} \qquad (7.7)$$

For those variables which do not appear at non-zero values in the now reduced ε-star portion of the design (non-ε variable), the value of the scale factor g, calculated from equation (5.3), is identical to that for the standard CCD, i.e.,

$$g = \left[N / (F + 2\alpha^2)\right]^{\frac{1}{2}}$$

In order that the variable which does appear at non-zero values in the ε-star portion (ε variable) achieves the criterion [ii] = 1, the value of the scale factor h is required to be the same as for the variables of the full ε-star design;

$$h = \left[N / (F + 2\alpha^2 + 2\varepsilon^2)\right]^{\frac{1}{2}}$$

The regressor matrix for the single ε-pair design is then of the form

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccccccccc} 1 & x_1 & x_2 & x_3 & x_1^2 & x_2^2 & x_3^2 & x_1x_2 & x_1x_3 & x_2x_3 \end{array} \\ \left[\begin{array}{cccccccccc} 1 & -g & -g & -h & g^2 & g^2 & h^2 & g^2 & gh & gh \\ 1 & -g & -g & +h & g^2 & g^2 & h^2 & g^2 & -gh & -gh \\ 1 & -g & +g & -h & g^2 & g^2 & h^2 & -g^2 & gh & -gh \\ 1 & -g & +g & +h & g^2 & g^2 & h^2 & -g^2 & -gh & gh \\ 1 & +g & -g & -h & g^2 & g^2 & h^2 & -g^2 & -gh & gh \\ 1 & +g & -g & +h & g^2 & g^2 & h^2 & -g^2 & gh & -gh \\ 1 & +g & +g & -h & g^2 & g^2 & h^2 & g^2 & -gh & -gh \\ 1 & +g & +g & +h & g^2 & g^2 & h^2 & g^2 & gh & gh \\ 1 & -\alpha g & 0 & 0 & \alpha^2g^2 & 0 & 0 & 0 & 0 & 0 \\ 1 & +\alpha g & 0 & 0 & \alpha^2g^2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -\alpha g & 0 & 0 & \alpha^2g^2 & 0 & 0 & 0 & 0 \\ 1 & 0 & +\alpha g & 0 & 0 & \alpha^2g^2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -\alpha h & 0 & 0 & \alpha^2h^2 & 0 & 0 & 0 \\ 1 & 0 & 0 & +\alpha h & 0 & 0 & \alpha^2h^2 & 0 & 0 & 0 \\ 1 & 0 & 0 & -\varepsilon h & 0 & 0 & \varepsilon^2h^2 & 0 & 0 & 0 \\ 1 & 0 & 0 & +\varepsilon h & 0 & 0 & \varepsilon^2h^2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

As for the previous design possessing a complete ε-star portion, calculation of the elements of the moment matrix $N^{-1}\mathbf{X}'\mathbf{X}$ shows that this design meets the following requirements, for all $i < j < p < q$, as described in Section 5.2.

$$[ij] = [ijj] = [iij] = 0$$
$$[iijp] = [ijjp] = [ijpp] = 0$$
$$[iiij] = [ijjj] = 0$$
$$[ijpq] = [ijp] = 0$$

and that for any non-ε variable i,

$$[ii] = N^{-1}[Fg^2 + 2\alpha^2g^2] = g^2N^{-1}[F + 2\alpha^2]$$

whilst for the ε variable;

$$[ii] = N^{-1}[Fh^2 + 2\alpha^2h^2 + 2\varepsilon^2h^2] = h^2N^{-1}[F + 2\alpha^2 + 2\varepsilon^2]$$

so that the values of the scaling factors g and h have resulted in the required value of [ii] = 1.

This difference in scaling factors for the two classes of variable leads to complications in the assessment of rotatability and orthogonality. The values of the non-zero fourth-order moments are as follows.

$[iijj] = N^{-1} Fg^4$ between two non-ε variables;
$[iijj] = N^{-1} Fg^2h^2$ between ε and non-ε variables;

$[iiii] = N^{-1}[Fg^4 + 2\alpha^4 g^4]$ for non-ε variables;
$[iiii] = N^{-1}[Fh^4 + 2\alpha^4 h^4 + 2\varepsilon^4 h^4]$ for the ε variable;

Thus to satisfy the orthogonality requirement [iijj] = 1, two separate conditions must be met:

(i) between two non-ε variables

$$\frac{1}{N} Fg^4 = 1$$

$$\frac{1}{N} F\left[N / (F + 2\alpha^2)\right]^2 = 1$$

$$FN = (F + 2\alpha^2)^2 \qquad (7.8)$$

(ii) between ε and non-ε variables;

$$\frac{1}{N} Fg^2h^2 = 1$$

$$\frac{1}{N} F\left[N^2 / (F + 2\alpha^2)(F + 2\alpha^2 + 2\varepsilon^2)\right] = 1$$

$$FN = (F + 2\alpha^2)(F + 2\alpha^2 + 2\varepsilon^2) \qquad (7.9)$$

Clearly, (7.8) and (7.9) cannot simultaneously be true unless $\varepsilon = 0$

Similarly, for rotatability [iiii] = 3[iijj], and a design must meet the four conditions:

$$Fg^4 = 3\,(Fg^4 + 2\alpha^4 g^4)$$
$$Fg^4 = 3\,(Fh^4 + 2\alpha^4 h^4 + 2\varepsilon^4 h^4)$$
$$Fg^2h^2 = 3\,(Fg^4 + 2\alpha^4 g^4)$$
$$Fg^2h^2 = 3\,(Fh^4 + 2\alpha^4 h^4 + 2\varepsilon^4 h^4)$$

which again cannot be simultaneously fulfilled unless $\varepsilon = 0$.

On considering the expressions for the two scaling variables g and h, however, it is apparent that the difference between the two is purely due to the additional term $2\varepsilon^2$. Since the value of ε itself is chosen to be small in magnitude, it is reasonable to neglect the difference in scale factors. As an example, Table 7.1 shows the required values of g and h for the seven

variable case in which F = 64, N = F + 2n + 2, when $\alpha = 2.8284$ (standard CCD rotatability criterion) or 1.0, and $\varepsilon = 0.05$ or 0.10.

α	ε	g	h	g - h	g / h	(g - h)/g %
2.8284	0.05	1.00000	0.99997	3.125×10^{-5}	1.0000312	0.00312
1.0000	0.05	1.10096	1.10092	4.170×10^{-5}	1.0000379	0.00379
2.8284	0.10	1.00000	0.99988	1.250×10^{-4}	1.0001250	0.01250
1.0000	0.10	1.10096	1.10080	1.668×10^{-4}	1.0001515	0.01515

Table 7.1 Values of the scaling variables for a CCD with single ε-pair

There will clearly be little error in the calculation of parameter requirements for orthogonality and rotatability if identical values are chosen for g and h. Parameter values can thus be selected as for a standard CCD with $2 + n_0$ centre points. The inclusion of an exact centre point in the design, thus incrementing the value of N by 1, has no effect on the ratio of the two scaling factors, and no significant effect on the magnitude of the difference between them unless (F + 2n + 2) is small: even with n = 3, $F = 2^3 = 8$, the change in (g - h) is $(17/16)^{1/2}$, or just 3.078 %, when the exact centre point is added.

The value of $\varepsilon = 0.10$ used in the above calculations might be considered to be a typical upper limit for the approximation of centre points. The difference in scaling factors is still negligible at this value, however, although its effect will of course increase with smaller values of F, resulting from either fewer variables or a lower resolution design. If a limiting value for (g - h)/g % were specified as, say, 0.1%, then calculation shows that, in the present example, a value of $\varepsilon = 0.2829$ is acceptable with $\alpha = 2.8284$, and $\varepsilon = 0.2570$ with $\alpha = 1.0$. Thus even this very tight constraint still allows ε points which are between a quarter and a third of the distance to the variable bound. Since this value will necessarily fall with decreasing F, it is appropriate to consider a severe case of n = 3, $F = 2^3 = 8$, $\alpha = 1.0$. For this design the 0.1% limit still allows a value of $\varepsilon = 0.100025$, and so it may be concluded that in general the introduction of an ε-star portion leads to insignificant changes in the scaling of the variables.

In summary, the results of this investigation lead to the same conclusions as for the CCD with complete ε-star portion, that the practical choice of α for orthogonality or rotatability is unchanged from that for the standard CCD. In this case, however, the value of α would be that chosen for a design with either 2 or 3 centre points, depending on whether a centre point test is included in the present design. Further investigation reveals the additional similarity, that designs which are both orthogonal and rotatable are not generally obtainable, and that neither property may be achieved if an axial parameter of $\alpha = 1.0$ is employed.

In the above analysis it was shown that the multiple criteria for orthogonality occur as a result of the different scaling factors which must be applied to different variables, in order to meet the requirement that [ii] = 1. An alternative method of meeting this requirement is to modify the axial portion of the design by choosing different values of the parameter α for the ε variable and non–ε variables, so that equal scaling factors are achieved for each variable. The advantage of a single scaling factor is that a single orthogonality criterion is obtained, with two rotatability criteria replacing the four obtained above. A detailed analysis of this alternative method of specifying the single ε–pair design is presented in Appendix 7B. The results of this analysis show that the difference between the two values of α is negligible for all acceptable values of ε, and hence confirm the conclusion that the parameters for the CCD with one additional ε pair may be specified as for the appropriate standard CCD.

7.4 Designs containing several ε pairs

Detailed analyses have now been presented for augmented Central Composite Designs in n variables for which the number of ε variables is zero, n or 1 (Chapter 6, Section 7.2 and Section 7.3 respectively). Assessment of the effect of the inclusion of ε pair tests on the rotatability and orthogonality characteristics of the CCD has been based solely on the criteria

$$[iijj] = 1 \qquad \text{for orthogonality}$$
$$[iiii] = 3[iijj] \qquad \text{for rotatability.}$$

Inspection of these criteria reveals that, irrespective of the total number of variables in the design, orthogonality and rotatability may be assessed by just considering combinations of pairs of variables. The above analyses have also shown that all non-ε variables will have the same pure second order moment [ii] as each other, as will all ε variables. When analysing a general CCD with $1 \leq k \leq n$ ε pairs, there are thus only three possible values of the mixed fourth order moment [iijj], in which either both variables are ε variables, both are non-ε variables, or one variable is of each type. Each of these combinations has been considered in the preceding analyses, so that the conclusions drawn concerning the CCD with one ε pair and the CCD with a full ε star may be generalised to designs in which $1 \leq k \leq n$ ε pairs are present.

7.5 Observations on the inclusion of ε-pair tests

Although the preceding analyses have shown that the value of the parameter ε must be relatively small (less than about 0.1) in order not to invalidate either the approximation to the centre point, or the comparison with the standard CCD, no further guidance as to a suitable choice has emerged. There remains the issue of what a lower bound to this value may be which still genuinely provides additional information on the variation of the response functions. This lower bound is likely to be largely dependent on the nature of the problem being addressed, and general recommendations are unlikely to prove reliable. It is recommended that if this class of design is to be employed, some preliminary investigation of the nature of the response surface for the particular application under consideration should be carried out prior to the execution of the main body of test work. Such an investigation into the variation around the centre of the design variable space of the radiated noise level predicted by the FE analysis program is presented in Section 7.6.2. The results of this investigation suggest that the value of the parameter ε may be reduced to at least as low as 0.025 without encountering the problem of identical test results which the use of the ε-star portion is intended to address. The choice of value within this range is discussed in Section 7.6.1, below, with reference to numerical examples.

The analyses of Sections 7.2 and 7.3 have shown that for an n-dimensional CCD, the inclusion of either 1 or n pairs of axial tests with parameter $\varepsilon \leq 0.1$ may be carried out without the need to modify either the value of the axial parameter α or the way in which the variable values are scaled. Orthogonality and rotatability may be assessed with little error by simply treating the design as a standard CCD in which the number of centre point tests n_0 is equal to the number of tests included in the ε-star portion, plus the exact centre point, if this is included. As described in Section 7.4, these conclusions may be generalised to cover all designs containing between 1 and n pairs of axial tests with parameter $\varepsilon \leq 0.1$.

The flexibility to specify the number of ε pairs to be included allows for a procedure analogous to that employed with the standard CCD, in which the number of centre point tests may be chosen in order to aid the fulfilment of, for example, orthogonality or rotatability criteria. Having ascertained from the analysis of a standard CCD the number of centre point tests which is required, the appropriate number of ε pairs can be included in the design in place of the replicate points. As an example, consider again the seven variable example with F = 64. If an axial parameter of $\alpha = 2.257$ were employed, then equation (6.7) shows that a total of 8 centre point tests would be required in order to attain rotatability of the design. When investigating a deterministic system, this requirement for replicate centre points may be approximately met by the inclusion of four pairs of ε-star points, with no exact centre point test. The choice of which axes the additional tests should lie on is unlikely to significantly alter the accuracy with which the model is constructed, due to the proximity of the points involved.

The only restriction on the selection of ε-points in order to fulfil a particular requirement for centre point tests is that both points of a pair must appear in the design. If an odd number of centre points is required, then the additional test may be supplied by including a single test at the exact centre of the design variable space.

As a further example, the condition that must be fulfilled in order to obtain a standard CCD which is both orthogonal and rotatable is given in equation (6.10). This shows that the seven variable problem with F = 64 requires a centre portion of $n_0 = 22$. Since the total number of points available within the ε-star is just 14, it is clearly not possible to meet this requirement using any of the designs so far described. Examination of the analysis of the previous sections reveals, however, that the construction of the standard portion of the CCD will need no modification as long as the ratio $\varepsilon^2 : \alpha^2$ remains small. Thus, in order to achieve the required centre point tests, two separate ε-star portions may be specified, with parameters ε_1 and ε_2 respectively, with ε-pairs selected from them in any combination up to the total number required. If, in general, *s* ε-star portions are included in the design, then the standard portion of the CCD will need no modification as long as the ratio $(\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_s^2) : \alpha^2$ remains small. The values of $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_s$ must also be selected in such a way that each lies within the allowed range for ε, and that the difference between them is also greater than the permitted lower bound. Together with the above ratio to the axial parameter α, these conditions may be expressed, for two ε-star portions, as follows.

$$\varepsilon_{min} \leq \varepsilon_1 , \varepsilon_2 \leq \varepsilon_{max}$$
$$| \varepsilon_1 - \varepsilon_2 | \geq \varepsilon_{min}$$
$$(\varepsilon_1^2 + \varepsilon_2^2) : \alpha^2 < 1\ \%, \text{ say}$$

Even for the case of α = 1 0, this last condition will always be met if both ε parameters are smaller than 0.07. Alternatively, one possible combination would be $\varepsilon_1 = 0.075$ and $\varepsilon_2 = 0.025$, giving $(0.075^2 + 0.025^2) : 1.0 = 0.625\%$. The requirement of $n_0 = 22$ could thus be met using a full ε star with parameter ε = 0.075, and four ε pairs with parameter ε = 0.025.

As noted in Section 6.4, the nature of the variables under investigation in the present application leads to a strictly cuboidal design variable space, with the result that the axial parameter is constrained to α = 1. The consequence of this is that the noise simulation problem, as currently posed, cannot be investigated with a standard CCD which is either orthogonal or rotatable. The analysis of the present chapter has shown that the characteristics of the standard portion of the augmented CCD under consideration here are unaffected by the addition of an ε star portion. It follows that the designs of this chapter share with the standard CCD the feature that the properties of orthogonality and rotatability are both unattainable with α = 1.

Since the exact fulfilment of these criteria is no longer an objective, the precise number of simulated replications which are to be performed need not be specifically tailored to achieve this goal. The aim in carrying out replicated centre points is now to collect additional information on the variation of the response functions at or around the centre of the design variable space. These data may then be used to obtain improved estimates of the pure quadratic components of a fitted model, and ultimately to develop an enhanced predictive model of the original surface throughout the region of interest. There is now no clear indication, however , as to the preferred number of tests to be included in order to produce such an improvement, although an initial consideration would suggest that the predictive accuracy will increase with the number of points tested.

7.6 Numerical tests using the CCD with additional ε portion

In order to investigate the many issues arising from the above discussions, a series of numerical tests has been carried out, again using the FE model of Appendix 1C as an example. To enable comparison with those tests involving the standard CCD which were carried out in the previous chapter, the same seven variable example has been used, with a 2^{7-1} factorial portion containing 64 tests, and an axial parameter of $\alpha = 1$.

In Section 7.6.1 numerical trials are carried out using a CCD with full ε star. A number of designs are investigated in which the parameter ε takes values in the range 0.025 to 0.1. The results of these tests show that there is no practical improvement in the predictive ability of the fitted model, and little change in the coefficient values of the model. There is a substantial reduction in the variance of the mean value, but no significant reduction in the variance of the pure quadratic terms, whilst the covariance between pairs of quadratic terms shows a slight rise.

In Section 7.6.2 it is shown that the reason for the lack of improvement in model fit on inclusion of the ε star tests is that the fit of the model is already good in this region. It is also demonstrated that genuine additional information concerning the original response function may be gained by testing at simulated centre-point replicates which have parameter $\varepsilon \geq 0.025$.

The inclusion of just one pair of ε tests is investigated in Section 7.6.3. Designs are tested in which each of the seven variables are included as the ε pair. The results of these trials confirm that the effect of including a single ε pair is substantially less than the effect of the full ε star.

7.6.1 The CCD with ε star

To assess the performance of the experimental designs described in Section 7.2, constructed using an additional star portion with axial parameter ε, a selection of such designs has first been investigated. Four different values of the parameter ε were chosen: 0.025, 0.05, 0.075 and 0.1, and tests conducted both with and without an exact centre point. Each of the resulting designs were used to fit the models of Sections 6.5.1 - 6.5.3, containing 71, 17 and 36 parameters, respectively, and the results of these tests then compared with those obtained using the standard CCD, as described in Section 6.5.1. Comparison was also made with designs containing the same number of exact centre point replications as contained in the ε-star portion (plus centre point, if applicable) of the modified designs, as discussed in Sections 6.4 and 6.5.4.

For the 36 coefficient strict quadratic model of Section 6.5.3, Table 7.4 shows the maximum and average lack of fit, for each design, which occurs within the three categories of test points described in Section 6.5.1. Each of the modified designs of Table 7.4 includes one exact centre point ($n_0 = 1$). The first line of the table refers to the standard CCD of Section 6.5.3, containing a single centre point, with the second line of data referring to a CCD which includes a total of $2n + n_0 = 15$ exact centre point tests (identical to the 15 centre point design used in Section 6.5.4). The corresponding lack of fit data for models containing 71 and 17 parameters are given in Appendix 7C, together with the results obtained with each model when no exact centre point is specified in the modified designs.

	Maximum lack of fit — E_R			Average lack of fit — E_R		
ε	Design	Factorial	Higher	Design	Factorial	Higher
standard	4.0151	2.6508	6.2354	0.67564	0.85803	2.6048
0.0	4.0190	2.6547	6.3404	0.58955	0.85819	2.6000
0.025	4.0195	2.6553	6.3491	0.59321	0.85819	2.5995
0.05	4.0185	2.6543	6.3192	0.60006	0.85814	2.6006
0.075	4.0183	2.6545	6.3255	0.60159	0.85819	2.6001
0.1	4.0180	2.6549	6.3244	0.60827	0.85822	2.5998

Table 7.4 Variation of lack of fit with ε-star parameter for augmented CCD with 93 tests and 36 coefficients ($n = 7$, $n_0 = 1$)

The results of Table 7.4 show that the prediction accuracy using each of the modified designs is almost identical to that achieved using the standard CCD. Taking as an example the average lack of fit in the third category of test points ('Higher'), an initial examination suggests that a slight improvement has been achieved by the use of each of the modified designs, with the

design in which $\varepsilon = 0.025$ performing best. The difference between the lack of fit value for this design and the result obtained using the standard CCD is, however, just 0.0053% of the variable range, with only 0.0011% of the range separating the best and the worst of the modified designs. Thus, the inclusion of the additional ε-star portion has not led to any practical improvement in the ability of the predictive model to represent the variation of the computed noise function within the design variable space. This feature is also displayed by each of the result sets given in Tables 7C.1 to 7C.6 of Appendix 7C.

The fact that the inclusion of the additional simulated centre point replicates has little effect on the performance of the model throughout the region of interest suggests that the parameter estimates for each of the coefficients within the model will also be largely unchanged by the introduction of the ε-star portion. Indeed, since each of the additional points has just one variable at a non-zero value, the estimates of all linear interaction terms, of which the 36 term model contains 21 two-way interaction terms, must be identical to those obtained with the standard CCD. The only parameters to be affected are the mean, main effect and pure quadratic terms. Additionally, only two of the ε-star tests will contribute to the estimation of each of the main effect parameters, with the calculation of these being dominated by the 64 tests of the factorial portion of the design. The extent to which the parameter estimates of the mean, main effect and pure quadratic terms are modified by the inclusion of the additional data is demonstrated in Table 7.5, in which these terms are compared for the standard CCD and the ε–star design with $\varepsilon = 0.1$.

Term	Standard CCD	$\varepsilon = 0.1$
MEAN	86.711	86.721
A	-2.2352	-2.2351
B	-0.20402	-0.20404
C	-0.38703	-0.38704
D	-0.22783	-0.22793
E	-0.15832	-0.15830
F	-0.23327	-0.23334
G	-0.33550	-0.33553
A^2	0.98271	0.98129
B^2	0.02250	0.02113
C^2	0.06593	0.06458
D^2	0.19729	0.19589
E^2	0.05800	0.05667
F^2	-0.07867	-0.07999
G^2	0.11884	0.11747

Table 7.5 Comparison of parameter estimates between standard CCD and CCD with ε-star portion ($\varepsilon = 0.1$) and $n_0 = 1$: 36 coefficients

This table confirms that the main effect terms are virtually unchanged, with their values only differing at the fourth or fifth significant figure. Changes to the mean and pure quadratic coefficients, although greater, are still relatively minor. The variance and covariance values for each class of terms are given in Tables 7.6 and 7.7, below.

Class of Term	Standard CCD	$\varepsilon = 0.1$
Mean	0.0158	0.0109
Main effects	0.0152	0.0151
Linear interactions	0.0156	0.0156
Pure quadratic	0.1915	0.1910

Table 7.6 Comparison of variance values between standard CCD and CCD with ε-star portion ($\varepsilon = 0.1$) and $n_0 = 1 : 36$ coefficients

Class of Term	Standard CCD	$\varepsilon = 0.1$
between mean and each pure quadratic term	-0.0018	-0.0002
between each pair of pure quadratic terms	-0.0308	-0.0312

Table 7.7 Comparison of covariance values between standard CCD and CCD with ε-star portion ($\varepsilon = 0.1$) and $n_0 = 1 : 36$ coefficients

These tables show that, when using a modified design with $\varepsilon = 0.1$, the variance value for the mean term has been greatly reduced, compared with that obtained using the standard CCD, reflecting the gain in precision which results from using an increased number of test points from which to estimate the parameter value. In contrast, the variance values for the pure quadratic terms remain virtually unchanged by the increase in information, whilst the covariance effects between them actually increase. The reason for this is that the additional test points provide function information at a point at which all but one of the variables are at their mean level, with the remaining variable only slightly altered from its mean value. It is thus extremely hard to identify explicitly the effect of each of the variables. Indeed, a larger proportion of the available third level tests now lies at or close to the centre of the design space, and this leads to greater interdependency of the parameter estimates, and hence higher covariance values. The change in the parameter estimates and variance values of the main effect and pure quadratic terms is not large enough to have any appreciable effect on the outcome of a probability plot analysis, which, using the full 71 parameter model, produces results which are essentially identical to those of Section 6.5.1.

Comparison with Table 6B.8 of Appendix 6B shows that the changes in variance and covariance values are almost identical to those produced using a standard CCD with 15 exact centre point tests. In contrast with the results of Appendix 6B, however, these are authentic changes caused by the genuine inclusion of additional function information, whereas those of Appendix 6B were illusory gains resulting from the inappropriate use of the standard analysis procedure, based upon an invalid assumption of random experimental error.

7.6.2 Investigation of noise variation at ε-star points

In order to determine the reason for the relatively poor improvement in the performance of the mathematical model which has resulted from the inclusion of the ε-star portion, it is necessary to investigate the relationship, at each of the additional test points, between the value of the noise function calculated by the FE analysis and the response surface prediction obtained using just a standard CCD. As an example, Figure 7.1 shows the comparison between these functions at the centre point of the design variable space ($\varepsilon = 0$), and at each of the ε-star points for which the variable A (skirt thickness) has a non-zero value. The four values of ε chosen are those used in the analysis of Section 7.6.1, above. The predictions at these points are obtained using the 36 parameter strict quadratic model described previously, constructed using a 79 test CCD with $\alpha = 1$ and $n_0 = 1$.

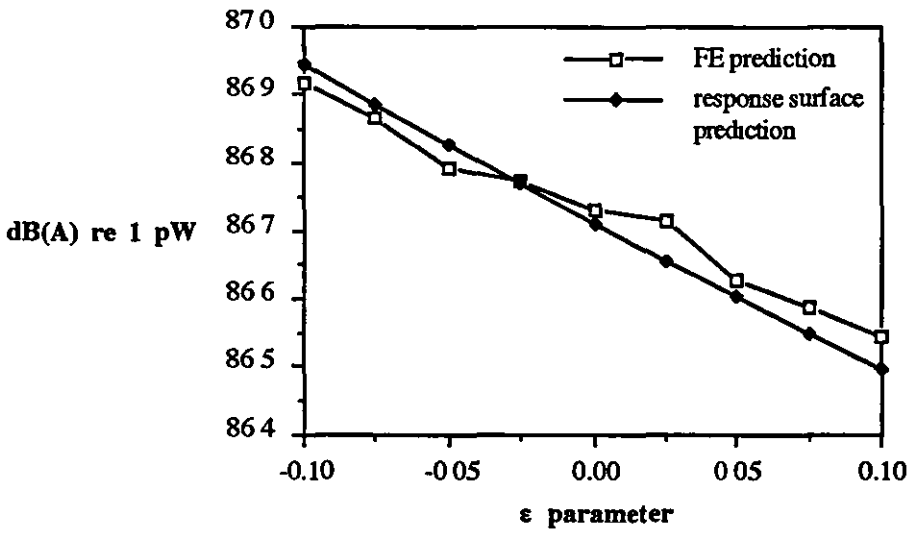

Figure 7.1 Comparison of FE analysis noise value with that predicted using a standard CCD with 79 tests and 36 coefficients ($n = 7$, $\alpha = 1$, $n_0 = 1$) at ε-star points with variable A non-zero

This graph shows that, over the range investigated, the slope of the predictive model is closely approximating that of the true response, with little error occurring at the ε-star points, even with $\varepsilon = 0.1$. The graph shows the middle 10% of the bound-to-bound variable range of the skirt thickness variable. Since there is such close correspondence between measured and predicted values, it is unsurprising that the approximating model is little changed by the inclusion of a pair of these data points.

This effect can be seen more clearly by considering the magnitude of the errors which occur at each of the ε-star points. Figure 7.2 shows the error values, in dB(A), for each of the points of Figure 7.1.

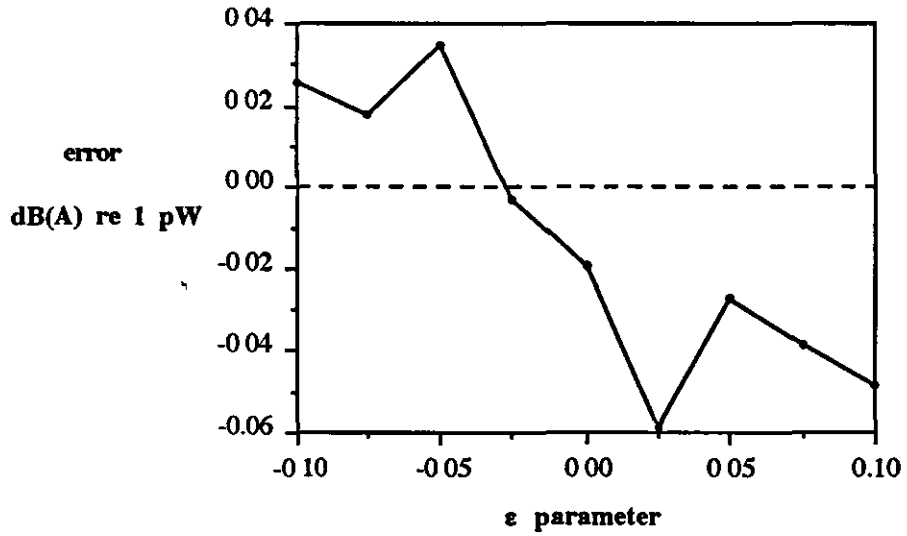

Figure 7.2 Prediction error using a standard CCD with 79 tests and 36 coefficients ($n = 7$, $\alpha = 1$, $n_0 = 1$) at ε-star points with variable A non-zero

The maximum error occurring at any of the ε-star points is thus of magnitude 0.06 dB(A), which is less than 0.8% of the function range. Comparing this with the lack of fit data given in the first line of Table 7.4, it can be seen that even this maximum error at the ε-star points is only slightly greater than the average error at the design points, and significantly less than the maximum error at these locations. This, then, is the reason that little modification to the parameter values is required in order to minimise the error at the design points, using a least squares criterion, when the additional ε-star points are introduced.

It is noticeable from Figure 7.2 that the maximum error occurring at the ε-star points with variable A non-zero (the dominant variable in the model), is obtained with a value of

$\varepsilon = 0.025$, which is also the value which yields the greatest change in overall prediction error (Table 7.4) and parameter values.

The above analysis may also be repeated for each set of ε-star points in which one of the other variables is at a non-zero level. These confirm that there is little lack of fit between the FE analysis value and the prediction based on the standard CCD at these points, and hence only minor modifications will be made to the model parameters when such points are included in the experimental design. As an example, Figure 7.3 shows the measured and predicted values for the variable C (bearing panels 2 & 4), in which the maximum error is 0.021 dB(A), which is just over 0.25% of the function range.

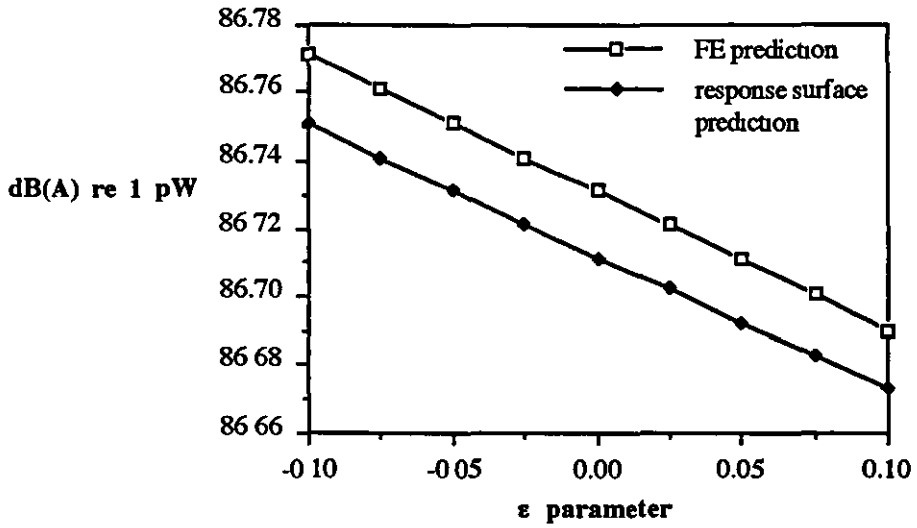

Figure 7.3 Comparison of FE analysis noise value with that predicted using a standard CCD with 79 tests and 36 coefficients ($n = 7$, $\alpha = 1$, $n_0 = 1$) at ε-star points with variable C non-zero

In addition to explaining the small improvement gained by including the ε-star points, these tests also demonstrate that it is possible to obtain genuine additional information concerning the variation of the noise function by carrying out tests close to the centre of the design variable space. This observation is valid for tests with an ε value as low as $\varepsilon = 0.025$.

7.6.3 The CCD with single ε pair

It has been shown in Section 7.6.1 that the inclusion of a complete ε-star portion in the experimental design leads to little improvement in the overall predictive ability of the mathematical model constructed from this design. It is therefore unlikely that the designs

outlined in Section 7.3, which include just two of the original 2n ε-star points, will prove to be of any practical use in addressing the noise simulation problem. Since the number of additional tests is now much smaller, the effect on the variance value is expected to be greatly reduced, with all parameter estimates remaining essentially unchanged from those of the standard CCD. In order to confirm this, and to investigate the effect of using ε pairs with different selections of the non-zero parameter, the following tests have been conducted. Seven experimental designs have been constructed, each containing the 79 tests of the standard CCD, as described previously, plus two additional points lying at $\pm\varepsilon = 0.025$ on one of the variable axes, to give a total of 81 test points. Each of these designs has then been used to fit the 36 term strict quadratic model used above, and the results compared with both the standard CCD and the CCD with full ε-star portion ($\varepsilon = 0.025$) described in the previous section.

The results of the lack of fit calculations in each of the three test categories are shown in Table 7.8. The first line of this table refers to the results obtained using the standard CCD. The next seven lines refer to the ε-pair designs currently being investigated, with the first column indicating the variable which takes a non-zero value in the two extra tests. The final line shows the results obtained using the full ε-star portion with parameter $\varepsilon = 0.025$, and is as shown in Table 7.4.

	Maximum lack of fit — E_R			Average lack of fit — E_R		
Variable	Design	Factorial	Higher	Design	Factorial	Higher
standard	4.0151	2.6508	6.2354	0.67564	0.85803	2.6048
A	4.0171	2.6528	6.2808	0.66716	0.85808	2.6026
B	4.0161	2.6519	6.2628	0.66363	0.85807	2.6035
C	4.0162	2.6519	6.2640	0.66385	0.85807	2.6035
D	4.0161	2.6519	6.2647	0.66398	0.85808	2.6034
E	4.0161	2.6519	6.2641	0.66387	0.85808	2.6034
F	4.0161	2.6520	6.2643	0.66390	0.85808	2.6034
G	4.0162	2.6519	6.2637	0.66381	0.85807	2.6035
full star	4.0195	2.6553	6.3491	0.59321	0.85819	2.5995

Table 7.8 Variation of lack of fit with ε-pair variable for augmented CCD with 81 tests and 36 coefficients ($n = 7$, $n_0 = 1$, $\varepsilon = 0.025$)

These lack of fit data show the expected result that the inclusion of an ε pair in the design yields virtually no improvement in predictive accuracy. Additionally, it can be seen that, although the use of an ε pair using the dominant variable A gives marginally better results in terms of minimising the average lack of fit at the 'Higher' test points, this design performs

worst in other categories, and in general there is no practical advantage in selecting a particular variable in preference to any of the others.

In analysing the variance values for each of the designs, it is found that, as may be expected, the variance value for the ε variable is the same in each design, as are the values for each of the non-ε variables. A similar correspondence is found between covariance values. These values are summarised in Tables 7.9 and 7.10, and are valid for all seven of the ε-pair designs.

Class of Term	Standard CCD	ε pair
Mean	0.015773	0.014434
Main effect (ε)	0.015152	0.015151
Main effect (non-ε)	0.015152	0.015152
Linear interaction	0.015625	0.015625
Pure quadratic (ε)	0.19147	0.19134
Pure quadratic (non-ε)	0.19147	0.19134

Table 7.9 Comparison of variance values between standard CCD and CCD with ε pair (ε = 0.025) and n_0 = 1 : 36 coefficients

Class of Term	Standard CCD	ε pair
mean / pure quadratic (ε)	-0.001757	-0.001351
mean / pure quadratic (non-ε)	-0.001757	-0.001341
pure quadratic (ε) / pure quadratic (non-ε)	-0.030754	-0.030881
pure quadratic (non-ε) / pure quadratic (non-ε)	-0.030754	-0.030884

Table 7.10 Comparison of covariance values between standard CCD and CCD with ε pair (ε = 0.025) and n_0 = 1 : 36 coefficients

These results confirm that the variance of the mean effect is altered much less than was the case for the ε-star design. All other variance and covariance effects are also largely unchanged from the values obtained using the standard CCD.

Table 7.11 shows the parameter estimates which are obtained when fitting the 36 term model to the standard CCD, the CCD with single ε pair in variable A (ε = 0.025), and the CCD with full ε star (ε = 0.025). The linear interaction terms are not shown, as these are identical in each case. The table shows that the degree of modification which occurs in each parameter

estimate, compared with the standard CCD, increases as the number of additional tests rises. As discussed in the previous section, the main effect terms are virtually unchanged by the small amount of additional data, and even the mean and quadratic terms are little modified by the inclusion in the experimental design of either a single ε pair or a complete ε-star portion.

Term	Standard CCD	ε pair (A)	full star
MEAN	86.711	86.716	86.723
A	-2.2352	-2.2352	-2.2352
B	-0.20402	-0.20402	-0.20402
C	-0.38703	-0.38703	-0.38703
D	-0.22783	-0.22783	-0.22784
E	-0.15832	-0.15832	-0.15832
F	-0.23327	-0.23327	-0.23328
G	-0.33550	-0.33550	-0.33550
A^2	0.98271	0.98204	0.98101
B^2	0.02250	0.02182	0.02080
C^2	0.06593	0.06524	0.06215
D^2	0.19729	0.19660	0.19558
E^2	0.05800	0.05732	0.05629
F^2	-0.07867	-0.07936	-0.08038
G^2	0.11884	0.11816	0.11713

Table 7.11 Comparison of parameter estimates between standard CCD, CCD with ε pair (ε = 0.025, variable A) and n_0 = 1 and CCD with ε-star portion (ε = 0.025) and n_0 = 1: 36 coefficients

7.7 Observations on the use of simulated centre point replications

The numerical results of the previous sections have shown that it is possible to obtain genuine additional information concerning the variation of the noise function by performing tests at points close to the centre of the design variable space. The use of an ε-star portion has been shown to yield a similar small increase in the accuracy of the fitted model as was obtained by simply including the same number of exact centre point replicates, with the advantage that the present results represent a real improvement in model performance, rather than an illusory gain based on the adoption of invalid error assumptions.

Additionally, although precluded in the present noise analysis example by the cuboidal nature of the design variable region, theoretical studies have shown that in general deterministic

experimental environments, the specification of an ε-star or ε-pair portion can be made without significantly compromising the orthogonal or rotatable nature of a standard CCD.

In the specific case investigated, little improvement in model accuracy has been obtained over that of the approximating model derived from the standard CCD, due to the small lack of fit which occurs at points close to the centre of the region. In such cases the advantage to be gained from including centre-point pseudoreplicates is minimal. If, however, the original response surface exhibited a substantial amount of high-frequency fluctuation, or 'peakiness', then repeated testing over a small area would be likely to result in a significant gain in the predictive accuracy of the fitted model. These multiple tests would provide a better picture of the overall trend of variation in the response function, rather than relying on data collected at a single point, which may be excessively influenced by the peaks and troughs of a heavily fluctuating surface. For the particular case of radiated noise, such a response may result from the use of a low value of structural damping within the engine model, as discussed by Milsted, Zhang and Hall (1993).

The numerical examples of the previous sections have shown, however, that if the response is smooth, with a close correspondence between the original response and the prediction based on the standard CCD, then little is to be gained from this type of multiple testing. In this case, the most effective use of additional test points is not to provide an increase in the density of testing in certain areas, but to improve the distribution of test points throughout the entire design variable space. A modified design could, for example, include some fraction of a second 2^n hypercube, with levels ±1/2. These points could be specified using a procedure similar to that followed in Section 6.5.1 to select additional points at which to test the lack of fit due to higher order terms. A design of this type would also allow the possibility of retaining the properties of orthogonality and rotatability.

The effect of such a modification, however, would be to substantially increase the total test requirement for the design. It should be borne in mind that the results of the previous chapter have shown that, even for the standard CCD, the ratio of significant terms to number of tests, for the seven variable example, may be as small as $17 / 79 \approx 22$ %, and is unlikely to be greater than $36 / 79 \approx 46$ %. With a full ε–star portion this falls to $36 / 93 \approx 39$ %, whilst with a second hypercube containing as few as 16 points, the ratio is $36 / 95 \approx 38$ %. If a second hypercube were to be included, then one could perhaps reconsider the function of the original star portion, since tests at additional variable levels would now be available. However, a design in which this level of modification has taken place is, of course, no longer a CCD. A further disadvantage of the CCD which needs to be addressed is the relative imprecision with which the pure quadratic components of the second order model are estimated, as discussed in detail in Section 6.5.1. What is required is an economic second order design, constructed specifically to estimate just the parameters of the strict quadratic model, and to estimate these with approximately equal accuracy. Such designs are considered in detail in the next chapter.

Appendix 7A

Values of α^2 and ε^2 for both orthogonality and rotatability of CCD with ε-star portion

n	F	Resolution	n_0	N	α^2	ε^2
3	8	Full	0	20	2.79	-0.465
3	8	Full	1	21	2 81	-0.329
4	16	Full	0	32	3.95	-0.635
4	16	Full	1	33	3 97	-0 482
5	32	Full	0	52	5 54	-1.14
5	32	Full	1	53	5 57	-0.980
5	16	V	0	36	4 00	0 000
5	16	V	1	37	4 00	0 169
6	64	Full	0	88	7.70	-2.18
6	64	Full	1	89	7.74	-2 01
6	32	VI	0	56	5 64	-0 471
6	32	VI	1	57	5 65	-0.295
7	128	Full	0	156	10 6	-3 95
7	128	Full	1	157	10 7	-3.78
7	64	VII	0	92	7 86	-1.49
7	64	VII	1	93	7 89	-1.32
8	256	Full	0	288	14 5	-6.74
8	256	Full	1	289	14 6	-6.58
8	128	VIII	0	160	10 8	-3.28
8	128	VIII	1	161	10 9	-3 10
8	64	V	0	96	7.96	-0.771
8	64	V	1	97	7 98	-0 583
9	512	Full	0	548	19 8	-11 0
9	512	Full	1	549	19.9	-10.8
9	256	IX	0	292	14 8	-6 09
9	256	IX	1	293	14 9	-5 92
9	128	VI	0	164	11 0	-2 57
9	128	VI	1	165	11.1	-2.39
10	1024	Full	0	1064	27.0	-17.1
10	1024	Full	1	1065	27.1	-17.0
10	512	X	0	552	20.1	-10.3
10	512	X	1	553	20.2	-10 2
10	256	VI	0	296	15 1	-5 42
10	256	VI	1	297	15.1	-5.25
10	128	V	0	168	11.2	-1 84
10	128	V	1	169	11.2	-1 65
11	2048	Full	0	2092	37.0	-26.1
11	2048	Full	1	2093	37 1	-25 9
11	1024	XI	0	1068	27 4	-16 5
11	1024	XI	1	1069	27.5	-16.4
11	512	VII	0	556	20 5	-9 68
11	512	VII	1	557	20 5	-9 52
11	256	V	0	300	15 3	-4.72
11	256	V	1	301	15 3	-4.55
11	128	V	0	172	11 3	-1.07
11	128	V	1	173	11.3	-0 876
12	4096	Full	0	4144	50 8	-38 9
12	4096	Full	1	4145	50 9	-38 7
12	2048	XII	0	2096	37 4	-25 5
12	2048	XII	1	2097	37.5	-25.3
12	1024	VIII	0	1072	27.8	-15.9
12	1024	VIII	1	1073	27.9	-15.7
12	512	VI	0	560	20.8	-9.02
12	512	VI	1	561	20 8	-8 85
12	256	V	0	304	15 5	-4 01
12	256	V	1	305	15 5	-3 82

Appendix 7B

The CCD with one ε pair and equal scaling

In order to satisfy the moment matrix condition [ii] = 1, whilst avoiding the need for two different scaling factors, the value of the axial parameter α of the original star portion for the ε variable may be modified from α to γ. The corresponding design matrix for a three variable example is then of the form

$$\mathbf{D} = \begin{bmatrix} -g & -g & -g \\ -g & -g & +g \\ -g & +g & -g \\ -g & +g & +g \\ -g & -g & -g \\ -g & -g & +g \\ -g & +g & -g \\ -g & +g & +g \\ -\alpha g & 0 & 0 \\ +\alpha g & 0 & 0 \\ 0 & -\alpha g & 0 \\ 0 & +\alpha g & 0 \\ 0 & 0 & -\gamma g \\ 0 & 0 & +\gamma g \\ 0 & 0 & +\varepsilon g \\ 0 & 0 & -\varepsilon g \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} \left.\vphantom{x}\right\} F \\ \\ \\ \\ \\ \\ \\ \\ \left.\vphantom{x}\right\} 2n \\ \\ \\ \\ \\ \\ \left.\vphantom{x}\right\} 2 \\ \left.\vphantom{x}\right\} n_0 \end{matrix} \qquad (7B.1)$$

For the non-ε variables the value of the scale factor g is again

$$g = \left[N / (F + 2\alpha^2)\right]^{\frac{1}{2}}$$

whilst for the ε variable its value must be

$$g = \left[N / (F + 2\gamma^2 + 2\varepsilon^2)\right]^{\frac{1}{2}}$$

Equating these two expressions yields

$$\gamma^2 + \varepsilon^2 = \alpha^2$$

and hence

$$\gamma = \sqrt{\alpha^2 - \varepsilon^2} \qquad (7B.2)$$

The appropriate value of γ can then be calculated for known values of α and ε. For the seven variable case, for example, the following would be applicable

α	ε	γ
2.8284	0.05	2 8280
2.8284	0.10	2 8267
1.0000	0 05	0 9987
1.0000	0 10	0 9950

Table 7B.1 Values of α and γ for a CCD with one ε-pair and equal scaling

This example illustrates the general feature that for the small values of ε required to simulate centre point replications the difference of ε^2 between the values of α^2 and γ^2 is of small magnitude.

The regressor matrix for this design is then of the form

$$
\mathbf{X} = \begin{array}{c}
\begin{array}{cccccccccc} 1 & x_1 & x_2 & x_3 & x_1^2 & x_2^2 & x_3^2 & x_1x_2 & x_1x_3 & x_2x_3 \end{array} \\
\left[\begin{array}{cccccccccc}
1 & -g & -g & -g & g^2 & g^2 & g^2 & g^2 & g^2 & g^2 \\
1 & -g & -g & +g & g^2 & g^2 & g^2 & g^2 & -g^2 & -g^2 \\
1 & -g & +g & -g & g^2 & g^2 & g^2 & -g^2 & g^2 & -g^2 \\
1 & -g & +g & +g & g^2 & g^2 & g^2 & -g^2 & -g^2 & g^2 \\
1 & +g & -g & -g & g^2 & g^2 & g^2 & -g^2 & -g^2 & g^2 \\
1 & +g & -g & +g & g^2 & g^2 & g^2 & -g^2 & g^2 & -g^2 \\
1 & +g & +g & -g & g^2 & g^2 & g^2 & g^2 & -g^2 & -g^2 \\
1 & +g & +g & +g & g^2 & g^2 & g^2 & g^2 & g^2 & g^2 \\
1 & -\alpha g & 0 & 0 & \alpha^2 g^2 & 0 & 0 & 0 & 0 & 0 \\
1 & +\alpha g & 0 & 0 & \alpha^2 g^2 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & -\alpha g & 0 & 0 & \alpha^2 g^2 & 0 & 0 & 0 & 0 \\
1 & 0 & +\alpha g & 0 & 0 & \alpha^2 g^2 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & -\gamma g & 0 & 0 & \gamma^2 g^2 & 0 & 0 & 0 \\
1 & 0 & 0 & +\gamma g & 0 & 0 & \gamma^2 g^2 & 0 & 0 & 0 \\
1 & 0 & 0 & -\varepsilon g & 0 & 0 & \varepsilon^2 g^2 & 0 & 0 & 0 \\
1 & 0 & 0 & +\varepsilon g & 0 & 0 & \varepsilon^2 g^2 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]
\end{array}
$$

Calculation of the elements of the moment matrix $N^{-1}\mathbf{X}'\mathbf{X}$ shows that this design also meets the following requirements of Section 5.2, for all $i < j < p < q$

$$[ij] = [iij] = [iij] = 0$$
$$[iijp] = [ijjp] = [ijpp] = 0$$
$$[iiij] = [ijjj] = 0$$
$$[ijpq] = [ijp] = 0$$

It can also be seen that the values of the pure second-order moments are as follows:

$[ii] = N^{-1}[Fg^2 + 2\alpha^2 g^2] = g^2 N^{-1}[F + 2\alpha^2]$ for non-ε variables
$[ii] = N^{-1}[Fg^2 + 2\gamma^2 g^2 + 2\varepsilon^2 g^2] = g^2 N^{-1}[F + 2\gamma^2 + 2\varepsilon^2]$ for the ε variable,

Since $2\alpha^2 = 2\gamma^2 + 2\varepsilon^2$, the value of the scaling factor g gives $[ii] = 1$ for each variable.

The values of the non-zero fourth order moments for this design are:

$[iijj] = N^{-1} Fg^4$ for all variables;

and

$[iiii] = N^{-1}[Fg^4 + 2\alpha^4 g^4]$ for non-ε variables;
$[iiii] = N^{-1}[Fg^4 + 2\gamma^4 g^4 + 2\varepsilon^4 g^4]$ for ε variable;

The condition for satisfying the orthogonality requirement $[iijj] = 1$ is thus identical to that for the standard CCD

$$FN = (F + 2\alpha^2)^2$$

For rotatability, however, two conditions are obtained.

$$3F = F + 2\alpha^4 \quad \rightarrow \quad F = \alpha^4 \tag{7B.3}$$

$$3F = F + 2\gamma^4 + 2\varepsilon^4 \quad \rightarrow \quad F = \gamma^4 + \varepsilon^4 \tag{7B.4}$$

These cannot be simultaneously fulfilled unless $\varepsilon = 0$ and $\gamma = \alpha$ (since otherwise $\gamma^4 + \varepsilon^4 \neq \alpha^4$).

The amount δ by which the right-hand side of (7B.4) differs from that of (7B.3) is

$$\delta = \alpha^4 - (\gamma^4 + \varepsilon^4)$$

and since $\gamma^2 = \alpha^2 - \varepsilon^2$,

$$\delta = \alpha^4 - (\alpha^2 - \varepsilon^2)^2 - \varepsilon^4$$

$$\delta = \alpha^4 - (\alpha^4 - 2\alpha^2\varepsilon^2 + \varepsilon^4) - \varepsilon^4$$

$$\delta = 2\alpha^2\varepsilon^2 - 2\varepsilon^4$$

$$\delta = 2\varepsilon^2\left[\alpha^2 - \varepsilon^2\right]$$

As an example of the small magnitude of this difference, the following table shows the values of δ and fractional difference in the right-hand side of the equation for the seven variable case, with F = 64 The values of the parameters used are $\alpha = 2.8284$ (standard CCD rotatability criterion) and $\alpha = 1.0$, $\varepsilon = 0.05$ and $\varepsilon = 0.10$.

α	ε	δ	α^4	δ / α^4 %
2 8284	0 05	0 0400	64	0.062
1 0000	0 05	0.0050	1	0 499
2 8284	0 10	0.1598	64	0 250
1 0000	0.10	0 0198	1	1 980

Table 7B.2 Values of δ for a CCD with one ε-pair and equal scaling

The amount by which the rotatability condition for the ε variable deviates from that for the standard CCD is thus extremely small for the seven variable example. This amount will of course increase with decreasing α, and increasing ε, although it is independent of the value of F. Even for the case of n = 3, $F = 2^3 = 8$, $\alpha = 1\ 0$ and $\varepsilon = 0.1$, the value of δ / α^4 is still 1.980%.

It may thus be concluded that in general the inclusion of an ε-pair with equally scaled variables leads to insignificant changes in the orthogonality and rotatability criteria of the standard CCD. Since the results of Table 7B 2 have shown that the difference between the two axial parameters α and γ is also negligible, it is reasonable to employ the same value of $\alpha = \gamma$ as would be used in a standard CCD. This alternative specification of the CCD with single ε pair and equal scaling is thus identical, in practice, to the original description of Section 7.3, in which different scaling factors were applied to the variable values, according to whether they appeared at non-zero levels within the ε-star portion.

Appendix 7C

Lack of fit data for CCD with ε-star portion

ε	Maximum lack of fit — E_R Design	Factorial	Higher	Average lack of fit — E_R Design	Factorial	Higher
standard	1 2483	2.7467	6.2519	0 15039	0.93156	2.7334
0.0	1 3658	2.7506	6.3547	0.14449	0 93130	2.7289
0 025	1.3764	2.7513	6.3638	0.14817	0 93130	2.7284
0 05	1.3410	2.7510	6 3323	0.15481	0 93139	2.7296
0 075	1 3492	2.7523	6 3390	0.15628	0.93135	2.7291
0.1	1 3489	2.7539	6 3379	0 16292	0.93138	2.7287

Table 7C.1 Variation of lack of fit with ε-star parameter for augmented CCD with 92 tests and 71 coefficients (n=7, n_0=0)

ε	Maximum lack of fit — E_R Design	Factorial	Higher	Average lack of fit — E_R Design	Factorial	Higher
standard	4 6395	3.1124	6.4331	0.77241	0.77967	2.6160
0.0	4 6404	3.1124	6.4521	0 66624	0.77978	2.6153
0 025	4 6411	3.1123	6.4612	0 67626	0 77979	2 6149
0 05	4 6396	3.1123	6 4298	0 68395	0 77954	2 6158
0.075	4 6396	3.1126	6 4365	0 68906	0 77952	2 6153
0.1	4 6393	3.1127	6 4354	0 69638	0.77938	2 6150

Table 7C.2 Variation of lack of fit with ε-star parameter for augmented CCD with 92 tests and 17 coefficients (n=7, n_0=0)

ε	Maximum lack of fit — E_R Design	Factorial	Higher	Average lack of fit — E_R Design	Factorial	Higher
standard	4 0151	2.6508	6 2354	0.67564	0.85803	2 6048
0 0	4 0189	2.6546	6.3369	0.59544	0.85818	2 6001
0.025	4 0194	2.6551	6.3460	0.59910	0 85818	2 5996
0 05	4.0183	2 6542	6.3148	0 60568	0 85814	2.6008
0 075	4.0182	2 6544	6 3214	0 60711	0 85818	2.6003
0.1	4 0179	2 6548	6 3203	0 61369	0 85821	2.6000

Table 7C.3 Variation of lack of fit with ε-star parameter for augmented CCD with 92 tests and 36 coefficients (n=7, n_0=0)

ε	Maximum lack of fit — E_R Design	Factorial	Higher	Average lack of fit — E_R Design	Factorial	Higher
standard	1 2483	2.7467	6 2519	0.15039	0 93156	2 7334
0.0	1.3698	2.7507	6 3582	0.14346	0 93129	2.7287
0 025	1 3800	2 7515	6 3670	0.14713	0 93129	2.7282
0 05	1.3460	2.7511	6 3367	0.15405	0 93138	2.7294
0 075	1.3539	2.7524	6.3431	0.15561	0.93134	2.7289
0.1	1 3536	2.7541	6.3420	0.16235	0 93136	2 7285

Table 7C.4 Variation of lack of fit with ε-star parameter for augmented CCD with 93 tests and 71 coefficients (n=7, n_0=1)

ε	Maximum lack of fit — E_R Design	Factorial	Higher	Average lack of fit — E_R Design	Factorial	Higher
standard	4 6395	3.1124	6 4331	0 77241	0 77967	2.6160
0 0	4 6404	3.1124	6 4528	0 65918	0.77979	2.6153
0 025	4 6411	3.1123	6 4615	0.66909	0.77979	2 6149
0 05	4 6397	3 1123	6 4313	0 67708	0 77955	2.6157
0 075	4 6396	3 1126	6 4377	0.68204	0 77953	2.6153
0 1	4 6393	3 1127	6 4367	0 68929	0 77939	2.6149

Table 7C.5 Variation of lack of fit with ε-star parameter for augmented CCD with 93 tests and 17 coefficients (n=7, n_0=1)

ε	Maximum lack of fit — E_R Design	Factorial	Higher	Average lack of fit — E_R Design	Factorial	Higher
standard	4 0151	2.6508	6 2354	0.67564	0 85803	2 6048
0 0	4 0190	2 6547	6 3404	0 58955	0 85819	2.6000
0 025	4 0195	2 6553	6 3491	0 59321	0 85819	2 5995
0 05	4 0185	2 6543	6.3192	0 60006	0 85814	2 6006
0.075	4 0183	2.6545	6.3255	0 60159	0.85819	2 6001
0.1	4 0180	2.6549	6.3244	0 60827	0.85822	2.5998

Table 7C.6 Variation of lack of fit with ε-star parameter for augmented CCD with 93 tests and 36 coefficients (n=7, n_0=1)

8. Economic second-order designs

The results of numerical tests using both the linear + interactions model of Chapter 4 and the 71 term quadratic model in Chapter 6 have shown that, in the present engine noise investigation, the effect of interactions of order greater than two is extremely small, but that pure quadratic terms contribute significantly to the overall shape of the response surface. Additionally, the results of Section 6.5.1 have shown that the distribution of test points in the Central Composite Design results in a much less precise estimate of the pure quadratic effects than of the linear and interaction components. This reveals a mismatch between the model which is to be fitted and the tests which are performed in order to estimate this model.

In order to achieve a better match between design and model, an experimental design is sought which is specifically constructed to assess only those terms which have been shown to be of interest, but to estimate them with approximately equal precision and a minimal number of tests. These terms are of the following form.

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} X_i X_j + \sum_{i=1}^{n} \beta_{ii} X_i^2 \qquad (8.1)$$

Here the second and third terms on the right hand side of the equation represent the main effect and two-way interaction terms respectively, of the linear + interactions model, with the final term representing the pure quadratic contributions. Since the two-way interaction terms are simply mixed quadratic terms, (8.1) is in fact identical to the strict quadratic model of equation (5.1).

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n} \sum_{j=i}^{n} \beta_{ij} X_i X_j$$

In the final term of this equation the coincidence of i and j produces the pure quadratic terms, which can alternatively be described as self-interaction terms.

As discussed in Section 5.4, comparison of (8.1) with the complete quadratic plus interactions model of equation (5.12), which may be constructed from a full three level factorial experiment, shows that the successful use of the strict quadratic model to represent the behaviour of the system depends on the assumption that no significant interactions occur which involve more than two factors, and, further, that the only two factor interactions which do occur are the linear by linear ones, $X_i X_j$, with the cubic $X_i^2 X_j$, $X_i X_j^2$ and quartic $X_i^2 X_j^2$ terms also omitted.

8.1 A survey of available second-order designs

Due to the widespread use of quadratic models, a substantial body of theoretical, numerical and experimental work has been published on the subject of economic second-order designs. However, a comprehensive comparison of all available designs in the context of the computer simulation of engine noise radiation is not within the scope of the current work. Of more importance is the identification of a suitable existing experimental scheme. Since a general approach is required, it is important that a particular design be available for a wide range of the number of design variables, n. This section outlines the characteristics of the more commonly used designs.

8.1.1 Box-Behnken

The class of designs introduced by Box and Behnken (1960) are widely used in general experimental work. Their two main advantages are that they require just three levels of each of the design variables, and that many of the designs may also be blocked orthogonally. The first of these factors may be important if quantitative variables are used, or if, for some or all of the variables, only a limited number of levels may feasibly be tested. In contrast with the Box-Behnken designs, a general CCD, as described in Section 6.1, requires that each of the variables be tested at five levels ($-\alpha$, -1, 0, 1, α), although this is of course reduced to three levels if $\alpha = 1$. As discussed in Section 2.10, blocking is necessary if it is not possible to conduct all experimental tests under identical conditions. Orthogonal blocking ensures that estimation of design variable parameters is independent of any unwanted effect which may vary between blocks. An additional advantage of the Box-Behnken design is that, although not available for all n, Box and Behnken (1960) list designs for a number of different problem sizes (n = 3-7, 9-12, 16).

The disadvantages of the Box-Behnken designs are that, although highly orthogonal, with only the mean and pure quadratic terms correlated (as for the CCD), they are not always block orthogonal, nor are they always rotatable. Perhaps of more importance in the present application is the fact that the test requirement is still substantially greater than the number of parameters to be estimated. For a seven variable problem, for example, 36 parameters are to be estimated, whereas the Box-Behnken design contains 62 test points, giving a ratio of 58%. This falls slightly with problem size, so that for 10 variables this saturation ratio is 39%.

In the computer analysis of radiated engine noise, blocking of tests is not required, as discussed in Section 2.10, since all relevant conditions are exactly repeatable. Additionally, there is no limit on the number of different variable levels which can be tested, although there is, of course, a restriction on the magnitude of each level. The rather high number of tests is thus not compensated for, in the present noise analysis example, by other beneficial qualities.

8.1.2 Box-Draper saturated designs

A number of saturated designs, in which the number of parameters to be estimated is equal to the number of tests in the design, were introduced by Box and Draper (1971). Designs involving n = 2 and n = 3 variables were given which are optimal in terms of maximising the determinant $|\mathbf{X}'\mathbf{X}|$. The designs were generalised for $n \geq 4$ by Box and Draper (1974), although these are not necessarily optimal, judged on the above criterion.

8.1.3 Hoke

A class of economical designs for fitting the strict quadratic model was developed by Hoke (1974), generalising earlier designs by Rechtschaffner (1967). These designs are based on partially balanced irregular fractions of the 3^n factorial, and are valid for any number of variables $n \geq 3$. The Hoke designs fall into two categories; firstly, a class of designs in which the number of tests is equal to the number of parameters to be estimated, known as a 'saturated' design; and secondly, a class of 'minimally augmented' designs, in which extra tests are added to the saturated design in order to improve efficiency and to provide a number of degrees of freedom for estimation of errors.

The price to be paid for obtaining such a small experimental design, and thus high saturation ratio, is that the designs are neither orthogonal nor rotatable, although as discussed in Section 2.8, neither of these properties is absolutely indispensable. One important feature of the Hoke designs, however, is that they are invariant under permutation of the factors in the model, so that, as described in Section 2.9, the estimates of the parameters do not change if the order of the variables is changed. Thus all the parameters of the same form are measured with equal precision; e.g. the variance of parameter β_{ii} is independent of the value of i.

Despite the small number of tests used, Hoke showed that his designs compared favourably with both Box-Behnken designs and a particular class of CCD employed by Hartley (1959), which, additionally, are not permutation-invariant. A further investigation by Lucas (1976) compared the Hoke designs with the CCD and Box-Draper designs, and found that the Hoke designs performed better than the saturated designs of Box-Draper, and nearly as well as the CCD, although requiring far fewer tests.

8.1.4 Uniform shell designs

These were developed by Doehlert (1970) and Doehlert and Klee (1972), and are generated from the points of a regular simplex by calculating the difference between the corresponding variable values of pairs of points. The location of these points is such that they are uniformly spaced and lie on concentric spherical shells. The disadvantages of these designs are that a large number of variable levels are required, and, more importantly, that the number of points on a each shell is $n^2 + n$. Thus even a single shell plus centre point would contain $0.5n^2 + 0.5n$ more tests than there are parameters in the model. For a seven variable test, for example, the number of tests required is 57, giving a significance ratio of 63%. Additionally, Lucas (1976) found that the uniform precision design did not perform as well as either the Box-Behnken design or the CCD.

8.1.5 Hybrid designs

Introduced by Roquemore (1976) these designs are constructed from a CCD of dimension n–1, augmented with an extra row. The specification of this extra test is determined in such a way as to achieve a similar degree of orthogonality as the CCD, whilst also being nearly rotatable. The principal disadvantages of the hybrid design are that it is only available for a fairly limited range of problem sizes (n = 3, 4, 6 or 7), and that many of its points lie well beyond the bounds of the unit hypercube.

8.1.6 Other second-order designs

Of the many other schemes which have been put forward for the selection of test points in a second-order design, the following are worthy of mention. An early example of a family of saturated designs is given in Koshal (1933), from which it is possible to construct specific designs of any order d in n variables. Designs which are based on irregular fractions of factorials, and are very nearly saturated, were introduced by Westlake (1965) for n = 5, 7 and 9 variables. Much simpler designs which achieved a similar, or in some cases improved, level of saturation were suggested by Draper (1985). A set of saturated designs which are constructed using tests which form part of the three level factorial design were developed by Notz (1982).

8.2 Selection of an appropriate design

In selecting an economic second-order design to use in the investigation of engine noise, the two principal requirements are that a better estimate be obtained for the pure quadratic terms of the model, and that the number of tests be kept as low as possible without compromising the precision with which each of the parameters is estimated. It is not clear, however, to what extent the number of test points in the design may be reduced before substantial errors occur in the construction of the predictive model. Box and Draper (1987, p.520) suggest that designs which are either saturated or near to saturation are likely to be of interest if one of the following conditions apply.

1. Runs might be extremely expensive.
2. The checking of assumptions, the need for an internal error estimate, the need to check fit might not be regarded as important in a particular application.
3. The objective might be to approximate a function that can be computed *exactly* at any given combination of the input variables; that is, there is no experimental error.

Of these, the first and third are certainly true for the process under current investigation, with the importance of the second criterion being substantially reduced by the lack of experimental error. It is thus likely that, for the noise analysis problem, a design which is close to saturation will provide sufficient information to allow the necessary parameters to be estimated with the required precision. Of those designs described in Section 8.1 which are of a saturated or near-saturated nature, the most appropriate in the present circumstances would appear to be the class of designs due to Hoke. These possess the advantage that both saturated and near-saturated designs are available within the same scheme, and that these are valid for all $n \geq 3$. A further advantage is that, in addition to the tests conducted by Hoke, independent investigators have found that these designs perform well in comparison with alternative designs having a similar, or greater, number of test points (Lucas, 1976, Khuri and Cornell, 1987). For these reasons, the Hoke designs have been chosen for use in the further investigation of the engine noise problem, and the remaining discussions of the present chapter will be restricted to designs of this class.

8.3 Description of the Hoke designs

The description of the Hoke designs presented in this section is based on that given by Hoke (1974). Before demonstrating the way in which the Hoke designs are generated, it is useful to adopt the following conventions. As with the designs of previous chapters, the scaling convention of Section 2.12 is adopted, which, for tests drawn from the three level factorial design, results in normalised variable levels of -1, 0 and +1. Additionally, each of the test points of the design is notated as $(x_1\ x_2\ x_3 \ldots\ x_n)$, at which each variable i (for i=1, ..., n) is set to the (normalised) level x_i, which may take the value -1, 0 or 1. Further, let π (...) represent the family of test points obtained by permuting the symbols enclosed within the brackets. As an example, π (-1 0 1) gives the set of points {(-1 0 1), (-1 1 0), (0 -1 1), (0 1 -1), (1 -1 0), (1 0 -1)}.

In forming the Hoke designs, the complete set of 3^n factorial tests S_{FS} is first divided into subsets, such that each member of the r^{th} subset S_r lies on the hypersphere of radius $r^{1/2}$ about the centre point of the design (0 0 0 ... 0), for $0 \le r \le n$. The total set is thus partitioned as

$$S_{FS} = \sum_{r=0}^{n} \oplus S_r \tag{8.2}$$

where $\sum\oplus$ represents disjoint union

In order to lie on the hypersphere of radius $r^{1/2}$, each member of the set S_r must have exactly r variables which take a non-zero value, such that

$$S_r = \pi(\overbrace{\pm 1\ \pm 1 \ldots \pm 1}^{r}\ \overbrace{0\ \ 0\ \ \ldots\ \ 0}^{n-r}) \tag{8.3}$$

Each S_r is then divided into further subsets $S_r(j)$, in which exactly j of the non-zero terms take the value +1, and the remaining r–j take the value -1. Each subset is then of the form

$$S_r(j) = \pi(\overbrace{+1\ +1 \ldots +1}^{j}\ \overbrace{0\ \ 0\ \ \ldots\ \ 0}^{n-r}\ \overbrace{-1\ -1 \ldots -1}^{r-j}) \tag{8.4}$$

Special notation is used for the following three test points, within each of which all variables are at the same level.

$S_n(0) = (-1\ -1\ \ldots\ -1) = \mathbf{-1}$
$S_0 = (0\ 0\ \ldots\ 0) = \mathbf{0}$
$S_n(n) = (+1\ +1\ \ldots\ +1) = \mathbf{1}$

As an example, the four variable subset $S_2(1) = \pi\ (1\ 0\ 0\ -1)$ consists of the following twelve test points

x_1	x_2	x_3	x_4
1	0	0	-1
1	0	-1	0
1	-1	0	0
0	1	0	-1
0	1	-1	0
0	0	1	-1
0	0	-1	1
0	-1	1	0
0	-1	0	1
-1	1	0	0
-1	0	1	0
-1	0	0	1

and the complete three level factorial is composed of the subsets

set	r	j	No.(1) = j	No.(0) = n-r	No.(-1) = r-j	
S_0 :	0	0	0	4	0	= (0 0 0 0)
$S_1(0)$:	1	0	0	3	1	= π(0 0 0 -1)
$S_1(1)$:	1	1	1	3	0	= π(1 0 0 0)
$S_2(0)$:	2	0	0	2	2	= π(0 0 -1 -1)
$S_2(1)$:	2	1	1	2	1	= π(1 0 0 -1)
$S_2(2)$:	2	2	2	2	0	= π(1 1 0 0)
$S_3(0)$:	3	0	0	1	3	= π(0 -1 -1 -1)
$S_3(1)$:	3	1	1	1	2	= π(1 0 -1 -1)
$S_3(2)$:	3	2	2	1	1	= π(1 1 0 -1)
$S_3(3)$:	3	3	3	1	0	= π(1 1 1 0)
$S_4(0)$:	4	0	0	0	4	= (-1 -1 -1 -1)
$S_4(1)$:	4	1	1	0	3	= π(1 -1 -1 -1)
$S_4(2)$:	4	2	2	0	2	= π(1 1 -1 -1)
$S_4(3)$:	4	3	3	0	1	= π(1 1 1 -1)
$S_4(4)$:	4	4	4	0	0	= (1 1 1 1)

As a comparison with previous designs, examination of the above table shows that the subset S_n comprises the vertices of the hypercube, which are also the members of the 2^n full factorial design introduced in Section 4.3. Similarly, the star points of the CCD form the set S_1 if a value of $\alpha = 1$ is used (see Sections 6.1 and 6.4). The set of points S_0 lying on a hypersphere of radius zero contains just the single point (0 0 ... 0), which is the centre point of the design space.

Working first with a seven variable example, and then generalising the results for $n \geq 3$, Hoke selected four classes of these subsets as follows:

C_1 : **-1, 0, 1**
C_2 : $S_1(0)$, $S_1(1)$, $S_{n-1}(0)$, $S_{n-1}(n-1)$, $S_n(1)$, $S_n(n-1)$
C_3 : $S_2(0)$, $S_1(2)$, $S_{n-2}(0)$, $S_{n-2}(n-2)$, $S_n(2)$, $S_n(n-2)$
C_4 : $S_3(0)$, $S_3(3)$, $S_{n-3}(0)$, $S_{n-3}(n-3)$, $S_n(3)$, $S_n(n-3)$

From these four classes, subsets were then combined so as to generate all of the possible designs having (k+1)(k+2)/2 tests, each of these saturated designs then being analysed to determine the most efficient in terms of minimizing the trace of the inverse information matrix, $tr[(\mathbf{X}'\mathbf{X})^{-1}]$. Note that, for a given design size, $\mathbf{X}'\mathbf{X}$ only differs by a constant from the moment matrix $N^{-1}\mathbf{X}'\mathbf{X}$ used in previous chapters. The following three designs were found to yield the best performance.

D_1 : **-1**, $S_n(n-1)$, $S_1(1)$, $S_n(2)$
D_2 : **-1**, $S_n(n-1)$, $S_1(0)$, $S_n(2)$
D_3 : **0**, $S_n(n-1)$, $S_{n-1}(0)$, $S_n(2)$

These designs are valid for all $n > 3$. When $n = 3$ the subsets $S_n(n-1)$ and $S_n(2)$ are identically $S_3(2)$, and the subset $S_n(2)$ is replaced with $S_n(1)$ in each of the above designs.

A series of minimally augmented designs was then derived by adding to each of the above designs that subset of C_2 not already performed which gives the sharpest rise in precision of estimation of parameters. The four best combinations were found to be

$D_4 = D_1 \oplus S_1(0)$
$D_5 = D_1 \oplus S_{n-1}(0)$
$D_6 = D_2 \oplus S_{n-1}(n-1)$
$D_7 = D_3 \oplus S_1(1)$

Substitution of $S_n(1)$ for $S_n(2)$ is again made when $n = 3$.

Hoke derives expressions for the parameters, variance multipliers var(β)/σ^2 for each parameter, and values of tr[$(\mathbf{X'X})^{-1}$] and $|\mathbf{X'X}|$ as a function of n, and quotes the following performance data for each design in the case of n = 7.

	var (β_0)	var (β_i)	var (β_{ij})	var (β_{ii})	tr[$(\mathbf{X'X})^{-1}$]	$\lvert\mathbf{X'X}\rvert$
D_1	0.047	0.050	0.050	0.101	2.165	0.5273×10^{48}
D_2	0.055	0.050	0.050	0.101	2.172	0.5273×10^{48}
D_3	0.152	0.049	0.049	0.106	2.274	0.3847×10^{48}
D_4	0.039	0.046	0.050	0.048	1.748	0.1319×10^{51}
D_5	0.043	0.047	0.049	0.051	1.757	0.2777×10^{51}
D_6	0.055	0.043	0.049	0.051	1.735	0.3725×10^{51}
D_7	0.041	0.047	0.049	0.051	1.758	0.2611×10^{51}

Table 8.1 Comparison of Hoke designs for seven variables

The best of the saturated designs, judged on a minimum tr[$(\mathbf{X'X})^{-1}$] criterion is thus D_1, with the best minimally augmented design being D_6. It is interesting to note that designs D_1, D_2 and D_4 contain only subsets of S_n and S_1, whose members are all contained in the 2^n factorial and CCD star portion respectively. In contrast with the CCD, however, the total number of points in each of the Hoke designs is substantially lower, so that for a seven variable example 36 tests are required for a saturated design and 43 for a minimally augmented design, compared with 79 tests for a CCD with half factorial (resolution VII) linear portion. A further contrast with the CCD is that the values of the variance parameters for each class of variable are of roughly similar magnitude. For the 36 term model derived from the 79 test CCD, Table 6.11 shows that the ratio of highest to lowest variance values was 12.6 : 1, whilst for the Hoke D_1 design, above, it is just 2.1 : 1, and for the minimally augmented design D_6 falls to less than 1.3 : 1. Taking the D_6 design as an example, this equalisation has been partly achieved at the expense of an approximately three-fold increase in the variances of the mean, main effect and interaction terms. More importantly, however, the variance of the pure quadratic parameters has decreased by a factor of over 3.75.

8.4 Application of the Hoke designs

Detailed numerical tests have been carried out using two of the Hoke designs described in the previous section. To enable comparison with previous results, the seven variable noise analysis problem introduced in Appendix 1C is again used as an example. The two Hoke designs selected are the best performing saturated design, D_1, containing 36 tests, and the best of the minimally augmented designs, D_6, which contains 43 tests. In each case, the model to be constructed using these designs is the strict quadratic model of equation (8.1), which in seven dimensions contains 36 parameters. The results of these tests are compared with those obtained in Section 6.5.3, also generated using the 36 parameter strict quadratic model, derived from the 79 test CCD.

The exact specification of the full set of test points used in each of the Hoke designs is given in Appendix 8A. It is interesting to note the following features of each of these designs, which may also be inferred directly from the description of Section 8.3.

1. Both designs include the subsets **-1**, $S_7(6)$ and $S_7(2)$. These are the points

   (-1 -1 -1 -1 -1 -1 -1) : 1 point
   π (1 1 1 1 1 1 -1) : 7 points
   π (1 1 -1 -1 -1 -1 -1) : 21 points

   which all lie at the vertices of the two-level hypercube, and are thus members of the 2^n factorial array. Each design thus includes 29 of the two-level factorial points.

2. Design D_1 contains the subset $S_1(1)$, whose members are π (1 0 0 0 0 0 0), which, for a CCD with $\alpha = 1$, are the 'upper' star points having parameter $+\alpha$. Similarly, design D_6 contains the subset $S_1(0)$, whose members are π (0 0 0 0 0 0 -1), which are the 'lower' star points of the CCD, having parameter $-\alpha$.

3. Design D_6 contains the additional seven points $S_6(6)$, or π (1 1 1 1 1 1 0), which lie at the centre of those edges of the hypercube which meet at (1 1 1 1 1 1 1). These points can be expected to yield a particular advantage, since each will estimate the effect of just one of the pure quadratic terms of the model, and will thus aid in distinguishing between these parameters; a particular problem of the CCD.

4. In summary, the difference between the two Hoke designs is that D_1 uses the upper star points, whilst D_6 uses the lower star points, and also contains the seven additional 'edge' points. Comparison with the 79 test CCD shows that each of the two Hoke designs contains just under half the number of two level factorial tests, half the number of star points, and no centre point test. D_6 also contains seven additional tests not featured in the CCD.

8.4.1 Comparison of model coefficients

The parameter estimates which are obtained using the CCD and each of the Hoke designs are shown in Table 8.2.

	Parameter	CCD	Hoke D_1	Hoke D_6
1.	MEAN	86.711	86.691	86.713
2.	A	-2.2352	-2 2425	-2.2389
3.	A^2	9 8271x10^{-1}	9.7734x10^{-1}	1 0363
4.	C	-3.8703x10^{-1}	-3.7112x10^{-1}	-3.7145x10^{-1}
5	G	-3.3550x10^{-1}	-3 2932x10^{-1}	-3.3063x10^{-1}
6.	F	-2.3327x10^{-1}	-2.3200x10^{-1}	-2.3522x10^{-1}
7.	D	-2.2783x10^{-1}	-2 2216x10^{-1}	-2 2337x10^{-1}
8.	B	-2 0402x10^{-1}	-1.9640x10^{-1}	-2 0124x10^{-1}
9.	D^2	1.9729x10^{-1}	2.6731x10^{-1}	1.3820x10^{-1}
10	E	-1.5832x10^{-1}	-1.5679x10^{-1}	-1.5687x10^{-1}
11.	G^2	1.1884x10^{-1}	1.1132x10^{-1}	1.3657x10^{-1}
12.	F^2	-7 8670x10^{-2}	-1.5393x10^{-1}	-8 0867x10^{-3}
13	AG	-7 7552x10^{-2}	-9 4246x10^{-2}	-9 4643x10^{-2}
14.	C^2	6.5926x10^{-2}	1.2771x10^{-1}	2.1857x10^{-2}
15.	E^2	5 8004x10^{-2}	9 8444x10^{-2}	3.7231x10^{-2}
16	AD	5 1217x10^{-2}	5 3462x10^{-2}	5 3090x10^{-2}
17.	FG	-4 6484x10^{-2}	-3 8167x10^{-2}	-4.0363x10^{-2}
18.	CG	-4 6205x10^{-2}	-5 6190x10^{-2}	-5 7624x10^{-2}
19.	AF	-4 4455x10^{-2}	-5.0221x10^{-2}	-5.1124x10^{-2}
20.	AE	-4 2211x10^{-2}	-4 9408x10^{-2}	-4 9483x10^{-2}
21	EF	-2.3002x10^{-2}	-4.3094x10^{-2}	-4 4967x10^{-2}
22.	B^2	2 2509x10^{-2}	-8 8783x10^{-3}	3 6786x10^{-2}
23.	AC	2.0781x10^{-2}	8.4686x10^{-3}	8 3275x10^{-3}
24.	DG	-1.5427x10^{-2}	-1.7113x10^{-2}	-1.8777x10^{-2}
25.	BF	-1.3916x10^{-2}	-5 6973x10^{-3}	-8 8220x10^{-3}
26.	BE	1.1641x10^{-2}	-1 6102x10^{-2}	-1 8399x10^{-2}
27.	CD	1.0701x10^{-2}	2.1987x10^{-2}	2.0578x10^{-2}
28.	EG	-9.9781x10^{-3}	-1 8383x10^{-2}	-1.9751x10^{-2}
29.	AB	6 5416x10^{-3}	-1 8266x10^{-3}	-3 1523x10^{-3}
30.	BD	6.4857x10^{-3}	1.5883x10^{-2}	1.3290x10^{-2}
31.	BC	-5 1977x10^{-3}	-4.4122x10^{-2}	-4 6485x10^{-2}
32	DF	-3 9314x10^{-3}	-3 0036x10^{-2}	-3.2207x10^{-2}
33.	CF	3.3634x10^{-3}	1.2399x10^{-2}	1.0459x10^{-2}
34.	DE	-2 0508x10^{-3}	1.9413x10^{-2}	1.8071x10^{-2}
35	BG	1.4971x10^{-3}	2 3847x10^{-2}	2.1228x10^{-2}
36.	CE	2 8026x10^{-4}	5.1286x10^{-2}	5 0174x10^{-2}

Table 8.2 Variation of parameter values for CCD and Hoke designs

It can be seen from this table that the use of either of the Hoke designs results in little modification to the parameter estimates for the largest eight terms, compared with the estimates obtained using the CCD. Many of the remaining estimates, however, differ substantially between the three designs, with the pure quadratic terms being especially sensitive to the

particular distribution of test points. As an example, the highest estimate for the D^2 term is over 35% greater than the lowest value obtained, with the ratio of highest to lowest estimates for C^2 being 5.8, and for F^2 over 19. There is also a large degree of fluctuation in the estimation of smaller interaction terms, with the parameter CE differing between designs by a factor of nearly 180.

The effect of different designs on the value of the estimated parameters can be seen more clearly in Figure 8.1, in which terms 6 to 36 of Table 8.2 are plotted against their rank order (using the CCD magnitudes as reference).

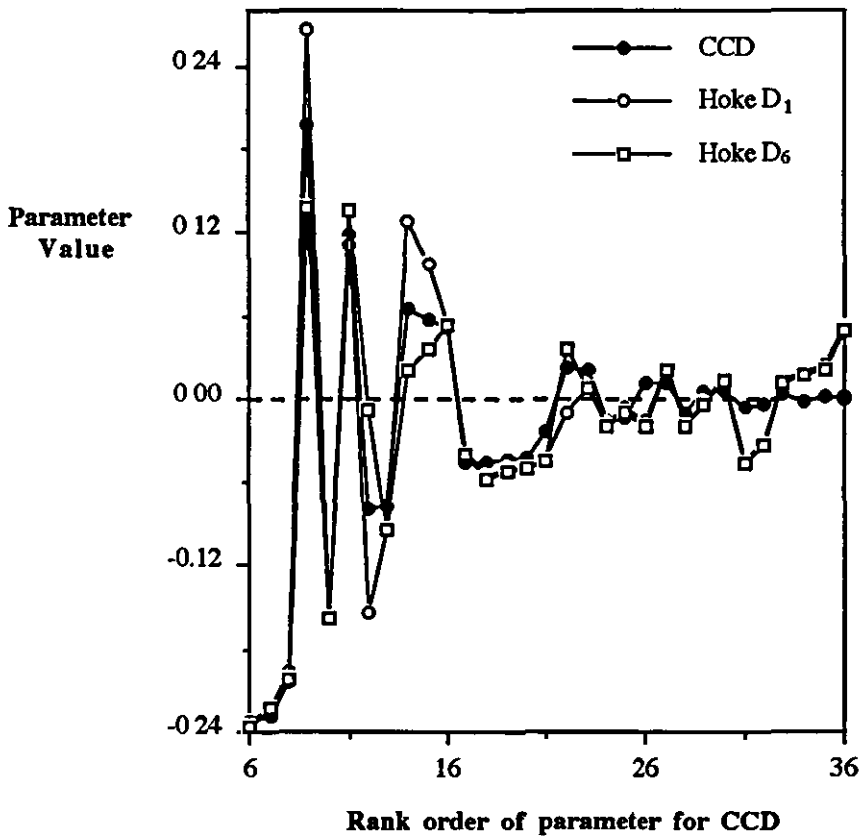

Figure 8.1 Variation of parameter values for CCD and Hoke designs

This graph shows that the general trend of values for each parameter is broadly similar for each of the designs, with the major differences occurring, with the exception of a number of the pure quadratic terms, in parameters of lower value, lying towards the right hand side of the graph. In interpreting these variations one should bear in mind the results of Sections 4.6.3 and 6.5.1. In the first of these it was shown that virtually all of the noise variation at two-level factorial points can be accounted for using just the main effect terms of the linear model, whilst Section 6.5.1 showed that, using the CCD, the only quadratic parameters which are found to be

statistically significant are the terms A^2 and D^2. With the sole exception of the D^2 effect, the estimate of each of these terms is largely unaffected by the use of either of the Hoke designs shown in Table 8.2, and it is thus likely, despite the discrepancy in the estimation of smaller parameters, that each of these models will yield similar predictions throughout the whole design variable space.

8.4.2 Variance values

The variance values which are obtained from the three designs for each class of term are shown in Table 8.3.

Class of Term	CCD	Hoke D_1	Hoke D_6
Mean	0.0158	0.0465	0.0547
Main effects	0.0152	0.0503	0.0425
Linear interactions	0.0156	0.0503	0.0489
Pure quadratic	0.1915	0.4048	0.2027

Table 8.3 Variance values for CCD and Hoke designs (n=7)

The variance values of the models derived from each of the Hoke designs which give most cause for concern are thus those of the pure quadratic terms, as was the case using the CCD. The halving of the number of star points in D_1, compared with the CCD, has resulted in an approximate doubling of the variance values of the quadratic terms. A considerable increase has also occurred in the variances of the mean, main effect and interaction terms, but these are still of relatively small magnitude. As expected, the inclusion of the seven additional 'edge' points in design D_6 has led to a large reduction in the variances of the quadratic terms, so that the performance of the Hoke D_6 design is roughly comparable with that of the CCD, despite containing little more than half the number of test points.

Comparison of these variance values with those derived by Hoke, given in Table 8.1, shows that, although the values for the mean, main effect and interaction terms are identical, those for the quadratic terms are substantially different. The reason for this is a difference in the scaling of the pure quadratic columns of the regressor matrix, a description of which is given in Appendix 8B. The result of this difference in the specification of the fitted model is that, whilst the coefficient values of the predictive model of equation (8.1) remain unchanged, as do predictions calculated using these coefficients, different values are obtained for the elements of the covariance matrix for the fitted model. An important effect of this difference in scaling is that

the pure quadratic variances shown in Table 8.3 are exactly four times those obtained by Hoke. This difference is of little importance in comparing the ability of the three designs of Table 8.3 to estimate accurately the pure quadratic parameters, since the adoption of an alternative scaling method would result in the pure quadratic variances for each design being modified, equally, by a factor of four.

What is of more importance, however, is that the variances obtained using Hoke's scaling suggest that, for D_6 , an error in estimation of one of the pure quadratic terms will lead to an approximately equal prediction error as would an error in the estimation of any of the mean, main effect or interaction terms (see Table 8.1). This is not the case, however, since, as discussed in Appendix 8B, the peak value of each of the (scaled) quadratic terms of Hoke's fitted model is twice that of the other terms in the model, leading to a prediction error which is proportional to four times the variance value, whilst for the other terms the prediction error is in direct proportion. In contrast, if the scaling used to obtain the results of Table 8.3 is adopted, the peak magnitude of each of the terms is equally 1.0, with the result that the error in prediction is directly proportional to each variance term, leading to a more informative comparison of the variance values for each class of term.

8.4.3 Effect of scaling on the selection of optimum designs

A further important consideration, which is related to the choice of scaling for the quadratic terms of the **X** matrix, is that Hoke's criterion for selecting the most efficient designs, within the general class outlined in Section 8.3, is that the trace of the inverse information matrix, $tr[(\mathbf{X}'\mathbf{X})^{-1}]$, be minimized. The value of the trace is numerically equal to the sum of the variances for each parameter in the fitted model, and will clearly change if the pure quadratic variances are modified by a change in scaling. The trace may be obtained, for a given design, by multiplying each of the appropriate variance values of Table 8.1 or 8.3 by the number of terms to which it applies, and summing the resulting four values. Since the change in scaling described above will only affect the magnitude of the pure quadratic variances, the relative contribution to the trace made by these values will vary by a factor of four. Since, additionally, the proportion of the trace value which is due to the quadratic variances differs between designs, it follows that the relative size of $tr[(\mathbf{X}'\mathbf{X})^{-1}]$ between different designs will also be modified if an alternative quadratic scaling is adopted. This is likely to lead to a different conclusion as to which of the full set of possible designs is the most efficient, in terms of the chosen criterion.

As an example, consider the two best saturated designs of Table 8.1, D_1 and D_2, and the two best augmented designs, D_6 and D_4. Table 8.4 shows the variance values and the resulting trace using each of the scaling methods, in which the numeral '1' refers to the pure quadratic variance and trace obtained using Hoke's scaling, as in Table 8.1, whilst '2' refers to

the values obtained using the scaling of Table 8.3, in which the pure quadratic variances are four times as large. The variances of the mean, main effect and interaction terms apply to both methods.

	var (β_0)	var (β_i)	var (β_{ij})	var (β_{ii})1	var (β_{ii})2	trace 1	trace 2
D_1	0.047	0.050	0.050	0.101	0.405	2.165	4.290
D_2	0.055	0.050	0.050	0.101	0.404	2.172	4.295
D_4	0.039	0.046	0.050	0.048	0.192	1.748	2.757
D_6	0.055	0.043	0.049	0.051	0.203	1.735	2.799

Table 8.4 The effect of quadratic scaling on the trace of Hoke designs in seven variables

Comparison of the traces for the two saturated designs shows that the advantage of D_1 over D_2 is slightly smaller using the scaling of Table 8.3 than if Hoke's scaling is used. For the augmented designs, Table 8.4 shows that the choice of most efficient design will change if the scaling of Table 8.3 is adopted, with D_4 now having a smaller trace than the previously optimal D_6.

If this alternative method of scaling is used, then it is clear that, since the effect of quadratic variance on the value of the trace is now much greater, the characteristics of an optimal design, whether saturated or augmented, are likely to be substantially different to those required previously. Therefore, since each of the designs listed in Table 8.1 are optimal, or nearly so, when using the previous scaling method, they are unlikely to be optimal if the alternative approach is used. In order to identify an optimum design based on the new scaling it is thus necessary to consider all possible combinations of the subsets of the 3^n full factorial array, as described in Section 8.3. An investigation of this nature would be a major undertaking in itself, and goes beyond the scope of the present work.

8.4.4 Covariance

It was shown in Section 6.5.1 that the only non-zero off-diagonal elements of the CCD covariance matrix are those representing covariance effects between the mean and pure quadratic terms, and between pairs of pure quadratic terms. Each of the two Hoke designs, however, have a dense covariance matrix, with covariance effects occurring between all pairs of terms of the fitted model. The value of all non-zero elements of the covariance matrix for each of the three designs are given in Table 8.5.

Class of Term		CCD	Hoke D_1	Hoke D_6
MEAN	β_i		0.0040	0.0065
MEAN	β_{ii}	-0.0018	-0.0009	-0.0013
MEAN	β_{ij}		-0.0048	-0.0060
β_i	β_j		-0.0043	-0.0033
β_i	β_{ii}		-0.0299	-0.0071
β_i	β_{jj}		0.0066	0.0020
β_1	β_{ij}		0.0043	0.0023
β_1	β_{jk}		-0.0035	-0.0031
β_{ii}	β_{jj}	-0.0308	-0.0640	-0.0309
β_{ii}	β_{ij}		-0.0048	-0.0055
β_{11}	β_{jk}		0.0004	0.0013
β_{ij}	β_{1k}		-0.0043	-0.0051
β_{1j}	β_{kl}		0.0035	0.0033

Table 8.5 Covariance values for CCD and Hoke designs (n=7)

This table shows that, although the covariance matrix for each of the Hoke designs is full, many of the terms are of a very small magnitude. The two largest terms for D_1 are the interaction between a given main effect and the quadratic effect in the same variable, and between pairs of quadratic effects. For D_6 only the covariance between pairs of quadratic terms is significant, and the value of this effect is similar to that obtained for the CCD.

8.4.5 Predictive ability of Hoke designs

To assess the prediction accuracy of the designs derived from each of the Hoke models, the lack of fit was calculated at locations specified by the three categories of test points described in Section 6.5.1. When comparison of these results is made with the results obtained using the CCD it should be borne in mind that the number of points which lie within each category is not necessarily the same in each case. Additionally, since each of the two Hoke designs only use half of the CCD star portion and do not contain a centre point, these extra eight tests, which lie at important locations within the design space, are included as a separate category. Note that design D_1 uses the 'upper' star points and D_6 the 'lower', so that the identity of the seven star points used solely to calculate lack of fit is different in each case. The numbers of tests which fall in each category for the three designs are shown in Table 8.6.

Design	Number of points in each category				
	Design	Factorial	Star	Higher	Total
CCD	79	64		16	159
Hoke D_1	36	99	8	16	159
Hoke D_6	43	99	8	16	166

Table 8.6 Number of points used to calculate lack of fit data for CCD and Hoke designs (n=7)

Since each of the Hoke designs contain 29 of the 2^n factorial points, compared with 64 for the CCD, the additional 35 points are included in the 'Factorial' lack-of-fit category. The extra seven tests which are used for the D_6 design are the 'edge' points which constitute the augmentation to the saturated design, as described in Section 8.4. The 16 points which form the 'Higher' category are identical for each design, and are as specified in Section 6.5.1 and Appendix 6A.

The maximum and average lack of fit values which occur within each of the categories of test point are given in detail in Appendix 8C. A summary of the average lack of fit values, in terms of the percentage of the function range, is given in Table 8.7.

Design	Average lack of fit — E_R				
	Design	Factorial	Star	Higher	All tests
CCD	0.68	0.86		2.61	0.94
Hoke D_1	0.00	1.40	1.08	2.66	1.20
Hoke D_6	0.13	1.44	1.01	2.62	1.20

Table 8.7 Summary of lack of fit data for CCD and Hoke designs (n=7)

As discussed in previous chapters, the lack of fit which occurs at design points is largely determined by the degree of saturation of the experimental design, with the saturated D_1 design giving zero lack of fit at these points. The performance of the two Hoke designs at the factorial points is substantially worse than the CCD, reflecting the fact that over 50% more tests are included in this category for the Hoke designs. The highest average lack of fit for each design occurs at the 'Higher' test points, and at these locations the Hoke designs are performing almost identically to the CCD, with D_6 marginally better than D_1. The star points and centre point are important measures of the performance of the model, since they lie at the highest distance from any of the design points of the Hoke designs, and at these points the average lack of fit is lower than at either the 'Factorial' or 'Higher' points. Over the full set of tests, the average lack of fit for the two Hoke designs is virtually identical, with the performance of each of them only slightly inferior to that of the CCD. This is despite the fact that the number of tests

in design D_1 is only 46% of that used in the CCD, whilst the augmented design contains just 54% of this number. These results show that the Hoke designs meet the requirements of minimising the required number of tests, whilst not compromising the prediction accuracy of the resulting model.

8.5 Observations on the use of Hoke's economical second order designs

The main conclusions to be drawn from the numerical tests carried out in the previous section are as follows.

- The prediction accuracy of each of the two Hoke designs tested was found to be similar to that of the 79 test CCD, despite requiring far fewer tests. On the basis of prediction errors, no reason was found for choosing the minimally augmented design, D_6, rather than the saturated design D_1, which requires fewer tests.
- The parameter estimates for the main effects of the predictive model were found to be very similar whichever of the three designs (CCD, D_1, D_6) was used. The value of the A^2 parameter, which is the dominant quadratic effect, was also found to vary little between designs, although other quadratic coefficients, and a number of small magnitude interaction parameters, were found to exhibit a high degree of variability.
- The largest of the variance terms obtained using the Hoke designs are those for the pure quadratic terms. A disadvantage of the saturated design D_1 is that these variances are twice the size of the corresponding values for the CCD, whilst the minimally augmented design D_6 yields variances of approximately equal magnitude to the CCD.
- The way in which the pure quadratic columns of the regressor matrix are scaled has a critical influence on the selection of an optimum design. If an alternative scaling factor to that used by Hoke is adopted, in which the quadratic terms have a peak value of 1.0, then the design D_4 has a lower trace than the previously optimum design D_6. Other designs of the full set investigated by Hoke are likely to perform better under this new scaling method.
- Hoke's economical second order designs have been found to provide an extremely useful alternative to the CCD for the approximation of the engine noise response surface which is used in the current investigation, in that they yield approximately the same accuracy of prediction throughout the design variable space, whilst having a much reduced test requirement. No improvement over the CCD has been found, however, in terms of reducing the variance of the pure quadratic terms of the fitted model. Under the alternative scaling method, which results in a greater contribution to the matrix trace from the quadratic variances, it is likely that improved designs may be found which reduce both the overall trace and the quadratic variance. Further work is required in this area.

Appendix 8A

Specification of test points for Hoke designs in seven variables
(Section 8.4)

Hoke D_1

Variable						
A	B	C	D	E	F	G
0060	0200	0200	.0040	0100	.0200	.0060
0060	0320	.0320	.0140	0250	.0320	.0120
0120	0200	.0320	0140	0250	0320	.0120
0120	0320	.0200	0140	.0250	0320	.0120
0120	.0320	.0320	0040	.0250	.0320	.0120
.0120	.0320	0320	0140	0100	.0320	.0120
0120	.0320	.0320	0140	0250	.0200	.0120
0120	.0320	.0320	0140	0250	.0320	.0060
0120	.0260	.0260	.0090	.0175	0260	.0090
.0090	0320	.0260	.0090	.0175	.0260	0090
0090	.0260	.0320	0090	.0175	.0260	.0090
0090	.0260	0260	0140	.0175	.0260	0090
.0090	.0260	.0260	0090	0250	.0260	.0090
0090	0260	.0260	0090	.0175	.0320	0090
0090	0260	.0260	0090	.0175	0260	0120
0120	0320	0200	0040	.0100	.0200	0060
0120	.0200	0320	0040	.0100	.0200	.0060
.0120	.0200	0200	.0140	0100	.0200	.0060
.0120	.0200	.0200	.0040	0250	.0200	.0060
0120	0200	0200	0040	0100	0320	0060
.0120	0200	0200	.0040	.0100	0200	.0120
0060	0320	.0320	0040	.0100	0200	0060
0060	0320	.0200	0140	.0100	.0200	0060
0060	0320	0200	.0040	0250	.0200	.0060
.0060	.0320	0200	.0040	0100	.0320	.0060
.0060	.0320	0200	.0040	.0100	.0200	.0120
.0060	.0200	.0320	.0140	.0100	0200	.0060
0060	0200	.0320	0040	.0250	.0200	0060
0060	.0200	0320	0040	.0100	.0320	.0060
.0060	.0200	.0320	.0040	0100	.0200	.0120
.0060	.0200	0200	.0140	.0250	0200	.0060
0060	0200	0200	.0140	.0100	0320	.0060
0060	0200	.0200	0140	.0100	.0200	0120
0060	0200	0200	0040	0250	.0320	.0060
0060	.0200	0200	0040	0250	.0200	.0120
.0060	.0200	0200	.0040	.0100	0320	.0120

Hoke D_6

Variable						
A	B	C	D	E	F	G
0060	0200	0200	.0040	0100	.0200	0060
0060	.0320	0320	.0140	0250	.0320	0120
.0120	0200	0320	0140	.0250	0320	.0120
0120	0320	0200	.0140	0250	.0320	.0120
0120	0320	.0320	0040	0250	0320	.0120
0120	0320	0320	.0140	0100	0320	.0120
0120	0320	0320	.0140	0250	0200	0120
.0120	0320	0320	0140	.0250	0320	0060
.0060	.0260	0260	.0090	0175	.0260	0090
.0090	.0200	0260	.0090	0175	0260	0090
.0090	0260	0200	0090	0175	0260	0090
.0090	.0260	0260	.0040	0175	.0260	0090
0090	0260	0260	0090	0100	0260	0090
.0090	0260	.0260	0090	0175	0200	0090
.0090	.0260	.0260	0090	.0175	0260	.0060
.0120	.0320	.0200	0040	.0100	0200	.0060
.0120	.0200	.0320	0040	.0100	.0200	.0060
0120	.0200	0200	.0140	0100	0200	.0060
0120	.0200	0200	.0040	0250	0200	0060
.0120	0200	.0200	0040	0100	.0320	0060
0120	0200	0200	0040	0100	0200	0120
0060	0320	0320	.0040	0100	.0200	0060
0060	0320	0200	0140	0100	0200	0060
0060	0320	0200	0040	0250	0200	0060
.0060	0320	0200	0040	0100	0320	0060
.0060	0320	0200	0040	0100	0200	0120
0060	.0200	.0320	.0140	0100	.0200	.0060
.0060	.0200	.0320	.0040	0250	.0200	.0060
.0060	.0200	.0320	0040	.0100	0320	.0060
.0060	.0200	.0320	0040	0100	0200	.0120
.0060	.0200	.0200	.0140	0250	.0200	.0060
0060	.0200	0200	.0140	0100	0320	.0060
.0060	0200	0200	0140	0100	.0200	0120
0060	.0200	0200	.0040	0250	0320	0060
0060	0200	0200	0040	0250	.0200	0120
0060	0200	0200	0040	0100	0320	0120
0090	0320	.0320	0140	.0250	0320	.0120
.0120	.0260	.0320	.0140	.0250	.0320	.0120
.0120	.0320	.0260	.0140	.0250	.0320	.0120
.0120	.0320	.0320	.0090	0250	.0320	.0120
.0120	.0320	0320	.0140	0175	0320	.0120
0120	.0320	0320	.0140	0250	0260	0120
.0120	0320	0320	0140	0250	0320	0090

Appendix 8B

Scaling of orthogonal polynomials

It was shown in Section 5 2 that a diagonal moment matrix cannot be obtained when fitting a quadratic model, so that the design cannot meet the orthogonality criteria outlined in Section 2 12. An alternative approach may be taken, however, by rearranging the predictive model in terms of certain polynomial functions of the input variables (Box and Hunter, 1957). These polynomial functions may be chosen in such a way as to achieve mutual orthogonality. As an example, consider a simple one-dimensional design consisting of three tests at levels -1, 0 and +1. If the model to be estimated from these tests is of the form

$$Y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 \tag{8B.1}$$

then the X matrix will be as follows:

$$\mathbf{X} = \begin{array}{c} \begin{array}{ccc} 1 & x_1 & x_1^2 \end{array} \\ \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \end{array} \tag{8B.2}$$

It can be seen that, whilst the second column is orthogonal to each of the other two, the first and third columns are not orthogonal. Orthogonality may be achieved by rewriting the predictive model in terms of polynomial functions of the original variable, as follows:

$$Y = \alpha_0 + \alpha_1 x_1 + \alpha_{11}(3x_1^2 - 2) \tag{8B 3}$$

where the original coefficients of (8B.1) may be expressed as

$$\beta_0 = \alpha_0 - 2\alpha_{11}$$
$$\beta_1 = \alpha_1$$
$$\beta_{11} = 3\alpha_{11}$$

The design matrix is now

$$\mathbf{X} = \begin{array}{c} \begin{array}{ccc} 1 & x_1 & 3x_1^2\text{-}2 \end{array} \\ \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix} \end{array} \tag{8B.4}$$

with each of the columns orthogonal to the other two. An additional effect of rewriting the prediction equation as (8B.3), however, is that the peak magnitude of the third term is twice that of each of the other two. The result of this is that the maximum contribution to the standard prediction error which is made by an error in the estimation of the third term is proportional to twice the standard error of the coefficient in that term. In general, this does not cause a particular problem for the investigator, since the values of the coefficients of the original predictive equation (8B.1) remain unchanged, as does the prediction at any chosen point in the design variable space. A problem arises, however, if the standard errors of the coefficient estimates (or their variance values, which are the squares of the standard errors) are used independently as a measure of the performance of an experimental design.

In particular, one criterion which may be adopted for the assessment of designs is the minimization of the prediction variance. If the design is orthogonal, with equal peak values in each of the fitted terms, then the maximum variance in prediction which is contributed by each term is in direct proportion to the variance on the estimate of that term (see Appendix 2B). One possible measure of prediction variance, then, is the trace of the covariance matrix, which is equal to the sum of the individual coefficient variances. If, however, as in the example of equation (8B 3), the peak value of each of the terms in the predictive equation is not equal, then the

trace of the covariance matrix is no longer a measure of the sum of contributions to the prediction variance of the individual terms. In this case, a weighted sum of individual contributions might perhaps be used, with the weighting reflecting the relative peak value of each of the terms involved.

Sections 8.3 and 8.4 are an example of the use of the trace of the covariance matrix in assessing competing designs. In this case, however, the orthogonal polynomials are chosen as shown in (8B.3) and (8B.4), whilst the criterion for judging designs is simply the sum of the unweighted coefficient variances. The use of this criterion is thus inappropriate to the choice of orthogonal polynomials, and does not give an accurate indication of the maximum likely error in prediction.

A simple means of rectifying the disparity in peak variable values of the model of equation (8B.3) is to scale the third polynomial function such that its peak magnitude is equal to those of the other two. The modified predictive model would then be as shown in equation (8B.5).

$$Y = \psi_0 + \psi_1 x_1 + \psi_{11}\left[\frac{1}{2}(3x_1^2 - 2)\right] \qquad (8B.5)$$

The original coefficients of (8B.1) may then be expressed as

$$\beta_0 = \psi_0 - \psi_{11}$$
$$\beta_1 = \psi_1$$
$$\beta_{11} = 1.5\psi_{11}$$

The design matrix now becomes

$$X = \begin{bmatrix} 1 & x_1 & {}^1/_2(3x_1^2-2) \\ 1 & -1 & {}^1/_2 \\ 1 & 0 & -1 \\ 1 & 1 & {}^1/_2 \end{bmatrix} \qquad (8B.6)$$

with the columns mutually orthogonal, and each having a peak magnitude of 1.

Even in this case, however, there still remains the question of whether a minimal covariance matrix trace is a suitable measure of the performance of a design, if it is the prediction variance which the investigator seeks to reduce. It should be noted that the trace of the covariance matrix provides an indication of the maximum prediction variance which could occur if all of the terms of the fitted equation were simultaneously to take their peak values. If, in the simple one-dimensional example demonstrated here, the terms to be estimated were those of equation (8B.1), (i.e. the mean, the linear term x_1, and the quadratic term x_1^2), then a value of $x = 1$ would indeed result in each term simultaneously taking a value of 1. If the orthogonal polynomials of either (8B.3) or (8B.5) are used, however, then (8B.4) and (8B.6) show that the maximum value of the third term of the predictive equation occurs when $x = 0$, so that maximum prediction error cannot be simultaneously contributed by all three terms. A better measure of maximum prediction error might thus be one which takes into account the relative magnitudes of the terms of the predictive equation at different levels of the original input variables. Further work is required in this area.

Appendix 8C

Lack of fit data for CCD and Hoke designs

In each of the following tables, the CCD used comprises 79 tests (see Section 6 5.1), whilst the Hoke D_1 design contains 36 tests and the Hoke D_6 has 43 (see Section 8.3). In each case the model constructed using the three designs is the 36 term strict quadratic model introduced in Section 6 5.3.

Design	No of lof tests	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
CCD	79	0 306	4 02	0 052	0.68
Hoke D_1	36	0 000	0 00	0 000	0 00
Hoke D_6	43	0 046	0 62	0 010	0 13

Table 8C.1 Summary of lack of fit at design points for CCD and Hoke designs (n=7)

Design	No of lof tests	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
CCD	64	0 202	2 65	0.065	0.86
Hoke D_1	99	0.317	4.14	0 107	1.40
Hoke D_6	99	0.338	4 49	0 108	1.44

Table 8C.2 Summary of lack of fit at factorial points for CCD and Hoke designs (n=7)

Design	No. of lof tests	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
Hoke D_1	8	0.188	2.45	0.083	1.08
Hoke D_6	8	0.164	2 18	0 076	1.01

Table 8C.3 Summary of lack of fit at star points for CCD and Hoke designs (n=7)

Design	No. of lof tests	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
CCD	16	0 476	6.24	0 199	2.61
Hoke D_1	16	0.510	6 65	0 204	2 66
Hoke D_6	16	0.520	6.90	0 197	2 62

Table 8C.4 Summary of lack of fit at additional points for CCD and Hoke designs (n=7)

9. Optimization using response surfaces

The purpose of the present chapter is to demonstrate the way in which practical optimization studies may be carried out using the response surface methods discussed in previous chapters. A number of case studies are presented which demonstrate the flexibility of this method in investigating a large number of combinations of different objective and constraint functions. Each example uses the finite element model of the four cylinder in-line diesel engine introduced in Appendix 1C, and the design variables used are selected from the seven which are described in that section.

Five examples are presented, the aims of which are as follows.

1. Unconstrained noise optimization.
2. Noise optimization with a known mass constraint.
3. Mass optimization with a known noise constraint.
4. Sweep through the full range of possible mass constraints to give the noise-mass trade-off within the design variable space.
5. Optimization using discontinuous variables.

These numerical trials have all been carried out using the computer program *optrsm*, which is described in Appendix 1B. The 'computer experiments' which were performed in order to obtain the function values at the required test points were carried out using the analyser program described in Appendix 1A. In many of the examples the mathematical models of the response surfaces which are used have been introduced and discussed in earlier chapters. These are referenced where appropriate, and may be consulted for further information on, for example, coefficient values and quality of fit.

9.1 The optimization algorithm

The optimization routine used for all of the numerical examples presented in this chapter is taken from the NAG software library (NAG, 1983). It is designated E04VDF, and is designed to perform minimization of any arbitrary smooth objective function, subject to a number of constraints. These constraints may include simple bounds on the input variables and functional constraints which are either linear or smoothly non-linear. It should be noted that this routine is different from that used in the example of Section 1.6. The principal reason for selecting a different routine is that, in contrast with the direct iterative method, the use of a

simplified mathematical representation of each function allows analytic calculation of gradient information. The use of gradient information is extremely beneficial, in that it yields a substantial increase in the efficiency of the numerical optimization procedure, since it is both more accurate and less computationally expensive than the calculation of finite difference approximations to the required gradients.

The information which must be supplied to the optimization routine is as follows.

1. Number of design variables
2. Upper and lower bounds on each variable.
3. A subroutine which calculates the value and gradients of the objective function at a given point.
4. A subroutine which calculates the value and gradients of each of the constraint functions at a given point.
5. Upper and lower bounds on each of the constraint functions.
6. The location of a point in the design variable space at which to commence.

Of these six requirements, the first two will already have been decided prior to carrying out the computer experiments. Items 3 and 4 access the mathematical representations of each function which have been generated using the response surface methods developed in previous chapters. Since each response surface is a simple polynomial expression, a general algorithm for the calculation of first derivative information with respect to each of the design variables is readily formulated, and has been included in the program *optrsm*. The choice of appropriate constraint bounds is discussed in Section 9.6, below. For both the design variables and the constraint functions, an equality constraint can be specified by setting the upper and lower bounds to identical values. Alternatively, each variable or constraint may be left unbounded in either direction.

The final piece of information which must be supplied to the optimization routine is the specification of a point in the design space at which to commence optimization. The location of this point is only critical if the response surface has a number of local minima, since the selection of an alternative starting point may then lead to a different, and better, minimum. If the routine can be relied upon to identify a global optimum under all conditions, however, then the only benefit to be gained from a better selection of this point is a reduction in computational effort – a point chosen close to the location of the required optimum can be expected to result in a smaller number of function calls. When using low order polynomial response surface, the scope for such local minima is very limited, and it has been found, in practice, that the selection of an alternative starting point is needed only rarely.

A procedure which has been developed for the selection of a starting point is to choose the test point which has the 'best' objective function value, whilst still observing each of the constraint conditions. The choice of this point is automated within the *optrsm* program, and has

the advantage that it is more likely to lie close to the final optimum than any of the other experimental test points, and much more likely than any random selection within the design variable space. It should be noted that this method of providing a 'head start' for the optimization cannot be used when using a direct iterative approach to optimization, since an essential prerequisite is a substantial amount of widely distributed function information. If it is found that the starting point chosen results in the identification of a local minimum, then an alternative starting point can be selected, such as the point at which all variables are at their low (or high) bound. An example of such a case is given in Section 9.8.

9.2 Unconstrained noise optimization

As an introduction to the use of response surfaces for optimization, the first example involves the search for the lowest possible noise value within the design variable space, irrespective of the value of any other function. All seven of the design variables described in Appendix 1C are included in this study. The design region is investigated using a full factorial design, involving 128 tests, which is the same as that presented in Section 4.1. Table 9.1 shows the specification of the optimum design which is identified in this study, and it can be seen, by comparison with Table 1C.1, that this design lies at the high bound of each of the design variables.

Optimum value of noise function	=	84.1	dB(A)
Total mass of block at optimum	=	185.3	kg
Optimum value of variable A	=	12.0	mm
Optimum value of variable B	=	32.0	mm
Optimum value of variable C	=	32.0	mm
Optimum value of variable D	=	14.0	mm
Optimum value of variable E	=	25.0	mm
Optimum value of variable F	=	32.0	mm
Optimum value of variable G	=	12.0	mm

Table 9.1 Unconstrained noise optimum using full factorial design

Before considering the implications of the location of this optimum, some observations are required on the nature of the approximating response surface used, and on the effect of this choice on the optimization process.

In this introductory example the response surface derived from the tests only contains terms which are either main effects or linear interaction terms, so that identification of the minimum noise design is extremely easy to carry out. Indeed, it is possible to make an estimate of the location of the optimum design by simply considering the coefficient values of the mathematical model. It was shown in Sections 4.2 and 4.3 that the noise surface is dominated by the largest fifteen coefficients, which are listed in Table 4.4, repeated here as Table 9.2.

1.	MEAN	$8.8082\times10^{+1}$
2.	A	-2.2377
3.	C	-3.8596×10^{-1}
4.	G	-3.3608×10^{-1}
5.	D	-2.3021×10^{-1}
6.	F	-2.2872×10^{-1}
7.	B	-1.9920×10^{-1}
8.	E	-1.5888×10^{-1}
9.	AG	-7.4246×10^{-2}
10.	AD	5.2756×10^{-2}
11.	AF	-4.6907×10^{-2}
12.	CG	-4.4956×10^{-2}
13.	AE	-4.4724×10^{-2}
14.	FG	-4.4068×10^{-2}
15.	ACG	3.5056×10^{-2}

Table 9.2 Significant noise coefficients for full factorial test (n=7)

It can be seen that all of the main effect coefficients have a negative sign, as do all but one of the two-way interaction terms. This suggests that the lowest noise value will be found when all variables have their largest values. If this is the case, then the optimum will occur at one of the original test points, since all of the vertices of the hyperspace are tested in a full factorial design. This presents a problem when selecting a starting point for demonstration purposes, since, if the test point which has the best objective function value is chosen, as suggested above, then this will result in the optimum point being used as the starting point, and no iterations will be performed. In order to demonstrate the search for an optimum value, an alternative staring point has been chosen for the example presented here. This is the point at which each of the variables has its lowest value, which lies at the furthest possible distance from the expected optimum.

The search history of this optimization test is presented in Table 9A.1 of Appendix 9A, and shown graphically in Figure 9.1.

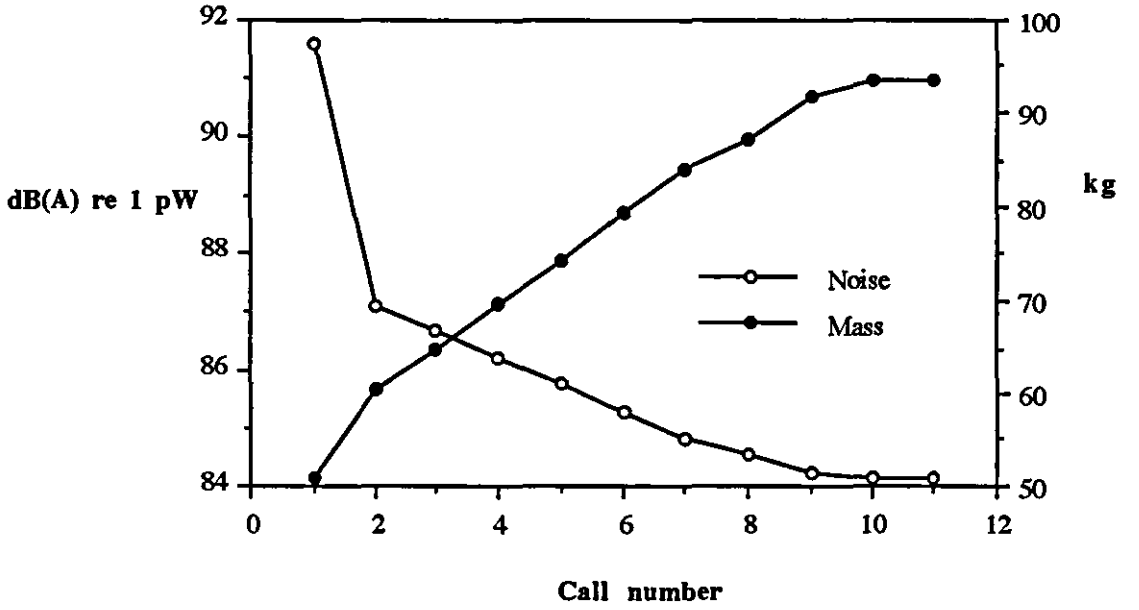

Figure 9.1 Search history - unconstrained optimization in seven variables using a full factorial design (128 tests)

The search history shows that the optimum has been identified using just 11 function calls, and that, as expected for a simple linear function, the noise decreases monotonically with increasing mass. The small number of iterations performed is in contrast with the example of direct iterative optimization shown in Section 1.6, which involves just three design variables, and yet requires 20 calls to locate the minimum, and a further 25 to confirm it.

Although the number of function calls required at the optimization stage is far fewer than the example of Section 1.6, it should be remembered that in order to construct the analytic response surface, 128 calls to the analyser program were performed. Additionally, the analysis of Section 4.1.2 showed that there was significant lack of fit at locations away from the vertices of the hypercube when using a linear model, so that there is the possibility that some inaccuracy in locating the optimum has occurred due to the presence of higher order components of the noise response. The analyses of Chapters 6 - 8 have shown that both of these issues can be addressed by using an economic second-order design, such as those of Hoke.

If the optimization is repeated using the minimally augmented Hoke design D_6 (see Section 8.3), then the optimum which is found is that shown in Table 9.3. Since the value of each of the quadratic terms is substantially lower than the main effect in the same variable, as shown by the parameter listing of Table 8.2, it is expected that the unconstrained optimum will still lie approximately at the upper bound of all variables. Table 9.3 shows that this is indeed the case, with only variable D moving away from its upper bound. The fact that variable D is the most likely to move away from its bound can be deduced from Table 8.2, which shows that

variable D has the highest ratio of quadratic coefficient magnitude to main effect magnitude. The search history is shown graphically in Figure 9.2.

Optimum value of noise function	=	84.0	dB(A)
Total mass of block at optimum	=	184.1	kg
Optimum value of variable A	=	12.0	mm
Optimum value of variable B	=	32.0	mm
Optimum value of variable C	=	32.0	mm
Optimum value of variable D	=	12.1	mm
Optimum value of variable E	=	25.0	mm
Optimum value of variable F	=	32.0	mm
Optimum value of variable G	=	12.0	mm

Table 9.3 Unconstrained noise optimum using Hoke design

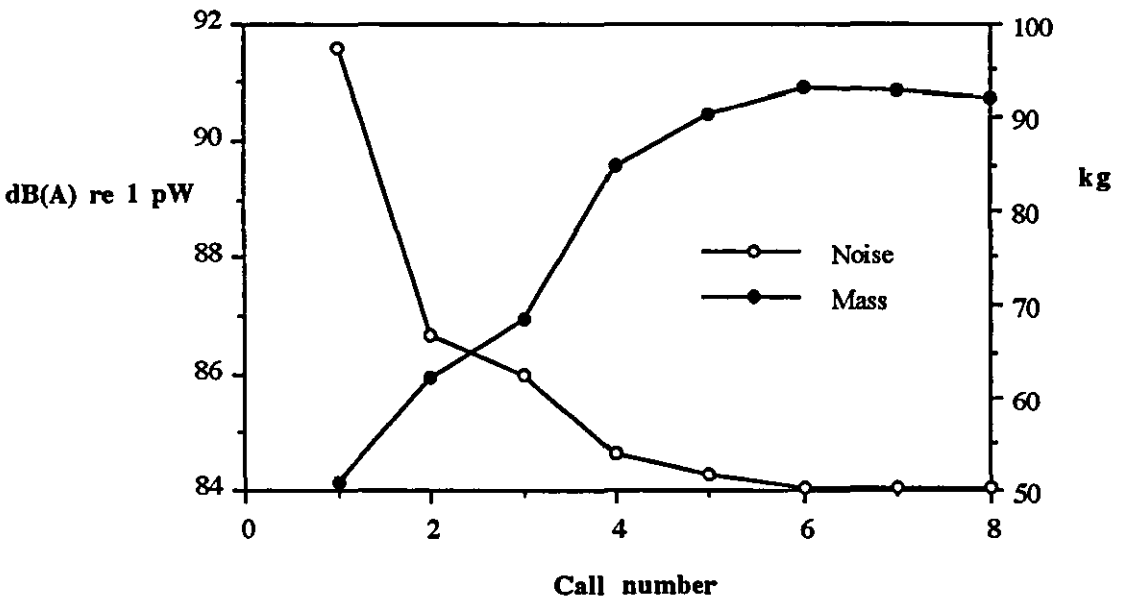

Figure 9.2 Search history - unconstrained optimization in seven variables using a minimally augmented Hoke design D_6 (43 tests)

Although a second-order model is now being used, this only contains 36 coefficients, rather than the 128 coefficients of the full factorial design, with the result that the solution converges after just 8 function calls. In addition, the number of calls to the analyser program required by the Hoke design is just 43 - two less than the 3 variable direct optimization problem of Section 1.6.

Returning to the location of the optimum design, this example suggests that, if no constraints are applied, the optimization routine will select a design which is as heavy as possible, so that the location of the optimum is determined solely by the specification of the upper bound of each variable. This agrees both with experimental findings and with vibration theory, since an increase in component dimension will result in a linear increase in mass, but a larger increase in stiffness, due to the squared increase in cross section. The natural frequencies of the structure will thus increase, so that fewer resonance conditions can occur in the frequency range of interest. Of more importance is that, at a given engine speed, each resonance will be excited by higher orders of engine excitation, which are in general of lower magnitude. These factors will together lead to a reduction in both vibration response and radiated noise.

In the present example, the original block design, whose specification is given in Appendix 1C, produced a noise level of 92.9 dB(A), with a mass of 139.2 kg. Although a predicted reduction in noise of nearly 9 dB(A) has been achieved, an increase in mass of nearly 45 kg has been incurred, which is nearly a one-third increase in the total block mass. A mass increase of this magnitude is unlikely to be acceptable to the engine designer, and it is for this reason that an unconstrained noise optimum is rarely of practical use. Only if the range of each variable is extremely limited, or the number of variables small, is the maximally heavy design likely to be an acceptable result. If this is the case, however, then there is little benefit in carrying out a numerical optimization study, and the upper bound of each variable can be selected without the need for detailed analysis.

9.3 Noise optimization with a known mass constraint

The second example expands on the previous study, in that a mass constraint will now be applied to the optimization problem. The aim is to identify which combination of the five variables included in the original design will have the lowest radiated noise value, whilst incurring no mass increase over the original engine block. A seven variable example with a zero-mass-change constraint is not possible in the present case, since, if all seven variables were included, the mass constraint would be violated even with each at its lower bound.

The response surface was investigated using a five-dimensional minimally augmented Hoke design D_4, which requires 26 tests. The parameter estimates which were obtained for the noise surface are given in Table 9A.2 of Appendix 9A. The lack of fit in the model was tested both at the design points and at a set of additional points, which are those tests of the appropriate CCD (with full factorial portion and $\alpha = 1$) not included in the Hoke design. A summary of the lack of fit calculations is given in Table 9.4, and suggests that the model constructed is providing a good representation of the original response.

Category	No. of tests	Maximum lack of fit dB(A)	Maximum lack of fit E_R	Average lack of fit dB(A)	Average lack of fit E_R
i) Design points	26	0.1584	2.6909	0.0488	0.8290
ii) Additional points	17	0.4647	7.8936	0.2119	3.5997
Average over all 43 tests				0.1133	1.9244

Table 9.4 Summary of lack of fit calculations for Hoke D_4 design with 26 tests and 21 coefficients (n=5) - Function : noise

The optimization trial was then carried out, with the mass constraint chosen to give a total block mass of 139.2 kg, which is the mass of the original design. A summary of the search history, starting from the test point which has the lowest noise, whilst obeying the mass constraint, is given in Table 9A.3 of Appendix 9A, and is shown in graphical form in Figure 9.3. This graph shows that the solution converges very rapidly. The reason for this can be seen from Table 9A.3, which shows that the search commences from a point at which three of the five variables are already at their optimum values.

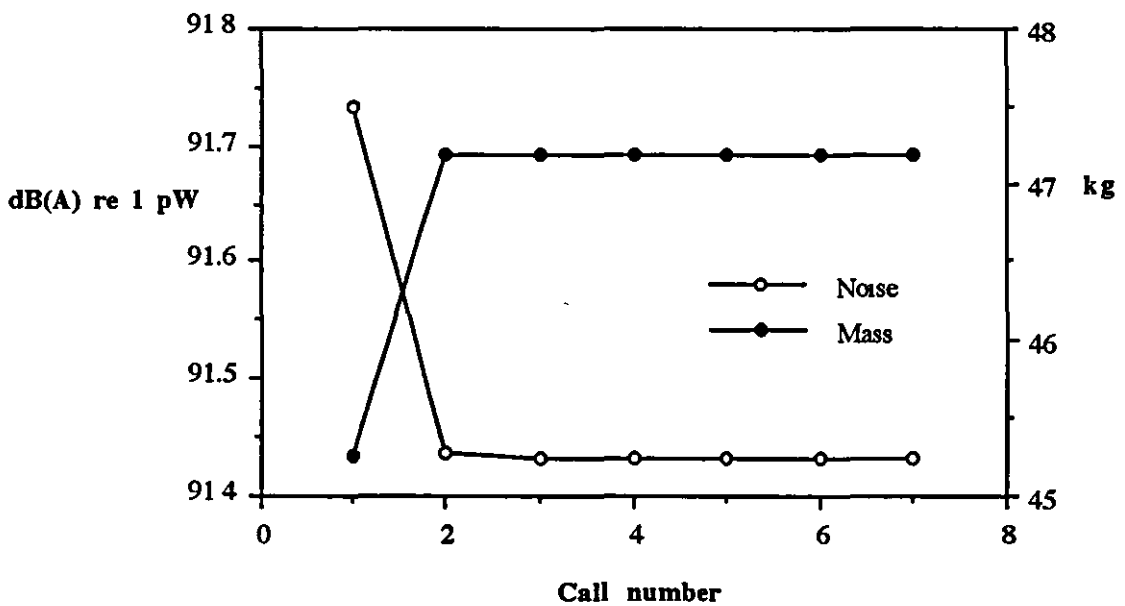

Figure 9.3 Search history - optimization with zero-change mass constraint starting from best constrained test point

This example demonstrates the advantage of commencing the optimization search from the analyser test which has the best constrained objective function value. If the search is started from the original design point, which has no variables at their optimum values, as shown in Table 9A.4, then nearly 50% more calls are required to locate the optimum, as shown in Figure 9.4.

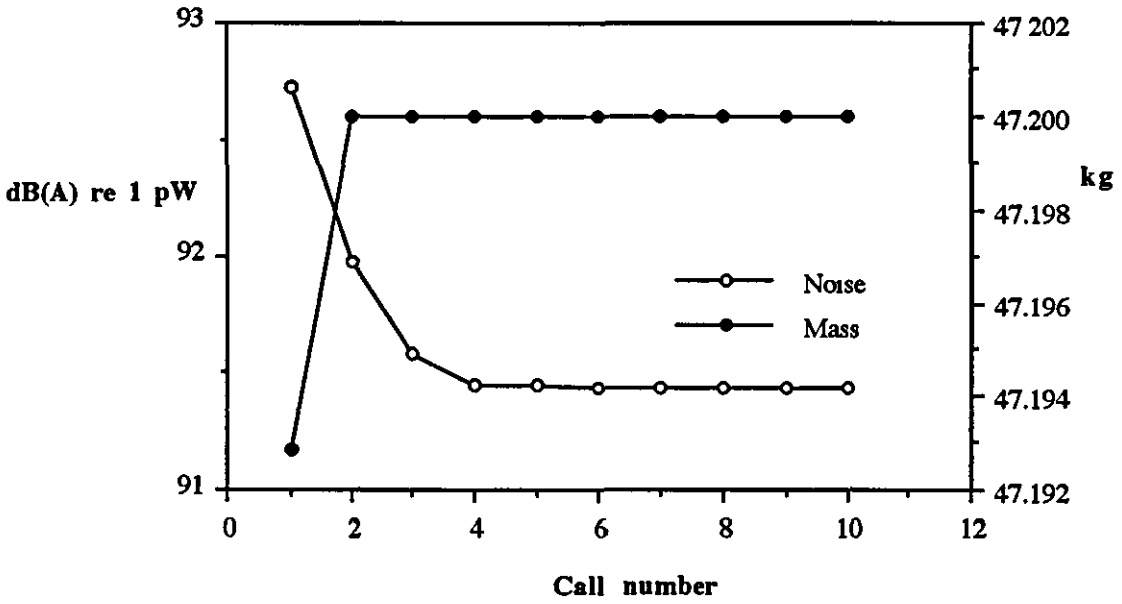

Figure 9.4 Search history - optimization with zero-change mass constraint starting from initial design point

The specification of the optimum which has been located, together with that of the initial design, is shown in Table 9.5. This shows that a predicted reduction in noise of 1.5 dB(A) has been gained, with no increase in mass. It is interesting to note that the optimizer has chosen to increase the thickness of bearings 1 and 5 (variable B), and yet reduce the thickness of the three middle bearings (variable C). The results also show that it is much more beneficial to increase the thickness of the crankcase skirt (variable A) than that of the longitudinal stiffener, even though this was included in the original design with the particular aim of reducing noise. These changes in variable value provide a good illustration of the need for a comprehensive numerical optimization approach to noise reduction, since without such a tool a designer would be extremely unlikely to choose such a combination of wall thicknesses.

	Variable					Mass	Noise
	A	B	C	D	G	kg	dB(A)
Initial	9.0	26.0	26.0	9.0	9.0	139.2	92.9
Optimum	12.0	32.0	20.0	6.0	6.6	139.2	91.4

Table 9.5 Initial and optimum designs under zero-change mass constraint

It can also be seen from Table 9.5 that, in contrast with the unconstrained optimization study, in which all design variables increased from their initial values, only some of the variables have gained in thickness. The others have not just remained unchanged, but have actually been reduced, even though, as shown by the coefficient values of Table 9A.2, their reduction leads to an *increase* in noise. This reduction is a direct result of the imposition of a functional constraint. A more detailed discussion of this characteristic is undertaken in Section 9.5, particularly with regard to its influence on the selection of design variables prior to optimization, and the investigation of multiple levels of a constraint function.

9.4 Mass optimization

In addition to carrying out a mass constrained optimization study, the flexibility of the response surface approach allows the engineer to investigate the optimum mass design for a given noise constraint without having to carry out any more expensive analyser calls. This ability may be of particular benefit if, for example, a design is required to conform to a maximum noise level imposed by legislation. Using the same response surfaces as those of the previous example, a mass optimization trial has been carried out using a noise constraint of 92.9 dB(A), which is equal to the noise level of the original design. The search history of this trial, starting from the initial design point, is shown in Figure 9.5, where, after allowing a temporary violation, the effect of the noise constraint can be clearly seen. The specification of the optimum design which results from this trial, together with that of the initial design is shown in Table 9.6.

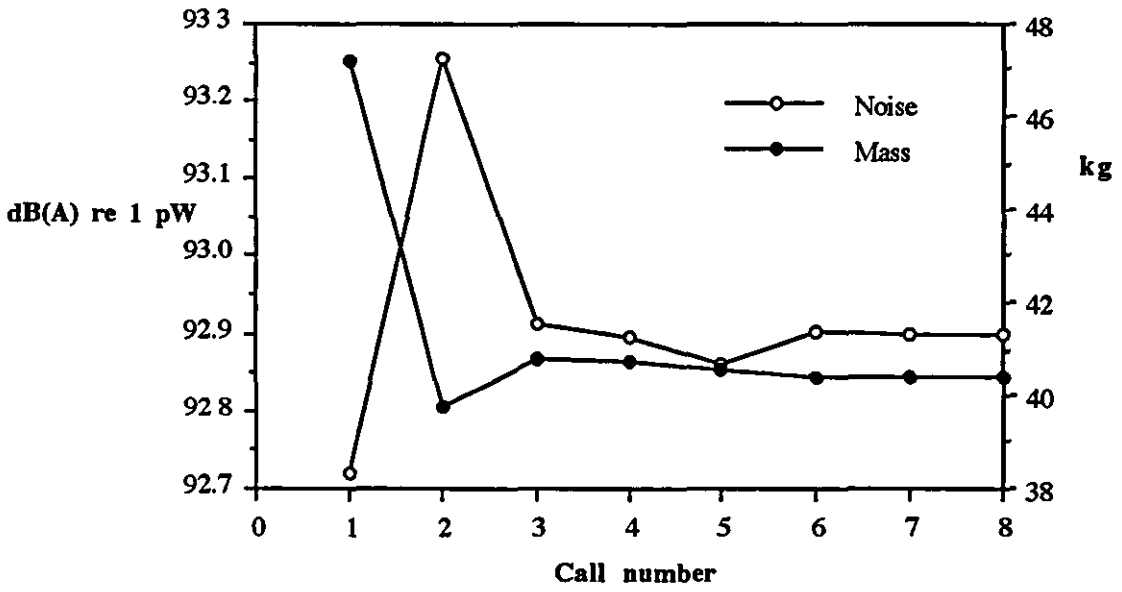

Figure 9.5 Search history - optimization with zero-change noise constraint starting from initial design point

	Variable					Mass	Noise
	A	B	C	D	G	kg	dB(A)
Initial	9.0	26.0	26.0	9.0	9.0	139.2	92.9
Optimum	12.0	20.0	20.0	6.4	6.0	132.4	92.9

Table 9.6 Initial and optimum designs under zero-change noise constraint

9.5 Selection of variables for optimization

The factors which influence the selection of design variables for various types of noise optimization study can be illustrated by considering in a simplified way a method by which an investigator might arrive at an optimum combination of design variable values. Suppose that a strictly linear (main effects only) response surface is being used, and that a minimum noise design is sought. Assume also that an increase in design variable value yields both an increase in mass and a decrease in noise, as in the examples presented above.

Suppose first that no functional constraints are applied, but that, as is always the case in practice, a limit is placed on the number of design variables which it is practical to investigate

with the available resources. In selecting which variables to include, the investigator simply seeks to choose those which will give the greatest reduction in noise with respect to changes in the parameter being varied, such as component thickness. Thus, variables which have a high noise sensitivity are of great interest, whilst those displaying a low noise sensitivity are of relatively little interest.

When carrying out a constrained optimization, however, the situation is substantially different, in that the investigator is not necessarily free to exploit all of the noise reduction potential of the design variables he has chosen. He must also consider the change which occurs in some other (constrained) function, such as structural mass, which must be kept within a predefined range. If a constraint level is applied which allows an increase in mass over the initial design, then, in modifying the component thicknesses of the initial design in order to reduce noise, the investigator might be considered as having a certain amount of 'free' mass to allocate. In choosing which variable(s) to allocate this 'free' mass to, he seeks to gain the maximum return, in terms of noise reduction, for the mass allocated. At this stage, the procedure for selection of variables is only a slight modification of that followed for unconstrained optimization, in that variables are now sought which will give the greatest reduction in noise with respect to changes in the structural mass, rather than in component thickness. Variables are thus selected which have a large noise/mass sensitivity.

Once this 'free' mass has been allocated, however, no more mass may be added to the system, and the only way in which further noise reductions can be gained is by reallocating that which is already there. The most efficient way in which to do this is to add mass to the variable which gives the *largest decrease* in noise per unit mass *increase*, whilst at the same time taking away mass from that variable which will give the *smallest increase* in noise per unit mass *decrease*. When either of these variables reaches its bound, mass is then added to (subtracted from) the next most (least) noise/mass sensitive variable. This process continues until a maximum of one variable remains unbounded, at which point the optimum has been identified for the mass constraint under consideration. Clearly, the noise gains to be derived from the 'high sensitivity' variables can only be exploited to the full if the number of 'low sensitivity' variables is sufficient to ensure that enough mass will be available for reallocation. Thus, in order to gain the 'best' optimum design for a given mass constraint, it is vital that the investigator include not only those variables which have the greatest noise/mass sensitivity, but also a number of variables which have a low noise/mass sensitivity.

This requirement is clearly substantially different from the method for selecting variables which is usually followed when carrying out a simple 'manual' or one-at-a-time optimization study. However, an understanding of this reallocation process is fundamental to an efficient use of any numerical tool for constrained optimization, whether it is based on the response surface methods used here, or employs a direct iterative approach.

The example used in this section is clearly very simplistic in nature, and differs from practical optimization studies in two main aspects. Firstly, the search methods employed by numerical optimization routines are far more sophisticated than that presented here. Additionally, when using response surfaces other than the simple main effects model, the noise/mass sensitivity of each variable will not be a fixed quantity, but will vary throughout the design region. However, neither of these considerations invalidate the observations made above.

When carrying out an optimization study which has multiple constraint functions, the method for selection of variables can be generalised as follows. Variables are first selected which are expected to have a beneficial effect on the objective function, although each of these may also incur substantial change in one or more of the constrained functions. In order to enable functional reallocation, a number of further variables are selected, each of which is expected to have a relatively low objective function sensitivity and a relatively high sensitivity in one or more of the constraint functions. In circumstances where a large number of constraint functions are to be applied, the selection of appropriate design variables is clearly an extremely complex procedure, and is made much more so by the fact that the relative functional sensitivities of various parts of the structure are unlikely to be known with great precision. In such circumstances, it may be useful to precede the formal numerical optimization methods considered in the present work with a sensitivity study of the initial design, at the finite element level, in order to ascertain which areas of the continuous block structure might most conveniently be grouped into a common design variable. This discussion highlights the fact that the expertise and experience of the individual investigator is still required when carrying out an optimization study, not so much to identify the best design for the chosen combination of variables, but to select these variables in the most advantageous manner.

As a practical example of the selection of design variables, consider again the mass-constrained optimization example shown in Section 9.3. Table 9.7 shows the bound-to-bound change in the noise and mass functions due to just the main effect coefficients, as well as the ratio of these. Note that these values would not change if the effect of pure quadratic terms were also included, since these would contribute an equal amount to the function value at each bound, and so would cancel; only the effect of interaction terms is omitted. This table shows that the noise/mass sensitivity of the first two variables is substantially higher than that of the others, as reflected in the optimum value of Table 9.5. The reason that the longitudinal stiffener was kept close to its low bound is also shown, even though the table confirms that it is having a significant effect in reducing noise. Its mass sensitivity is the greatest of the five variables, so that its ratio of noise reduction to mass increase is much reduced, and this explains why it is not favoured by the optimization routine.

	Variable	Mass kg range	Noise dB(A) range	dB(A)/kg
A	Crankcase skirt	6.40	-2.696	-0.421
B	1 & 5, end panels	6.38	-1.891	-0.296
C	Bearings 2, 3 and 4	3.98	-0.399	-0.100
D	Longitudinal stiffener	6.45	-0.831	-0.129
G	Water jacket sidewall	6.23	-0.223	-0.036

Table 9.7 An example of functional sensitivity

Of these five variables, the water jacket sidewall, variable G, was particularly selected because it was expected to have relatively low noise sensitivity, whilst also having high mass sensitivity due to its large area. This has been a particularly successful choice, in that, using the above figures, the penalty for removing mass from variable G is only just over one-third of that for the least sensitive of the remaining variables.

9.6 Investigating a range of constraint criteria

The above examples have served to confirm that the response surface approach can be successfully used to carry out single optimization studies of the type which are usually performed using direct iterative optimization techniques. When carrying out a single optimization trial, however, the response surface method does not have any particular advantages over the direct optimization method. Indeed, the discussions of Chapter 1 suggest that an advanced software program for direct iterative optimization is likely to find a constrained optimum with fewer function evaluations than would be required to establish an approximating response surface of sufficient accuracy. Thus, in cases where the value of each functional constraint can be precisely determined prior to optimization, so that the only requirement is the efficient identification of a single optimum, the direct iterative approach is almost certainly the best method to follow.

When carrying out a practical optimization study, however, especially in the early 'concept' stages of an engine design, it is unlikely that the constraint levels which are to be imposed will be known with any great precision. Considering the present application of noise optimization subject to a mass constraint, for example, the exact mass of an initial engine block specification is likely to be a result of the design process, and strength/durability considerations, rather than being a precisely defined target in itself. Thus, rather than being a rigid constraint which must be strictly adhered to, the mass of the structure can more appropriately be thought of as an additional variable in the optimization process. The goal of an optimization study, therefore, should not be restricted to the identification of one optimum, subject to a single mass

constraint, but should be to provide some indication as to how the optimum noise level will vary over a range of possible constraints, so that the designer has the necessary information to enable him to select an appropriate trade-off between these conflicting requirements.

If one were to seek to provide this information using the direct iterative approach, however, it would be necessary to carry out a separate optimization trial for each constraint value, or for each combination of values if multiple constraint functions were being considered. Even if the number of function evaluations required for each optimization trial were very low, the number of different constraint criteria which could practically be investigated would still be extremely limited. It is in this role that the response surface approach has been found to have the greatest advantage over the direct method, since, once an approximating response surface has been generated, a large number of different objective function and constraint level combinations can be investigated without the need to perform further evaluations of the original functions.

The power of this method is such that it is not limited to the investigation of a small number of discrete constraint values at, for example, ±2 kg mass change, ±4 kg mass change, etc., but can be used to generate what is effectively a continuous trade-off of optimum noise against structural mass throughout the entire range of feasible mass constraints. This may be carried out by performing a separate optimization trial at closely spaced increments of mass constraint which vary between the lowest possible mass value (which occurs when all variables are at their lower bound), and the highest possible mass value (when all variables are at their upper bound). This procedure may be termed a 'sweep' through the possible range of the constraint, and a numerical example of this method is given in the next section. The method can of course be extended to other constraint types and multiple constraint functions. Additionally, the flexibility of the procedure is such that, with no further analyser calls, the investigator may also carry out a mass optimization study of the type shown in Section 9.4, using radiated noise as a constraint function, or may simply evaluate the noise and mass functions for a variable combination which is of particular interest.

9.7 Numerical example of constraint sweep

To demonstrate the way in which a sweep throughout the entire range of a constraint may be performed in practice, the five variable example of Section 9.3 is again used, in which a 26 test Hoke design was employed to find a single optimum value, subject to a zero-change mass constraint. Using the same response surface models, based solely on the 26 tests of the Hoke design, Figure 9.6 shows the range of optimum noise values which results from carrying out optimization trials at intervals of 0.1 kg between the lowest possible mass of 124.5 kg and the highest possible mass of 154.0 kg. Figure 9.5 is thus the result of trials at 296 different constraint levels. The process of performing the complete sequence of optimization trials is

automated within the *optrsm* computer program, and using a Hewlett Packard 700 series Unix workstation takes just a few seconds of elapsed time, enabling it easily to be carried out interactively. It is notable that, although graphs of this type could be produced when using response surface methods to analyse the results of physical experimental programmes, no reference to such a use has been found in the published literature.

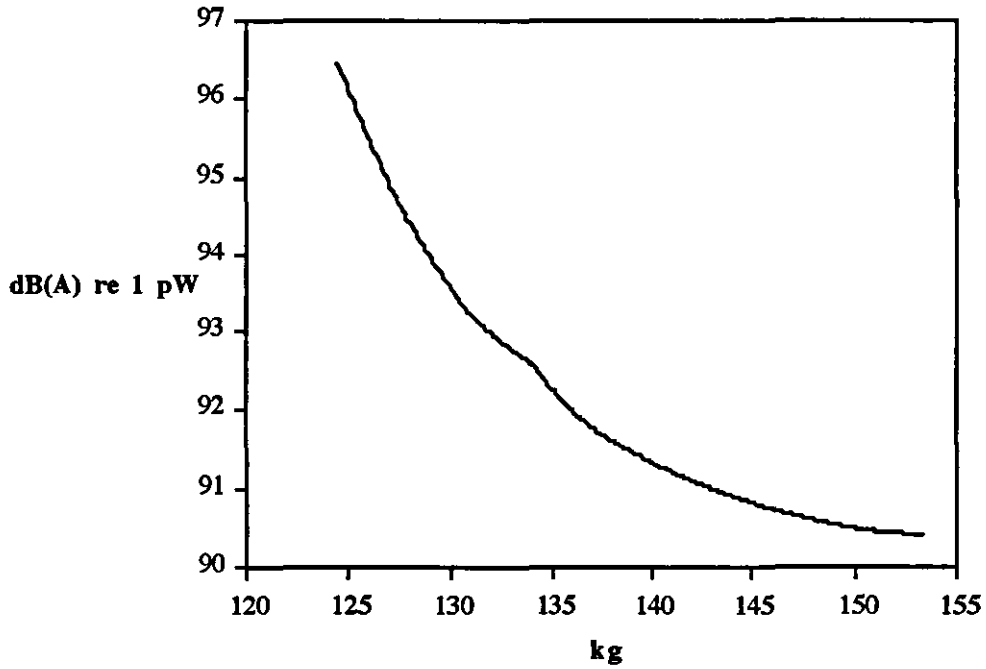

Figure 9.6 Noise/mass trade-off over complete range of possible mass constraints

The specification of each of the optimum designs which results from this investigation is given in Appendix 9B. Comparison of this table with Figure 9.6 shows that, although the change in the optimum noise value is very smooth, the change in the value of each of the variables can often be much more abrupt, as the optimizer switches to a different area of the design variable space in order to attain a slightly better function value. Figure 9.6 also shows that, as mass is added to the low-bounded design, the noise level falls fairly rapidly, since mass in being allocated to the most noise/mass sensitive of the variables. As the mass increases, these variables reach their upper bounds, and mass is then added to variables which have a lower noise/mass sensitivity, giving a smaller reduction in noise. The rate of noise reduction thus slows with increasing mass, as fewer variables remain to be increased.

In addition to the minimum-noise/mass trade-off, it is also possible to construct a graph of maximum noise against constraint value, as shown in Figure 9.7. This trade-off line is the result of a further 296 optimization trials, again using the mathematical models derived from the

same 26 function calls. The convergence of the two lines as the mass constraint reaches either of its extremes is, of course, a consequence of the fact that there is only one possible combination of variables which gives the maximum, or minimum, mass. The angle of inclination of these lines reflects the observation that an increase in mass will, in general, yield a decrease in radiated noise.

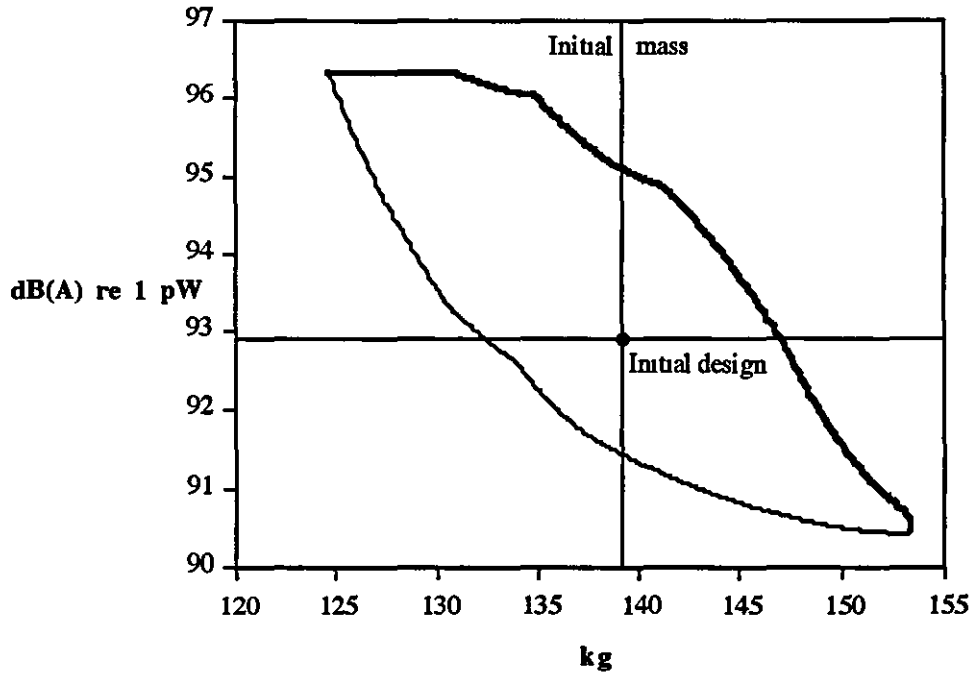

Figure 9.7 Maximum and minimum noise against structural mass

Although maximum-noise designs are clearly of little practical use, the graph of Figure 9.7 is of considerable interest, since it shows the noise range within which all possible designs must fall. The magnitude of this noise range highlights the need for a formal optimization tool, since, for a particular mass constraint, the choice of an inappropriate combination of design variables could result in a design which is up to 3.75 dB(A) worse than the optimum. As an example, consider the initial design, which lies close to the centre of both the mass and noise ranges, reflecting the fact that each of the five variables is at its mid-bound value. The vertical cross hair represents the zero-change mass constraint, and demonstrates that the worst no-mass-change design yields a noise level of 95.1 dB(A) – over 2 dB(A) worse than the original design and 3.7 dB(A) worse than the no-mass-change optimum. The specifications of these three designs are given in Table 9.8, showing the expected result that the variable values of the worst design are close to the opposite bound of those for the optimum. Figure 9.7 also shows that, as an alternative to a noise reduction of 1.5 dB(A) with no mass increase, it is also possible to reduce the mass by 7 kg without incurring a noise increase. If the distribution of

mass is not carried out correctly, however, it is also possible to add over 7.5 kg without achieving any reduction in noise. The range of possible no-noise-change designs is shown in Table 9.9. The optimum designs of Tables 9.8 and 9.9 are of course those designs identified in Sections 9.3 and 9.4, respectively.

	Variable					Mass	Noise
	A	B	C	D	G	kg	dB(A)
Worst	6.0	20.0	26.2	14.0	12.0	139.2	95.1
Initial	9.0	26.0	26.0	9.0	9.0	139.2	92.9
Optimum	12.0	32.0	20.0	6.0	6.6	139.2	91.4

Table 9.8 The range of possible no-mass-change designs

	Variable					Mass	Noise
	A	B	C	D	G	kg	dB(A)
Worst	6.0	31.0	32.0	14.0	12.0	146.8	92.9
Initial	9.0	26.0	26.0	9.0	9.0	139.2	92.9
Optimum	12.0	20.0	20.0	6.4	6.0	132.4	92.9

Table 9.9 The range of possible no-noise-change designs

9.8 Local minima problems

The previous constraint sweep study provides an example of the occurrence of local minima problems when using response surfaces. As discussed in Section 9.1, there is much less scope for local minima to exist in a low-order polynomial response surface than in the often complex function which it represents. Indeed, when carrying out an unconstrained optimization using a simple main effects model, no such local minima can occur. If the approximating polynomial is more complex, however, and especially if functional constraints are being applied, then a number of local minima may be found.

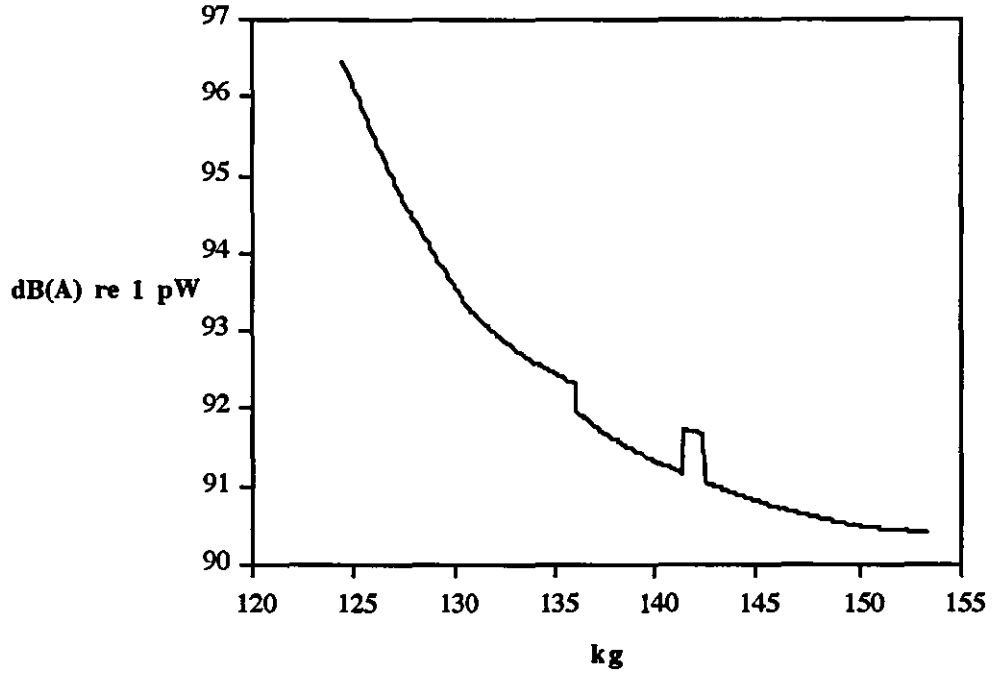

Figure 9.8 Noise/mass trade-off starting from 'best' constrained test point

As an example, Figure 9.8 shows the noise/mass trade-off graph which is obtained if each optimization trial of Section 9.7 is started from the 'best' test point which obeys the constraint, as described in Section 9.1. Because the full range of optimum designs is shown, any local minima which may occur are extremely easy to identify. There are clearly a number of constraint levels between 140 and 145 kg, for example, where a local minimum has been located, since it is impossible for an optimization trial with a more relaxed constraint to yield a worse optimum. A sudden drop in optimum objective function value is another indication of local minima which is displayed by this example. The sudden drop in function value just after 135 kg, preceded by a gradual swing away from what otherwise seems to be a well-defined curve, suggests that a problem is also occurring in this area.

As is the case when carrying out direct iterative optimization, the usual procedure for avoiding a local minimum is to restart the optimization trial from a different location. When using response surface methods, however, there are two major factors which give substantial assistance in solving this problem. These are as follows.

1. As shown by the example of Figure 9.8, the availability of a constraint sweep graph results in local minima being very easy to identify. This is in contrast with the direct iterative method, where one of the greatest problems with local minima is that it is often not possible to know whether the optimum which has been located is a global optimum within the design variable space, or is simply a local minimum.

2. The computational cost of carrying out additional optimization trials using an approximating response surface is negligible, since no further analyser calls are required. It is thus feasible to carry out a number of different trials for a given set of constraint criteria, each one starting from a different location in the design variable space. This procedure can easily be automated within a computer program, which might, for example, choose a number of random starting points and then select the best of these multiple optima in a way which is totally transparent to the user. This is again in contrast with the direct iterative technique, where the computational cost of performing each successive trial is approximately equal to the cost of the original.

To illustrate the way in which local minima problems may be addressed using response surface methods, Figures 9A.1 and 9A.2 of Appendix 9A show the noise/mass trade-off graphs which result from starting each optimization trial at the low bound of all variables and high bound of all variables, respectively. Whilst it can be seen that the second of these has its own minor local minima problem at around 134 kg, the combination of these two graphs with that of Figure 9.8 is likely to give a complete picture throughout the constraint range. These three plots are overlaid in Figure 9A.3 in order to confirm that they are coincident at all points other than those at which local minima have been identified. The 'true' noise/mass trade-off which results from this exercise is then the graph of Figure 9.6, used in the constraint sweep example of Section 9.7. Thus, although it is true, as stated in Section 9.7, that this graph is 'the result of trials at 296 different constraint levels', the number of trials carried out to produce this graph is in fact $296 \times 3 = 888$.

Even if a direct iterative optimization program could locate a single optimum in an average of 5 function calls, generation of this graph would still require 4440 calls. If just two calls per trial were needed, the number would reduce to 1776. Using the response surface approach, the required number of calls is 26.

9.9 Structural additions using discontinuous variables

The above examples have illustrated that the ability to identify the optimum combination of existing portions of an engine structure is an extremely powerful analytic tool when used at the concept design stage of an engine development programme. An important extension to this method, however, would be the ability to investigate the most beneficial combination of a number of structural additions to the base design, such as alternative means of main bearing support, or the relative merits of structural and non-structural sumps. In order to be able to carry out studies of this nature, however, the methods used in the above examples need to be extended in two main areas, as follows:

9.9.1 Modifications to the noise analysis program

The first area in which extensions to the existing methodology are required is in the evaluation of the function values, using the separate finite-element analyser program. In order to automate the process of function evaluation at each of the required test points, it is necessary that the specification of the initial model include the required information concerning each of the removable variables, with the removal of the appropriate elements being carried out within the finite-element program. This capability has been implemented within the analyser program by Zhang (1992). As noted in Section 1.8, it is also possible to perform manual modification of the design, if inclusion within an automated procedure proves infeasible.

9.9.2 Response surface representation of discontinuous variables

In order to investigate the addition or removal of component parts of the engine structure, the response surface method must be able to represent the effect of variables that can either take a zero value, or, if non-zero, can vary between upper and lower bounds in the normal manner. There would appear to be two distinct methods of approaching this problem, the first of which uses mathematical models of the type used in previous examples, and provides detailed information at the cost of a significant increase in function calls. The second aims at minimizing the number of function evaluations required, but provides less detailed information, and requires substantial modification to the methodology which has been used within the present work. The two approaches are as follows.

1. Investigation of discontinuous variables can be achieved within the current methodology by treating the problem as a main design variable space, in which all of the

variables vary between their bounds, together with a number of subspaces of reduced dimension, in which one or more variables are removed. An example of this approach is shown in Figure 9.9, where two of the three variables (x_1 and x_2) may be removed from the design. The main design variable space is thus three-dimensional, with the individual removal of variables x_1 and x_2 resulting in subspace 'a' and subspace 'b', respectively, each of which are two-dimensional. Additionally, the one-dimensional subspace 'c' is obtained when both x_1 and x_2 are removed together.

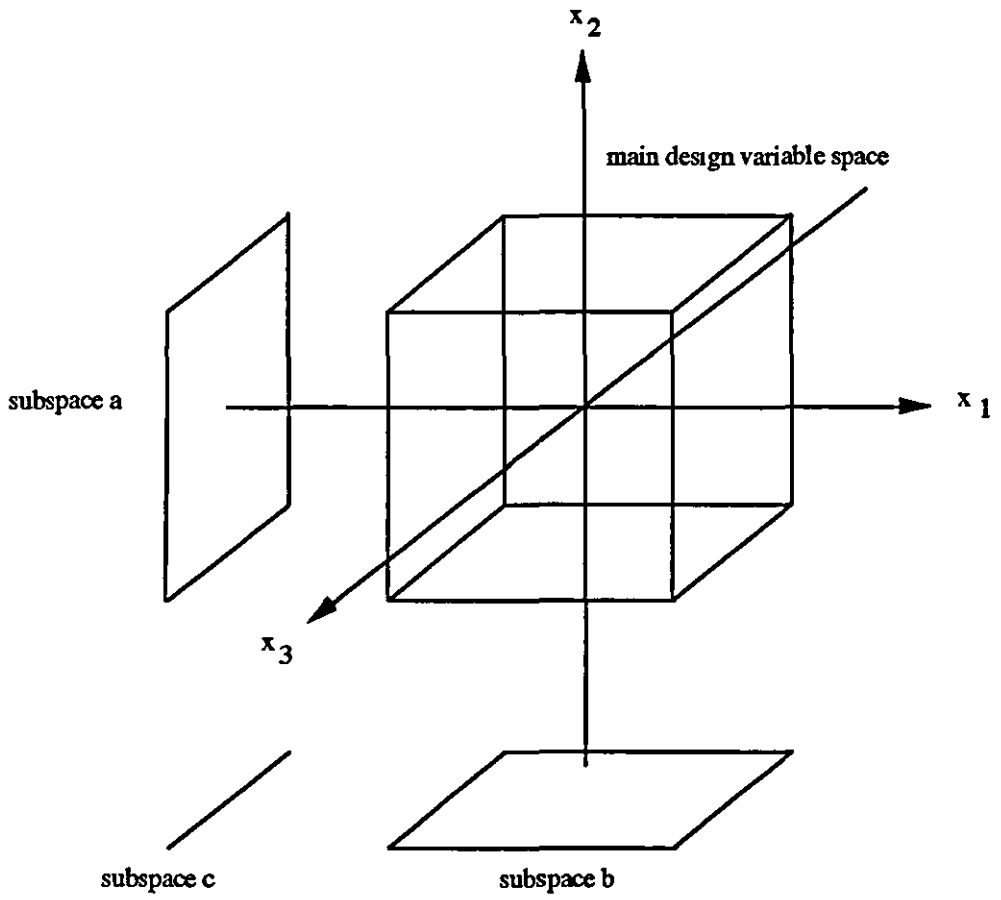

Figure 9.9 Investigation of discontinuous variables by division into subspaces

Although the use of this approach can potentially result in a large increase in the number of analyser calls which are required, there are two factors which suggest that the method will often be feasible in practise. Firstly, the number of subspaces which need to be investigated is determined by the number of variables which may take a zero value (for k such variables there are 2^k - 1 additional subspaces), and for practical problems the number of variables which may be removed from the structure will often be a small proportion of the total number of variables,

giving only a small number of subspaces. Additionally, each of these subspaces is of a reduced dimension, so that the number of tests required to investigate each diminishes rapidly as the number of variables which are simultaneously removed increases. A seven variable example of the use of this approach, in which minimally augmented Hoke designs are used to investigate each of the subspaces, is presented in the next section.

2. An alternative approach is to reformulate the definition of the response surface within the main design variable space in order to account explicitly for the effect of variable removal. Although this method is likely to require far fewer analyser calls than the subspace-investigation method, the use of such an approach raises the following two issues:

i) To minimize the required number of tests, the current subspace models are derived from Hoke designs, which are either saturated or minimally augmented. There is, therefore, little scope for further reduction of the number of tests, without sacrificing some of the terms of these subspace models. In order to represent the subspace response by the addition of a relatively small number of extra parameters to the main response surface definition, it would thus be necessary to make certain assumptions as to which of these effects could be neglected. This clearly calls for a detailed analysis of the nature of these subspace response surfaces, particularly with regard to their relationship with the main design variable space, and with each other.

ii) The inclusion of the effect of variable removal will necessarily lead to a response surface which exhibits a number of discontinuities. This a particular problem for iterative optimizer of the type which has been used here to carry out optimization on the response surfaces, since their methods of search are dependent on the analysis of local gradient information. The use of a single response surface model to represent variable discontinuity would thus require a substantially different approach to the final numerical optimization problem.

For these reasons, development of this second method of investigating the addition and removal of structural parts has not been addressed within the present work. It should be emphasised, however, that the subspace-investigation approach used here will, in fact, give a superior indication of the effect of variable removal, since the response in each subspace is being represented by a complete quadratic model. The disadvantage of this procedure is the large increase in the number of analyser calls which must be carried out if several discontinuous variables are to be investigated.

9.10 Numerical example using discontinuous variables

To illustrate the way in which an optimization study may be carried out using discontinuous variables, the seven variable example of Section 9.2 is again investigated. In the present example, however, each of the three methods of supporting the main bearing bulkheads (variables D, E and F) may be removed from the design. In order that all possible combinations of these stiffeners were investigated, the problem was broken down into eight separate sub-problems, as shown in Table 9.10. A Hoke D_4 design of the appropriate dimensionality was used to investigate each of the subspaces, with the number of tests used for each shown in Table 9.10. The mass range of each subspace is also shown.

subspace	Variable D	E	F	No. of variables	No. of tests	Mass range kg
main	•	•	•	7	43	142.9 – 185.3
a	–	•	•	6	34	140.3 – 176.3
b	•	–	•	6	34	140.7 – 179.7
c	•	•	–	6	34	126.7 – 159.4
d	–	–	•	5	26	138.1 – 170.7
e	–	•	–	5	26	124.2 – 150.4
f	•	–	–	5	26	124.5 – 153.9
g	–	–	–	4	19	121.9 – 144.8
					242	121.9 – 185.3

• variable included
– variable omitted

Table 9.10 subspaces used for investigation of main bearing support methods

A separate sweep through the range of possible mass constraints, at intervals of 0.1 kg, was carried out for each of the subspaces. Commencing optimization at three different locations for each of the 2612 different combinations of mass constraint and subspace, this gave a total of 7836 separate optimization trials, resulting in eight noise/mass trade-off graphs. These eight separate analyses were then combined by selecting the lowest optimum design for each mass constraint from among the subspaces, to give a single overall noise/mass trade-off graph. This graph is shown in Figure 9.10.

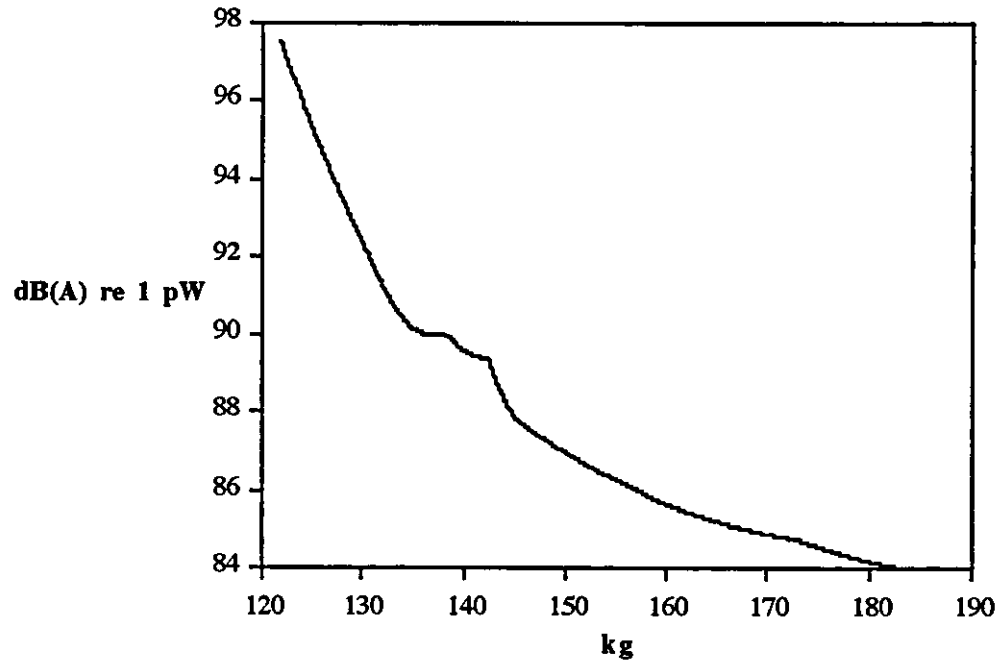

Figure 9.10 Noise/mass trade-off showing percentage change from initial mass

Figure 9.10 shows that, when compared with the noise/mass trade-off graph of Figure 9.6, a substantial 'kink' occurs at about 143 kg, somewhat similar to the local minima symptom of Figure 9.8. Figure 9.10 differs from this, however, in that there is no sudden drop back to the 'true' trade-off line. In this example, the reason for the apparent deviation is that Figure 9.10 is effectively displaying two curves of the form shown in Figure 9.6, with the sudden change in characteristic occurring when the bearing-tie (variable F) is brought into the design. Table 9.11 shows that this variable has a bound-to-bound mass change of 9.7 kg, equal to a rate of 0.81 kg/mm. Since the lower bound of this variable is 20 mm, the mass constraint must be at least (121.9 + 20 × 0.81) = 138.1 kg before it can be brought into the design.

	Variable	Mass kg range	Noise dB(A) range	dB(A)/kg
A	Crankcase skirt	6.40	-4.489	-0.702
B	1 & 5, end panels	6.38	-0.402	-0.063
C	Bearings 2, 3 and 4	3.98	-0.733	-0.184
D	Longitudinal stiffener	6.45	-0.433	-0.067
E	Lateral bearing support	3.34	-0.308	-0.090
F	Bearing cap tie	9.70	-0.481	-0.050
G	Water jacket sidewall	6.23	-0.663	-0.106

Table 9.11 Functional sensitivity when all variables are present

Although Table 9.11 shows that the noise/mass sensitivity of this variable is low once included in the design, there is clearly a large advantage in introducing it. This can be seen from the list of optimum designs in the range 142.1 to 143.0 kg given in Table 9.12, which show that in order to bring in variable F, the longitudinal stiffener (D) is removed from the design, the five bearing bulkheads (B and C) are reduced to their lowest bound, and the crankcase skirt also reduced significantly. The only way in which variable D could be introduced earlier would be to remove variable E from the design. This, however, is the main bearing lateral stiffener, and the choice of optimum design clearly shows that, like the bearing cap tie, there is much advantage in its presence, even at the lowest bound.

subspace	Variable - mm A	B	C	D	E	F	G	Mass kg	Noise dB(A)
c	12.0	32.0	27.9	4.0	10.0	0.0	6.0	142.1	89.40
c	12.0	32.0	28.2	4.0	10.0	0.0	6.0	142.2	89.39
c	12.0	32.0	28.5	4.0	10.0	0.0	6.0	142.3	89.38
c	12.0	32.0	28.8	4.0	10.0	0.0	6.0	142.4	89.38
c	12.0	32.0	29.1	4.0	10.0	0.0	6.0	142.5	89.37
a	8.2	20.0	20.0	0.0	10.0	20.0	6.0	142.6	89.35
a	8.3	20.0	20.0	0.0	10.0	20.0	6.0	142.7	89.26
a	8.4	20.0	20.0	0.0	10.0	20.0	6.0	142.8	89.18
a	8.5	20.0	20.0	0.0	10.0	20.0	6.0	142.9	89.11
a	8.6	20.0	20.0	0.0	10.0	20.0	6.0	143.0	89.03

Table 9.12 Introduction of the bearing cap tie (variable F)

Table 9.12 also shows the reason for the change in slope of the noise/mass trade-off graph when the bearing cap tie is introduced. As the mass constraint nears 142.5 kg, the most noise/mass sensitive variables of subspace 'c' are already at their upper bounds, so that the rate of noise reduction has slowed significantly. Once variable F is introduced, however, all variables are free to increase, and the rate of noise reduction suddenly rises as mass is again added to the most noise/mass sensitive variable (variable A).

A further investigation shows that the slope change at around 139 kg is due to a similar cause. Table 9.13 shows that the optimizer has located a minimum in subspace 'e', which it cannot escape from until enough mass is available to introduce another variable, in this case variable D, the longitudinal stiffener, which of course is removed again shortly afterwards to make way for the bearing cap tie (variable F).

subspace	Variable - mm A	B	C	D	E	F	G	Mass kg	Noise dB(A)
e	12.0	31.44	20.0	0.0	10.0	0.0	7.3	45.9	89.9817
e	12.0	31.46	20.0	0.0	10.0	0.0	7.3	46.0	89.9815
e	12.0	31.48	20.0	0.0	10.0	0.0	7.4	46.07	89.9814
e	12.0	31.48	20.0	0.0	10.0	0.0	7.4	46.07	89.9814
e	12.0	31.48	20.0	0.0	10.0	0.0	7.4	46.07	89.9814
e	12.0	31.48	20.0	0.0	10.0	0.0	7.4	46.07	89.9814
e	12.0	31.48	20.0	0.0	10.0	0.0	7.4	46.07	89.9814
e	12.0	31.48	20.0	0.0	10.0	0.0	7.4	46.07	89.9814
e	12.0	31.48	20.0	0.0	10.0	0.0	7.4	46.07	89.9814
c	11.7	31.29	20.0	4.0	10.0	0.0	6.0	46.8	89.9450
c	11.7	31.44	20.0	4.0	10.0	0.0	6.0	46.9	89.9073
c	11.8	31.59	20.0	4.0	10.0	0.0	6.0	47.0	89.8695
c	11.8	31.74	20.0	4.0	10.0	0.0	6.0	47.1	89.8317

Table 9.13 Introduction of the longitudinal stiffener (variable D)

The noise/mass trade-off graph of Figure 9.10 is again shown in Figure 9.11, with lines of percentage mass change superimposed. These highlight the fact that the inclusion of seven variables, of which three can be completely removed, leads to an extremely wide range of mass constraints. Even within the range of ±6% mass change, however, the noise level varies over a range of 4.5 dB(A). The specification of the optima which lie on these mass constraint lines are given in Table 9.14.

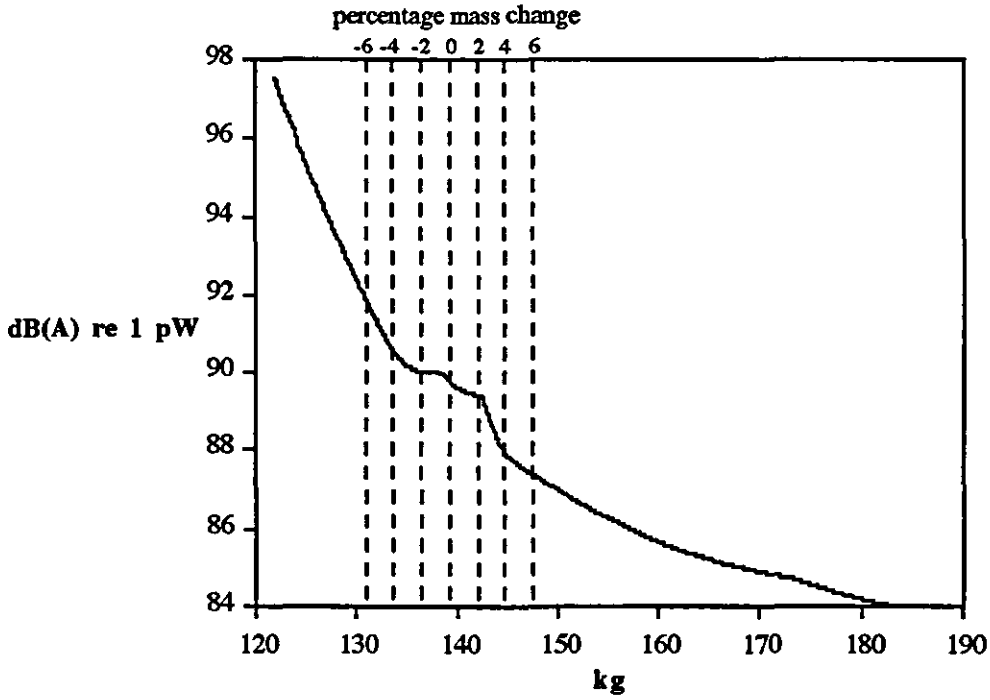

Figure 9.11 Noise/mass trade-off showing percentage change from initial mass

Variable - mm							Mass	%	Noise
A	B	C	D	E	F	G	kg	change	dB(A)
11.1	22.4	20.0	0.0	10.0	0.0	6.0	130.9	-6	91.9
12.0	25.8	20.0	0.0	10.0	0.0	6.0	133.6	-4	90.6
12.0	31.1	20.0	0.0	10.0	0.0	6.0	136.4	-2	90.0
11.8	31.9	20.0	4.0	10.0	0.0	6.0	139.2		89.8
12.0	32.0	27.3	4.0	10.0	0.0	6.0	142.0	+2	89.4
10.1	20.0	20.0	0.0	10.0	20.0	6.0	144.8	+4	88.0
11.3	20.0	24.8	0.0	10.0	20.0	6.0	147.6	+6	87.4

Table 9.14 A selection of optimum designs having mass change within ±6%

These optimum designs show that, as expected from the noise/mass sensitivity information of Table 9.11, the crankcase skirt and outside bearings (variables A and B) are identified by the optimization algorithm as the most efficient parts of the structure to which to add structural mass. As discussed above, the bearing cap tie (variable F) is introduced as soon as sufficient mass is available, but this is not until nearly a 2.5% mass increase has been incurred. Note that, even with a mass reduction of 6% over the initial design, a noise reduction of 1 dB(A) can still be achieved. Comparison of the no-mass-change optimum with that obtained using the five variables of Section 9.3, as shown in Table 9.15, reveals that the introduction of the lateral bearing support (variable E) has led to a further reduction of 1.6 dB(A), giving a design which is over 3 dB(A) quieter than the original.

	Variable - mm							Mass	Noise
	A	B	C	D	E	F	G	kg	dB(A)
Initial	9.0	26.0	26.0	9.0	0.0	0.0	9.0	139.2	92.9
n = 5	12.0	32.0	20.0	6.0	0.0	0.0	6.6	139.2	91.4
n = 7	11.8	31.9	20.0	4.0	10.0	0.0	6.0	139.2	89.8

Table 9.15 Initial design and no-mass-change optima with five or seven design variables

The increased rate of noise reduction which follows the introduction of the bearing cap tie suggests that further reductions in noise could be obtained at the lower mass constraints if this variable could be introduced with a lower thickness, and hence lower mass penalty. If this were the case, the optimum designs shown in Tables 9.12 and 9.13 suggest that there would no

longer be a need to introduce the longitudinal stiffener (variable D) in this mass range, thus allowing further mass reallocation to more noise/mass sensitive variables. In order to investigate these possibilities, a further series of tests were carried out, in which variable D was eliminated from the analysis, and variable F allowed to vary between bounds of 8 and 20 mm. The use of the lower bound of the original test range as the new upper bound allows a significant number of the original tests to be used. Although this additional study was investigated with a six-dimensional Hoke D_4 design comprising 34 tests, only 23 more calls to the analyser program were needed, cutting the additional computation time by one third.

A sweep through the mass range of the additional subspace was carried out, and the results combined with the previous trade-off graphs to give the family of optimum designs shown in Table 9.16. Comparison of these results with the optimum designs of Table 9.14 reveals that, except at the -6% mass constraint, all of the optimum designs lie in the additional subspace, resulting in a further reduction of 1.6 dB(A) for the no-mass-change constraint.

Variable - mm							Mass	%	Noise
A	B	C	D	E	F	G	kg	change	dB(A)
11.1	22.4	20.0	0.0	10.0	0.0	6.0	130.9	-6	91.9
8.8	20.0	20.0	0.0	10.0	8.0	6.0	133.6	-4	90.2
10.7	20.0	20.0	0.0	10.0	9.1	6.0	136.4	-2	89.0
11.3	20.0	20.0	0.0	10.0	11.6	6.0	139.2		88.2
12.0	20.0	20.0	0.0	10.0	14.1	6.0	142.0	+2	87.7
12.0	25.7	20.0	0.0	10.0	13.8	6.0	144.8	+4	87.4
11.6	20.0	32.0	0.0	10.0	14.1	8.0	147.6	+6	86.9

Table 9.16 Optimum designs obtained with an available bearing cap tie thickness of 8 - 20 mm

The noise levels of these optimum designs are shown graphically in Figure 9.12, where they are contrasted with that of the original design and the mass of the "worst" no-mass-change design, identified by carrying out a maximum-noise sweep through each of the design variable subspaces, as described in Section 9.7. The no-mass-change optimum is now 4.7 dB(A) quieter than the original design, and there is a range of nearly 7 dB(A) between the best and worst designs.

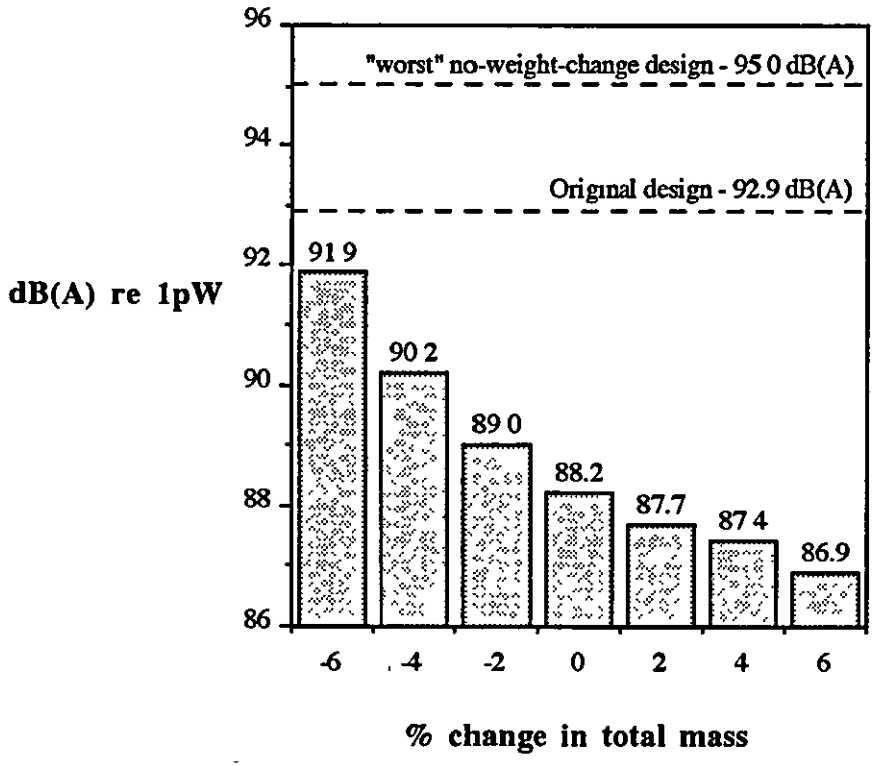

Figure 9.12 A range of optimum designs for varying mass constraints

9.11 Enhancements to the noise optimization procedure

The aim of the present chapter has been to demonstrate a number of the features of the response surface method which make it particularly suitable for carrying out an in-depth noise optimization study of a concept-stage engine design. In addition to these features, however, a number of extensions to the methodology are outlined below, which, if incorporated into the suite of computer programs developed as part of the present work, would significantly enhance the capabilities of the method.

9.11.1 Use of other function information

All of the examples presented in this chapter involve the use of just two functions of the design variables — radiated noise and structural mass. However, any function of the design variables can be used as either an objective or constraint function, as long as it fulfils the following three criteria:

1. Computable at the required test points.
2. Continuous function of the design variables.
3. Capable of being approximated by a low order polynomial.

The following are examples of additional functions which may be obtained from the noise analysis program used in the present work.

- Mass or weight of any individual component of the analysis structure.
- Dynamic response, such as displacement, velocity or acceleration, at specified nodes of the finite element mesh. This would be of particular interest when investigating, for example, the vibration transmission characteristics at engine mounting points, or the attachment points of engine covers.
- Maximum dynamic response within the structure.
- Individual natural frequency. If using a model of a complete powertrain, for example, an important use of this facility would be to ensure that the fundamental bending frequency exceeded the required design target. When carrying out noise optimization it is necessary to specify this requirement explicitly, since the low frequency of powertrain bending (usually below 400 Hz for light vehicles) is such that minimization of vibration in this range is of little consequence from a noise standpoint.

Although no detailed numerical investigations using the above functions have been carried out, preliminary trials suggest that they have considerable potential for inclusion in response surface-based optimization studies. In addition to the above, future work might investigate the inclusion of the following finite element analysis-based functions

- Alternative measures of radiated noise. This may include the use of unweighted noise values, different weighting scales, or a number of measures of sound 'quality' which are currently being used by investigators.
- Subdivision of the single sound power level into contributions from different areas of the structure surface. This would enable the use of a scaling factor to take into account, for example, variation in noise attenuation effects which occur when the unit is installed in a vehicle.

Although, in principal, measures of stress could be used as functions in the optimization process, the calculation of these values would require a finite element model of much greater sophistication than those currently used for vibration analysis of concept-stage engine structures. It is thus unlikely that the inclusion of stress information will be practical within the foreseeable future.

In addition to the use of alternative information supplied by the finite element program, the flexibility of the response surface method allows any other non-FE function which obeys the criteria outlined above to be included at the optimization stage. Functions of this nature might include, for example, some measure of cost, such as material cost, manufacturing cost, the effect of increased sales revenue resulting from a quieter engine, or some combination of these. The use of such functions is, of course, dependent on an ability to estimate them with the required accuracy.

In considering the inclusion of alternative functions, it is useful to bear in mind the following two features of the response surface method.

1. When the test points of the experimental design are analysed, all functions which can be calculated with little or no additional computational cost should be evaluated. If these are found to be of no use at the optimization stage, then little is lost. If even a single function is omitted, however, and later found to be required, then the cost of re-analysing each of the test points could be considerable.

2. At the analysis stage, there is no differentiation between the objective function and the constraint functions. It is not until the optimization trials are carried out, using the mathematical models of the original functions, that a choice of objective function, constraint functions and constraint levels must be made. Thus, because this optimization process incurs virtually no computational cost, it is feasible to carry out multiple studies of objective/constraint trade-offs using, for example, different measures of sound power and noise quality. This may yield a better understanding of the necessary compromises which must be made to give acceptable performance against a number of different criteria.

9.11.2 Multiple objective and constraint functions

Optimization of a single objective function subject to multiple constraints can be performed using the optimization routine employed in the above examples. A number of methods are currently available which allow the simultaneous optimization of two or more objective functions, subject to multiple constraints, although no routines of this nature are presently included in the NAG library. Discussion of this topic may be found in Khuri and Cornell (1987, Chapter 7), Hill and Hunter (1966), Myers, Khuri and Carter (1989). The

implementation of such a routine would, for example, allow the optimization of both radiated noise level and engine mount vibration velocities subject to a minimum powertrain bending frequency and a range of possible mass constraints.

9.11.3 Analytic optimization on the response surface

Since an exact specification of each of the approximating response surfaces is available, it would be feasible to solve these equations directly to give the required optimum, rather than using an iterative search algorithm. However, preliminary investigations of possible methods which have been undertaken within the present work have indicated that techniques such as the Kuhn-Tucker theory, whilst guaranteeing to find a constrained minimum, cannot guarantee to locate a global optimum (Banks, 1986). Since the routines currently used do not do so either, however, there would still seem to be scope for the use of such a method, especially if a reduction in solution time of the final optimization trial could be gained.

Appendix 9A

Optimization Information

Call	Variable - mm A	B	C	D	E	F	G	Mass kg	Noise dB(A)
1	6.0	20 0	20 0	4 0	10 0	20 0	6 0	142.87	91.603
2	12.0	21 4	22.1	5 3	10 6	20.7	6 3	152.56	87 117
3	12.0	22 5	24 3	6 2	12.1	22 0	7.4	156 99	86 691
4	12.0	23.7	26 5	7.2	13.7	23 4	8.5	161.60	86 242
5	12.0	24.9	28.7	8.1	15.2	25 0	9.6	166.37	85.767
6	12.0	26.1	31.0	9 0	16 7	26 8	10 8	171.32	85.260
7	12.0	27 3	32.0	9.9	18 2	28.7	12 0	175 96	84.808
8	12 0	28 5	32.0	10 7	19 7	30.8	12.0	179 23	84 567
9	12.0	31 4	32.0	12 5	23 2	32.0	12 0	183 66	84.236
10	12 0	32.0	32.0	14 0	25 0	32.0	12.0	185.32	84.121
11	12 0	32.0	32.0	14 0	25 0	32.0	12 0	185.34	84.120

Table 9A.1 Search history - unconstrained optimization in seven variables using a full factorial design (128 tests)

1.	MEAN	92.722	12.	G^2	1.0723×10^{-1}
2.	A	-1.3483	13.	BG	-6.8169×10^{-2}
3.	B	-9.4553×10^{-1}	14.	CD	-6.6992×10^{-2}
4.	D	-4.1588×10^{-1}	15.	AC	5.3423×10^{-2}
5.	A^2	2.6124×10^{-1}	16.	C^2	4.6248×10^{-2}
6.	AB	2.4908×10^{-1}	17.	CG	4.3133×10^{-2}
7.	C	-1.9983×10^{-1}	18.	AG	-3.5197×10^{-2}
8.	D^2	1.7826×10^{-1}	19.	AD	-1.5580×10^{-2}
9.	B^2	-1.6608×10^{-1}	20.	BC	7.3936×10^{-3}
10.	BD	1.2992×10^{-1}	21.	DG	-2.6639×10^{-3}
11.	G	-1.1148×10^{-1}			

Table 9A.2 Noise coefficients for Hoke D_4 design with 26 tests (n=5)

Call	Variable - mm A	B	C	D	G	Mass kg	Noise dB(A)
1	12.0	32 0	20 0	4 0	6 0	137.25	91.733
2	12.0	32 0	20 0	5 7	6.7	139 20	91.435
3	12 0	32 0	20 0	5 9	6 7	139.20	91.432
4	12.0	32 0	20 0	5 9	6.7	139 20	91.432
5	12.0	32.0	20 0	6 0	6.6	139.20	91 431
6	12 0	32.0	20 0	6 0	6.6	139 20	91 431
7	12 0	32.0	20 0	6.0	6 6	139 20	91.431

Table 9A.3 Search history, starting from best constrained test point

Call	Variable - mm A	B	C	D	G	Mass kg	Noise dB(A)
1	9 0	26 0	26 0	9 0	9 0	139 19	92.722
2	11 0	27 6	24.7	7.7	7.4	139 20	91 983
3	12 0	30 8	22 8	6.6	6.0	139 20	91 578
4	12 0	32 0	21.7	6 1	6 0	139 20	91 450
5	12.0	32 0	21 3	6 2	6 1	139 20	91.445
6	12.0	32 0	20.0	6 3	6 4	139 20	91.432
7	12.0	32 0	20.0	6 3	6.5	139.20	91.432
8	12 0	32 0	20 0	6 0	6.6	139.20	91.431
9	12 0	32 0	20 0	6.0	6.6	139.20	91.431
10	12 0	32.0	20 0	6 0	6 6	139.20	91 431

Table 9A.4 Search history, starting from initial design point

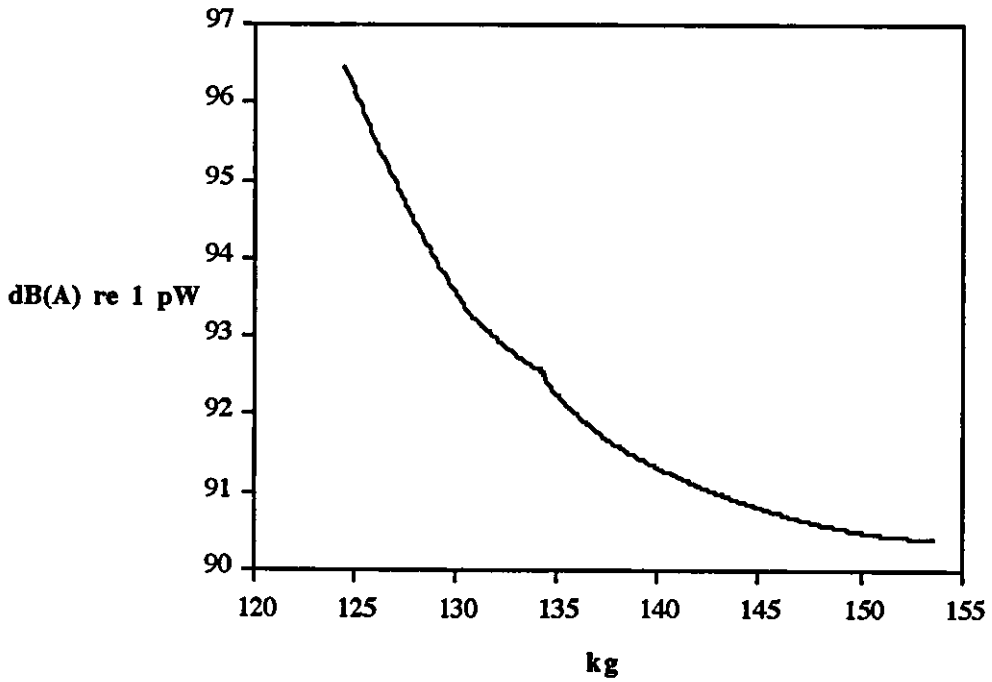

Figure 9A.1 Noise/mass trade-off starting from low variable bounds

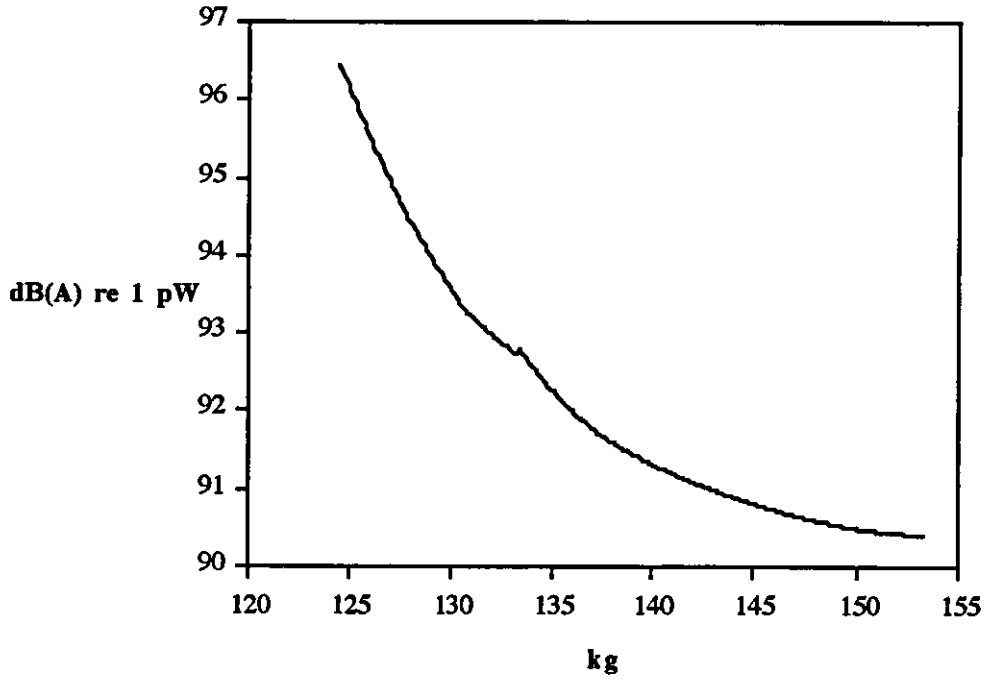

Figure 9A.2 Noise/mass trade-off starting from high variable bounds

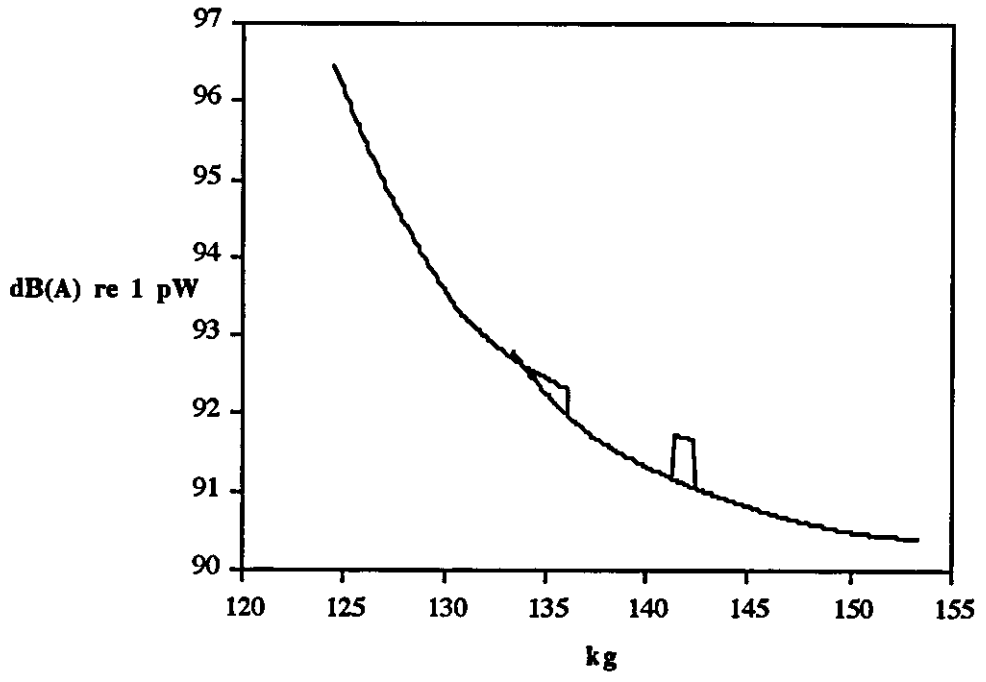

Figure 9A.3 Comparison of noise/mass trade-off graphs

Appendix 9B

Sweep through range of mass constraints
Five variable example

Variable - mm					Mass	Noise
A	B	C	D	G	kg	dB(A)
6 03	20 00	20 00	4 00	6 00	124 5	96 447
6 12	20 00	20 00	4 00	6 00	124 6	96 381
6 21	20 00	20 00	4 00	6 00	124 7	96 315
6 31	20 00	20 00	4 00	6 00	124 8	96 250
6 40	20 00	20 00	4 00	6 00	124 9	96 186
6 49	20 00	20 00	4 00	6 00	125 0	96 122
6 59	20 00	20 00	4 00	6 00	125 1	96 058
6 68	20 00	20 00	4 00	6 00	125 2	95 995
6 78	20 00	20 00	4 00	6 00	125 3	95 933
6 87	20 00	20 00	4 00	6 00	125 4	95 871
6 96	20 00	20 00	4 00	6 00	125 5	95 810
7 06	20 00	20 00	4 00	6 00	125 6	95 749
7 15	20 00	20 00	4 00	6 00	125 7	95 689
7 24	20 00	20 00	4 00	6 00	125 8	95 629
7 34	20 00	20 00	4 00	6 00	125 9	95 569
7 43	20 00	20 00	4 00	6 00	126 0	95 511
7 53	20 00	20 00	4 00	6 00	126 1	95 452
7 62	20 00	20 00	4 00	6 00	126 2	95 395
7 71	20 00	20 00	4 00	6 00	126 3	95 337
7 81	20 00	20 00	4 00	6 00	126 4	95 280
7 90	20 00	20 00	4 00	6 00	126 5	95 224
7 99	20 00	20 00	4 00	6 00	126 6	95 168
8 09	20 00	20 00	4 00	6 00	126 7	95 113
8 18	20 00	20 00	4 00	6 00	126 8	95 059
8 28	20 00	20 00	4 00	6 00	126 9	95 004
8 37	20 00	20 00	4 00	6 00	127 0	94 951
8 46	20 00	20 00	4 00	6 00	127 1	94 897
8 56	20 00	20 00	4 00	6 00	127 2	94 845
8 65	20 00	20 00	4 00	6 00	127 3	94 793
8 74	20 00	20 00	4 00	6 00	127 4	94 741
8 84	20 00	20 00	4 00	6 00	127 5	94 690
8 93	20 00	20 00	4 00	6 00	127 6	94 639
9 03	20 00	20 00	4 00	6 00	127 7	94 589
9 12	20 00	20 00	4 00	6 00	127 8	94 539
9 21	20 00	20 00	4 00	6 00	127 9	94 490
9 31	20 00	20 00	4 00	6 00	128 0	94 441
9 40	20 00	20 00	4 00	6 00	128 1	94 393
9 50	20 00	20 00	4 00	6 00	128 2	94 346
9 59	20 00	20 00	4 00	6 00	128 3	94 299
9 68	20 00	20 00	4 00	6 00	128 4	94 252
9 78	20 00	20 00	4 00	6 00	128 5	94 206
9 87	20 00	20 00	4 00	6 00	128 6	94 161
9 97	20 00	20 00	4 00	6 00	128 7	94 116
10 06	20 00	20 00	4 00	6 00	128 8	94 071
10 15	20 00	20 00	4 00	6 00	128 9	94 027
10 25	20 00	20 00	4 00	6 00	129 0	93 984
10 34	20 00	20 00	4 00	6 00	129 1	93 941
10 43	20 00	20 00	4 00	6 00	129 2	93 898
10 53	20 00	20 00	4 00	6 00	129 3	93 856
10 62	20 00	20 00	4 00	6 00	129 4	93 815
10 72	20 00	20 00	4 00	6 00	129 5	93 774
10 81	20 00	20 00	4 00	6 00	129 6	93 733
10 90	20 00	20 00	4 00	6 00	129 7	93 693
11 00	20 00	20 00	4 00	6 00	129 8	93 654
11 09	20 00	20 00	4 00	6 00	129 9	93 615
11 18	20 00	20 00	4 00	6 00	130 0	93 577
11 28	20 00	20 00	4 00	6 00	130 1	93 539
11 37	20 00	20 00	4 00	6 00	130 2	93 502
11 47	20 00	20 00	4 00	6 00	130 3	93 465
11 56	20 00	20 00	4 00	6 00	130 4	93 428
11 65	20 00	20 00	4 00	6 00	130 5	93 392
11 75	20 00	20 00	4 00	6 00	130 6	93 357
11 84	20 00	20 00	4 00	6 00	130 7	93 322
11 93	20 00	20 00	4 00	6 00	130 8	93 288
12 00	20 00	20 00	4 05	6 00	130 9	93 256
12 00	20 00	20 00	4 20	6 00	131 0	93 230
12 00	20 00	20 00	4 36	6 00	131 1	93 205
12 00	20 00	20 00	4 51	6 00	131 2	93 179
12 00	20 00	20 00	4 67	6 00	131 3	93 154
12 00	20 00	20 00	4 82	6 00	131 4	93 130
12 00	20 00	20 00	4 98	6 00	131 5	93 105
12 00	20 00	20 00	5 13	6 00	131 6	93 081
12 00	20 00	20 00	5 29	6 00	131 7	93 058
12 00	20 00	20 00	5 44	6 00	131.8	93 035
12 00	20 00	20 00	5 60	6 00	131 9	93 012
12 00	20 00	20 00	5 75	6 00	132 0	92 989
12 00	20 00	20 00	5 91	6 00	132 1	92 967
12 00	20 00	20 00	6 06	6 00	132 2	92 945
12 00	20 00	20 00	6 22	6 00	132 3	92 923
12 00	20 00	20 00	6 37	6 00	132 4	92 902
12 00	20 00	20 00	6 53	6 00	132 5	92 881
12 00	20 00	20 00	6 68	6 00	132 6	92 861
12 00	20 00	20 00	6 84	6 00	132 7	92 841
12 00	20 00	20 00	6 99	6 00	132 8	92 821
12 00	20 00	20 00	7 15	6 00	132 9	92 801
12 00	20 00	20 00	7 30	6 00	133 0	92.782

12 00	20 00	20 00	7 46	6 00	133.1	92 763
12 00	20 00	20 00	7 61	6 00	133 2	92 745
12 00	20 00	20 00	7 77	6 00	133 3	92 727
12 00	20 00	20 00	7 92	6 00	133 4	92 709
12 00	20 00	20 00	8 07	6 00	133 5	92 691
12 00	20 00	20 00	8 23	6 00	133 6	92 674
12 00	20 00	20 00	8 39	6 00	133.7	92 658
12 00	20 00	20 00	8 54	6 00	133 8	92 641
12 00	20 00	20 00	8 69	6 00	133 9	92 625
8 95	32 00	20 00	4 00	6 00	134 0	92 592
9 05	32 00	20 00	4 00	6 00	134 1	92 557
9 14	32 00	20 00	4 00	6 00	134 2	92 523
9 23	32 00	20 00	4 00	6 00	134 3	92 490
9 33	32 00	20 00	4 00	6 00	134 4	92 457
9 42	32 00	20 00	4 00	6 00	134 5	92 424
9 51	32 00	20 00	4 00	6 00	134 6	92 393
9 61	32 00	20 00	4 00	6 00	134 7	92 361
9 70	32 00	20 00	4 00	6 00	134 8	92 330
9 80	32 00	20 00	4 00	6 00	134 9	92 300
9 89	32 00	20 00	4 00	6 00	135 0	92 270
9 98	32 00	20 00	4 00	6 00	135 1	92 241
10 08	32 00	20 00	4 00	6 00	135 2	92 212
10 17	32 00	20 00	4 00	6 00	135 3	92 184
10 27	32 00	20 00	4 00	6 00	135 4	92 156
10 36	32 00	20 00	4 00	6 00	135 5	92 128
10 45	32 00	20 00	4 00	6 00	135 6	92 102
10 55	32 00	20 00	4 00	6 00	135 7	92 075
10 64	32 00	20 00	4 00	6 00	135 8	92 049
10 73	32 00	20 00	4 00	6 00	135 9	92 024
10 83	32 00	20 00	4 00	6 00	136 0	91 999
10 92	32 00	20 00	4 00	6 00	136 1	91 975
11 01	32 00	20 00	4 00	6 00	136 2	91.951
11 11	32 00	20 00	4 00	6 00	136 3	91 928
11 20	32 00	20 00	4 00	6 00	136 4	91 906
11 30	32 00	20 00	4 00	6 00	136 5	91 883
11 39	32 00	20 00	4 00	6 00	136 6	91 862
11 48	32 00	20 00	4 00	6 00	136 7	91 840
11 58	32 00	20 00	4 00	6 00	136 8	91 820
11 67	32 00	20 00	4 00	6 00	136 9	91 800
11 77	32 00	20 00	4 00	6 00	137 0	91 780
11 86	32 00	20 00	4 00	6 00	137 1	91 761
11 95	32 00	20 00	4 00	6 00	137 2	91 742
12 00	32 00	20 00	4 08	6 00	137 3	91 724
12 00	32 00	20 00	4 23	6 00	137 4	91 706
12 00	32 00	20 00	4 39	6 00	137 5	91 689
12 00	32 00	20 00	4 54	6 00	137 6	91 671
12 00	32 00	20 00	4 70	6 00	137 7	91 655
12 00	32 00	20 00	4 85	6 00	137 8	91 638
12 00	32 00	20 00	5 01	6 00	137 9	91 622
12 00	32 00	20 00	5 16	6 00	138 0	91 606
12 00	32 00	20 00	5 32	6 00	138 1	91 590
12 00	32 00	20 00	5 44	6 02	138 2	91 575
12 00	32 00	20 00	5 50	6 08	138 3	91 560
12 00	32 00	20 00	5 56	6 14	138 4	91 546
12 00	32 00	20 00	5 62	6 20	138 5	91 531
12 00	32 00	20 00	5 68	6 25	138 6	91 516
12 00	32 00	20 00	5 74	6 31	138 7	91 502
12 00	32 00	20 00	5 80	6 37	138 8	91 487
12 00	32 00	20 00	5 87	6 43	138 9	91 473
12 00	32 00	20 00	5 93	6 49	139 0	91 459
12 00	32 00	20 00	5 99	6 55	139 1	91 445
12 00	32 00	20 00	6 05	6 60	139 2	91 431
12 00	32 00	20 00	6 11	6 66	139 3	91 418
12 00	32 00	20 00	6 17	6 72	139 4	91 404
12 00	32 00	20 00	6 23	6 78	139 5	91 391
12 00	32 00	20 00	6 29	6 84	139 6	91 378
12 00	32 00	20 00	6 35	6 90	139 7	91 364
12 00	32 00	20 00	6 41	6 96	139 8	91 351
12 00	32 00	20 00	6 47	7 01	139 9	91 339
12 00	32 00	20 00	6 54	7 07	140 0	91 326
12 00	32 00	20 00	6 60	7 13	140 1	91 313
12 00	32 00	20 00	6 66	7 19	140 2	91 301
12 00	32 00	20 00	6 72	7 25	140 3	91 288
12 00	32 00	20 00	6 78	7 31	140 4	91 276
12 00	32 00	20 00	6 84	7 36	140 5	91 264
12 00	32 00	20 00	6 90	7 42	140 6	91 252
12 00	32 00	20 00	6 96	7 48	140 7	91 241
12 00	32 00	20 00	7 02	7 54	140 8	91 229
12 00	32 00	20 00	7 08	7 60	140 9	91 217
12 00	32 00	20 03	7 14	7 65	141 0	91 206
12 00	32 00	20 15	7 20	7 67	141 1	91 195
12 00	32 00	20 28	7 25	7 69	141 2	91.183
12 00	32 00	20 41	7 31	7 71	141 3	91 172
12 00	32 00	20 54	7 36	7 73	141 4	91 161
12 00	32 00	20 66	7 42	7 76	141 5	91 150
12 00	32 00	20 79	7 47	7 78	141 6	91 139
12 00	32 00	20 92	7 53	7 80	141 7	91 128
12 00	32 00	21 05	7 58	7 82	141 8	91 118
12 00	32 00	21 18	7 64	7 84	141 9	91 107
12 00	32 00	21 30	7 70	7 86	142 0	91 096
12 00	32 00	21 43	7 75	7 88	142 1	91 086
12 00	32 00	21 56	7 81	7 90	142 2	91 075
12 00	32 00	21 69	7 86	7 93	142 3	91 065
12 00	32 00	21 82	7 92	7 95	142 4	91 055
12 00	32 00	21 94	7 97	7 97	142 5	91 044
12 00	32 00	22 07	8 03	7 99	142 6	91 034
12 00	32 00	22 20	8 08	8 01	142 7	91 024
12 00	32 00	22 33	8 14	8 03	142 8	91 014
12 00	32 00	22 45	8 19	8 05	142 9	91 004
12 00	32 00	22 58	8 25	8 07	143 0	90 994
12 00	32 00	22 71	8.30	8 09	143 1	90 985
12 00	32 00	22 84	8 36	8 12	143 2	90 975
12 00	32 00	22 96	8 41	8 14	143 3	90 965
12 00	32 00	23 09	8 47	8 16	143 4	90 956
12 00	32 00	23 22	8 52	8 18	143.5	90 946

12 00	32 00	23 35	8 58	8 20	143 6	90 937
12 00	32 00	23 48	8 63	8 22	143 7	90 928
12 00	32 00	23 60	8 69	8 24	143 8	90 919
12 00	32 00	23 73	8 74	8 26	143 9	90 909
12 00	32 00	23 86	8 80	8 28	144 0	90 900
12 00	32 00	23 99	8 86	8 31	144 1	90 891
12 00	32 00	24 11	8 91	8 33	144 2	90 882
12 00	32 00	24 24	8 96	8 35	144 3	90 874
12 00	32 00	24 37	9 02	8 37	144 4	90 865
12 00	32 00	24 50	9 08	8 39	144 5	90 856
12 00	32 00	24 63	9 13	8 41	144 6	90 848
12 00	32 00	24 75	9 19	8 43	144 7	90 839
12 00	32 00	24 88	9 24	8 45	144 8	90 831
12 00	32 00	25 01	9 30	8 48	144 9	90 822
12 00	32 00	25 14	9 35	8 50	145 0	90 814
12 00	32 00	25 27	9 41	8 52	145 1	90 806
12 00	32 00	25 39	9 46	8 54	145 2	90 798
12 00	32 00	25 52	9 52	8 56	145 3	90 789
12 00	32 00	25 65	9 57	8 58	145 4	90 782
12 00	32 00	25 78	9 63	8 60	145 5	90 774
12 00	32 00	25 91	9 68	8 62	145 6	90 766
12 00	32 00	26 03	9 74	8 64	145.7	90 758
12 00	32 00	26 16	9 79	8 67	145 8	90 750
12 00	32 00	26 29	9 85	8 69	145 9	90 743
12 00	32 00	26 42	9 90	8 71	146 0	90 735
12 00	32 00	26 54	9 96	8 73	146 1	90 728
12 00	32 00	26 67	10 01	8 75	146 2	90 720
12 00	32 00	26 80	10 07	8 77	146 3	90 713
12 00	32 00	26 93	10 12	8 79	146 4	90 706
12 00	32 00	27 06	10 18	8 81	146 5	90 699
12 00	32 00	27 18	10 24	8 84	146 6	90 692
12 00	32 00	27 31	10 29	8 86	146 7	90 685
12 00	32 00	27 44	10 35	8 88	146 8	90 678
12 00	32 00	27 57	10 40	8 90	146 9	90 671
12 00	32 00	27 69	10 46	8 92	147 0	90 664
12 00	32 00	27 82	10 51	8 94	147 1	90 658
12 00	32 00	27 95	10 57	8 96	147 2	90 651
12 00	32 00	28 08	10 62	8 98	147 3	90 644
12 00	32 00	28 20	10 68	9 01	147 4	90 638
12 00	32 00	28 33	10 73	9 03	147 5	90 632
12 00	32 00	28 46	10 79	9 05	147 6	90 625
12 00	32 00	28 59	10 84	9 07	147 7	90 619
12 00	32 00	28 72	10 90	9 09	147 8	90 613
12 00	32 00	28 84	10 95	9 11	147 9	90 607
12 00	32 00	28 97	11 01	9 13	148 0	90 601
12 00	32 00	29 10	11 06	9 15	148 1	90 595
12 00	32 00	29 23	11 12	9 18	148 2	90 589
12 00	32 00	29 36	11 17	9 20	148 3	90 584
12 00	32 00	29 48	11 23	9 22	148 4	90 578
12 00	32 00	29 61	11 28	9 24	148 5	90 572
12 00	32 00	29 74	11 34	9 26	148 6	90 567
12 00	32 00	29 87	11 39	9 28	148 7	90 561
12 00	32 00	30 00	11 45	9.30	148 8	90 556
12 00	32 00	30 12	11 50	9 32	148 9	90 551
12 00	32 00	30 25	11 56	9 34	149 0	90 545
12 00	32 00	30 38	11 61	9 37	149 1	90 540
12 00	32 00	30 51	11 67	9 39	149 2	90 535
12 00	32 00	30 63	11 73	9 41	149 3	90 530
12 00	32 00	30 76	11 78	9 43	149 4	90 525
12 00	32 00	30 89	11 84	9 45	149 5	90 521
12 00	32 00	31 02	11 89	9 47	149 6	90 516
12 00	32 00	31 15	11 95	9 49	149 7	90 511
12 00	32 00	31 27	12 00	9 51	149 8	90 507
12 00	32 00	31 40	12 06	9 53	149 9	90 502
12 00	32 00	31 53	12 11	9 56	150 0	90 498
12 00	32 00	31 66	12 17	9 58	150 1	90 493
12 00	32 00	31.78	12 22	9 60	150 2	90 489
12 00	32 00	31.91	12 28	9 62	150 3	90 485
12 00	32 00	32 00	12 33	9 65	150 4	90 481
12 00	32 00	32 00	12 40	9 71	150 5	90 477
12 00	32 00	32 00	12 46	9 77	150 6	90 473
12 00	32 00	32.00	12 52	9 83	150 7	90 469
12 00	32 00	32 00	12 58	9 89	150 8	90 465
12 00	32 00	32 00	12 64	9 94	150 9	90 462
12 00	32 00	32 00	12 70	10 00	151 0	90 459
12 00	32 00	32 00	12 76	10 06	151.1	90 455
12 00	32 00	32 00	12 82	10 12	151 2	90 452
12 00	32 00	32 00	12 88	10 18	151 3	90 449
12 00	32 00	32 00	12 94	10 24	151 4	90 447
12 00	32 00	32 00	13 00	10.30	151 5	90 444
12 00	32 00	32 00	13 07	10 35	151 6	90 441
12 00	32 00	32 00	13 13	10 41	151 7	90 439
12 00	32 00	32 00	13.19	10 47	151 8	90 437
12 00	32 00	32 00	13 25	10 53	151 9	90 435
12 00	32 00	32 00	13 31	10 59	152 0	90 433
12 00	32 00	32 00	13 37	10 65	152 1	90 431
12 00	32 00	32 00	13 43	10 70	152 2	90 429
12 00	32 00	32 00	13 49	10 76	152 3	90 427
12 00	32 00	32 00	13 55	10 82	152 4	90 426
12 00	32 00	32 00	13 61	10 88	152 5	90 424
12 00	32 00	32 00	13 67	10 94	152 6	90 423
12 00	32 00	32 00	13 74	11 00	152 7	90 422
12 00	32 00	32 00	13 80	11 06	152 8	90 421
12 00	32 00	32 00	13 86	11 11	152 9	90 420
12 00	32 00	32 00	13 92	11 17	153 0	90 420
12 00	32 00	32 00	13 98	11 23	153 1	90 419
12 00	32 00	32 00	14 00	11.31	153 2	90 419
12 00	32 00	32 00	14 00	11 41	153.3	90 419
12 00	32 00	32 00	14 00	11 44	153 3	90 419
12 00	32 00	32 00	14 00	11 44	153 3	90 419
12 00	32 00	32 00	14 00	11 44	153 3	90 419
12 00	32 00	32 00	14 00	11 44	153 3	90 419
12 00	32 00	32 00	14 00	11 44	153 3	90 419
12 00	32 00	32 00	14 00	11 44	153 3	90 419
12 00	32 00	32 00	14 00	11 44	153.3	90 419

10. Summary and conclusions

The work presented within this thesis has served to demonstrate the potential of response surface methods for carrying out numerical optimization of engine structures for reduced noise. Investigation of a selection of existing experimental designs and classes of response surface model has identified a suitable combination with which to address the noise optimization problem. Using this approach, the variation of the radiated noise function throughout the design variable space can be accurately represented with an acceptable number of analyser calls. Optimization using the low-order mathematical model of the original responses offers a level of functionality not currently available using traditional direct iterative optimization techniques.

10.1 Summary of results

The main conclusions which may be drawn from the theoretical studies and numerical tests which have been carried out within the present work are as follows. Conclusions drawn from numerical studies are derived from the particular engine structure which has been investigated.

Chapter 4: Experiments using first-order designs and models

- The mass function is exactly representable using mean and main effect terms.
- Virtually all of the variation in the noise function at the vertices of the design variable space is attributable to main effect and two-way interaction terms.
- The use of probability plots provides a convenient method of identifying statistically significant model coefficients.
- An independent analysis confirms that the significant parameters are the only ones which substantially influence the accuracy of the predictive model.
- Such a model can be constructed from a fractional factorial design of at least resolution V. This also allows for the inclusion in the model of a number of higher order interaction terms.
- The use of a first-order model results in substantial lack of fit at locations which are distant from the test points. The inclusion of higher order terms is required to improve the predictive ability of the model .

Chapter 6: The Central Composite Design (CCD)

- The usefulness of the CCD is adversely affected by the constraints which the nature of the noise analysis problem place upon the axial parameter α and the number of centre point tests n_0. The result of these constraints is that it is not possible to specify designs which are either orthogonal or rotatable.
- Numerical tests have shown that a seven variable CCD with a resolution VII factorial portion, an axial parameter of $\alpha = 1$, and a single centre point can be used to construct a second order model which gives a close approximation to the original noise response throughout the region of interest.
- The predictive ability of the model can be further enhanced by removing those coefficients which the probability plot technique has found to be statistically insignificant.
- Further results suggest that when the identity of the significant terms is not known, little lack of performance results from the substitution of a strict quadratic model, in which no linear interaction terms are included of order higher than two.
- The small number of tests which involve a third level of each of the design variables leads to a low level of precision in the estimation of the pure quadratic terms of the model. This is of particular importance, since many of the quadratic coefficients have a dominant effect on the performance of the model, and yet, due to the relative imprecision of their estimated values, it is not possible to say with certainty whether their inclusion is justified by the available data.
- The inclusion of replicate centre points gives no additional information when investigating a deterministic application .

Chapter 7: Centre point replication

- It is possible to obtain genuine additional information concerning the variation of the noise function by performing tests at points close to the centre of the design variable space.
- Modification of the CCD to include an ε-star or ε-pair portion can be made without significantly compromising the orthogonal or rotatable nature of the design.
- In the specific case investigated, little improvement in model accuracy has been obtained over that of the approximating model derived from the standard CCD, due to the small lack of fit which occurs at points close to the centre of the region.
- If the original response surface exhibits a substantial amount of high-frequency fluctuation, then repeated testing over a small area may result in a significant gain in the

predictive accuracy of the fitted model.

- The most effective use of additional test points is not to provide an increase in the density of testing in certain areas, but to improve the distribution of test points throughout the entire design variable space.

Chapter 8: Economic second-order designs

- The prediction accuracy of each Hoke design is similar to that of the CCD, despite requiring far fewer tests. On the basis of prediction errors, no reason was found for choosing the minimally augmented design rather than the saturated design, which requires fewer tests.
- The use of a minimally augmented design is recommended in order to provide lack-of-fit information.
- The variance for the pure quadratic terms of the saturated Hoke designs are twice the size of the corresponding values for the CCD, whilst the minimally augmented design yields variances similar to the CCD.
- The way in which the pure quadratic columns of the regressor matrix are scaled has a critical influence on the selection of an optimum design. If an alternative scaling factor is adopted, then the design D_4 is superior to the previously optimum design D_6. Other designs of the full set investigated by Hoke are likely to perform better under this new scaling method.
- Hoke's economical second order designs have been found to provide an extremely useful alternative to the CCD, yielding approximately the same accuracy of prediction whilst having a much reduced test requirement.

Chapter 9: Optimization

- A standard iterative optimization algorithm can successfully be used to locate optima on the approximating response surface.
- Unconstrained noise optimization is seldom of practical use due to the associated increase in structural mass.
- The response surface approach can be successfully used to carry out single optimization studies of the type usually performed using direct iterative optimization.
- In order to gain the 'best' optimum noise design for a given mass constraint, it is important to include not only variables which have a high noise/mass sensitivity, but also a number of variables which have a low noise/mass sensitivity.

- The response surface approach can investigate a wide range of different constraint criteria without the need to perform additional finite element analyses. A 'sweep' through the possible range of constraints gives an explicit trade-off between competing design objectives, such as low noise and low mass.
- The availability of a constraint sweep graph allows local minima to be easily identified. When using the response surface approach, the cost of carrying out additional optimization trials to avoid local minima is very small.
- Discontinuous variables can be incorporated within an optimization study, thus allowing investigation of structural additions to the base engine design.
- A number of additional classes of response function have the potential for incorporation within a response surface-based optimization study.

10.2 Recommendations for further work

- All of the statistical analyses which have been carried out within the present work have been based on the assumption that errors due to underspecification of the approximating model can be treated as though they are randomly distributed and uncorrelated. Whilst adoption of this assumption has allowed the application of a number of standard statistical techniques which are based on this error model, it was recognised in Appendix 4C that the assumption is not strictly valid in the deterministic environment of computer simulation. A more rigorous approach to this problem, together with the development of an alternative approach to the modelling of errors, would represent an important extension to the investigations made within the current work.
- As with the adoption of a more appropriate error model, it is possible that the deterministic nature of computer experimentation can be better addressed by alternative approaches to experimental designs. The field of computer simulation would, for example, appear to hold considerable potential for the application of sequential experimentation techniques, in which the location of future test points is determined by information obtained from previous tests.
- Within the present work, little use has been made of the various measures of design optimality which are employed to judge competing designs. Further work in this area is required, particularly with regard to the effect of scaling on the performance of the Hoke designs, as discussed in Section 8.4.3 and Appendix 8B.
- If a different optimality criterion is adopted for judging the Hoke designs, it is likely that the full range of possible designs will need to be re-examined in order to identify which perform best under the new criterion.

- The method used for dealing with discontinuous variables provides detailed information on the characteristics of the response functions within the design variable subspaces, but at the cost of a substantial increase in the number of function evaluations which are required. Further work is necessary in order to assess whether the removal of structural parts can be accounted for within the response surface model of the main design space without sacrificing overall predictive ability.
- Preliminary work has been carried out to investigate the inclusion of a number of different response functions within a response surface-based optimization study, although no results have been presented within the present work. Further investigation is required in order to establish the range of such functions which lend themselves to approximation using low-order mathematical models.
- A number of different approaches to optimization of multiple objective functions have been reported in the literature. Incorporation of these techniques is likely to lead to a significant increase in functionality when carrying out a comprehensive optimization study involving a large range of potential objectives.
- Optimization on the response surface is presently performed using a standard iterative algorithm from a commercial subroutine library. Further work is required to assess whether more appropriate techniques might be applied to this final optimization problem. In particular, the capability to carry out analytic optimization on the response surface is likely to offer a substantial advantage in efficiency and functionality.

References

Addleman, S., (1961), "Irregular fractions of the 2^n factorial experiments", *Technometrics*, 3, 479–496.

Addleman, S., (1962a), "Orthogonal main-effect plans for asymmetrical factorial experiments", *Technometrics*, 4, 21–46.

Addleman, S., (1962b), "Symmetrical and asymmetrical fractional factorial plans", *Technometrics*, 4, 47–57.

Addleman, S., (1963), "Techniques for constructing fractional replicate plans", *J. Am. Statist. Assoc.*, 58, 45–71.

Addleman, S., (1964), "Designs for the sequential application of factors", *Technometrics*, 6, 365–370.

Addleman, S., (1969), "Sequences of two-level fractional factorial plans", *Technometrics*, 11, 477–509.

Arap Koske, J.K., (1987), "A fourth order rotatable design in four dimensions", *Communications in Statistics — Theory and Methods*, A16, 2747–2753.

Arap Koske, J.K., Patel, M.S. (1987), "Construction of fourth order rotatable designs with estimation of corresponding response surface", *Communications in Statistics — Theory and Methods*, A16, 1361–1376.

Atkinson, A.C., (1982), "Developments in the designs of experiments", *Int. Statist. Rev.*, 50, 161–177.

Banks, S.P., (1986), *Control Systems Engineering*, Prentice-Hall, London.

Bandemer, H., (1980), "Problems in foundation and use of optimal experimental design in regression models", *Mathematische Operationsforschung und Statistik Series Optimization*, 11, 89–113.

Bates, D.M., Watts, D.G., (1985), "Multiresponse Estimation With Special Application to Linear Systems of Differential Equations", *Technometrics*, 27, 329–360.

Biles, W.E., (1975), "A Response Surface Method for Experimental Optimization of Multi-Response Processes", *Industrial and Engineering Chemistry Process Design and Development*, 14, 152–158.

Birnbaum, A., (1959), "On the Analysis of Factorial Experiments Without Replication", *Technometrics*, 1, 343–357.

Box, G.E.P., (1952), "Multifactor designs of first order", *Biometrika*, 39, 49–57.

Box, G.E.P., (1954), "The exploration and exploitation of response surfaces: some general considerations and examples", *Biometrics*, 10, 16–60.

Box, G.E.P., Behnken, D.W., (1960), "Some New Three Level Designs for the Study of Quantitative Variables", *Technometrics*, 2, 455–475.

Box, G.E.P., Draper, N.R., (1959), "A basis for the selection of a response surface design", *J. Am. Statist. Assoc.*, 54, 622–654.

Box, G.E.P., Draper, N.R., (1965), "The Bayesian estimation of common parameters from several responses", *Biometrika*, 52, 355–365.

Box, G.E.P., Draper, N.R., (1987), *Empirical Model-Building and Response Surfaces*, Wiley, New York.

Box, G.E.P., Hunter, W.G., Hunter, J.S., (1978), *Statistics For Experimenters*, Wiley, New York.

Box, G.E.P., Hunter, J.S., (1957), "Multi-Factor Experimental Designs for Exploring Response Surfaces", *The Annals of Mathematical Statistics*, 28, 195–242.

Box, G.E.P., Hunter, J.S., (1961a), "The 2^{k-p} fractional factorial designs, I", *Technometrics*, 3, 311–351.

Box, G.E.P., Hunter, J.S., (1961b), "The 2^{k-p} fractional factorial designs, II", *Technometrics*, 3, 449–458.

Box, G.E.P., Jones, S.P., (1989), "Designs for minimizing the effect of environmental variables", *33rd Annual Fall Technical Conference of the ASQC*, Houston, Texas.

Box, G.E.P., Wilson, K.B., (1951), "On the Experimental Attainment of Optimum Conditions", *Journal of the Royal Statistical Society*, B13, 1–45.

Box, G.E.P., Youle, P.V., (1955), "The exploration and exploitation of response surfaces: an example of the link between the fitted surface and the basic mechanism of the system", *Biometrics*, 11, 287–323.

Box, M.J., Draper, N.R., (1971), "Factorial Designs, the |X'X| Criterion and Some Related Matters", *Technometrics*, 13, 731–742.

Box, M.J., Draper, N.R., (1974), "On Minimum-Point Second-Order Designs", *Technometrics*, 16, 613–616.

Bradley, R.A., (1958), "Determination of optimum operating conditions by experimental methods, Part I: Mathematics and statistics fundamental to the fitting of response surfaces", *Indus. Qual. Control.*, 15(4), 16–20.

Cook, R.D., Nachtsheim, C.J., (1980), "A comparison of algorithms for constructing D-optimal designs", *Technometrics*, 22, 315–324.

Coulson, J., Southall, R., (1976), "Diesel Engine Noise — Basic Studies Lead to Practical Reductions", *SAE 760550.*

Crowther, E.M., Yates, F., (1941), "Fertiliser policy in war-time", *Empire J. Exp. Agric.*, 9, 77–97.

Daniel, C., (1959), "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments", *Technometrics*, 1, 311–341.

Daniel, C., (1976), *Applications of Statistics to Industrial Experimentation*, Wiley, New York.

Daniel, C., Wood, F.S., (1971), *Fitting Equations to Data*, Wiley, New York.

Davies, O.L., ed., (1954), *Design and Analysis of Industrial Experiments*, Oliver and Boyd, Edinburgh.

Davies, O.L., Hay, W.A., (1950), "The construction and uses of fractional factorial designs in industrial research", *Biometrics*, 6, 233–249.

DeBaun, R.M., (1959), "Response surface designs for three factors at three levels", *Technometrics*, 1, 1–8.

DeJong, R.G., (1976), *Vibrational Energy Transfer in a Diesel Engine*, Sc.D. Thesis, Massachusetts Institute of Technology.

DeJong, R.G., Manning, J.E., (1979), "Modelling of vibration transmission in engines to achieve noise reduction", *SAE 790360.*

Derringer, G.C., Suich, R., (1980), "Simultaneous optimization of several response variables", *J. Qual. Technol.*, 12, 214–219.

Doehlert, D.H., (1970), "Uniform Shell Designs", *Journal of the Royal Statistical Society*, C19, 231–239.

Doehlert, D.H., Klee, V.L., (1972), "Experimental Designs Through Level Reduction of the d–Dimensional Cuboctahedron", Discrete Mathematics, 2, 309–334.

Draper, N.R., (1960a), "Second order rotatable designs in four or more dimensions", *Ann. Math. Statist.*, 31, 23–33.

Draper, N.R., (1960b), "Third order rotatable designs in three dimensions", *Ann. Math. Statist.*, 31, 865–874.

Draper, N.R., (1960c), "A third order rotatable designs in four dimensions", *Ann. Math. Statist.*, 31, 875–877.

Draper, N.R., (1961), "Third order rotatable designs in three dimensions: some specific designs", *Ann. Math. Statist.*, 32, 910–913.

Draper, N.R., (1962), "Third order designs in three factors: analysis", *Technometrics*, 4, 219–234.

Draper, N.R., (1963), "'Ridge analysis' of response surfaces", *Technometrics*, 5, 469–479.

Draper, N.R., (1982), "Center Points in Second-Order Response Surface Designs", *Technometrics*, 24, 127–133.

Draper, N.R., (1985), "Small composite designs", *Technometrics*, 27, 173–180.

Draper, N.R., Lin, D.K.J., (1990a), "Small Response Surface Designs", *Technometrics*, 32, 187–194.

Draper, N.R., Lin, D.K.J., (1990b), "Connection Between Two-Level Designs of Resolutions III* and V", *Technometrics*, 32, 283–288.

Draper, N.R., Pukelsheim, F., (1990), "Another Look at Rotatability", *Technometrics*, 32, 195–202.

Federov, V.V., (1972), *The Theory of Optimal Experiments*, Academic Press, New York.

Finney, D.J., (1945), "The fractional replication of factorial arrangements", *Ann. Eugen.*, 12, 291–301.

Galil, Z., Kiefer, J., (1980), "Time and space saving computer methods, related to Mitchell's DETMAX for finding D-optimum designs", *Technometrics*, 22, 302–313.

Gardiner, D.A., Grandage, A.H.E., Hader, R.J., (1959), "Third order rotatable designs for exploring response surfaces", *Ann. Math. Statist.*, 30, 1082–1096.

Hall, R.A., (1992), *OPTRSM (optimization and response surface methods): Introductory User Guide*. Department of Transport Technology Publication, Loughborough University of Technology.

Hall, R.A., Zhang, T., (1992), *Noise Analysis Program: Introductory User Guide*. Department of Transport Technology Publication, Loughborough University of Technology.

Harrington, E.C., (1965), "The desirability function", *Indus. Qual. Control*, 21, 494–498.

Hartley, H.O., (1959), "Smallest Composite Designs for Quadratic Response Surfaces", *Biometrics*, 15, 611–624.

Herzberg, A.M., (1964), "Two third order rotatable designs in four dimensions", *Ann. Math. Statist.*, 35, 445–446.

Hill, W.J., Hunter, W.G., (1966), "A Review of Response Surface Methodology: A Literature Survey", *Technometrics*, 8, 571–590.

Hoerl, A.E., (1959), "Optimum solution of many variables equations", *Chem. Eng. Prog.*, 55, 69–78.

Hoke, A.T., (1974), "Economical Second-Order Designs Based on Irregular Fractions of the 3^n Factorial", *Technometrics*, 16, 375–384.

Huda, S., (1987), "The construction of third order rotatable designs in k dimensions from those in lower dimensions", *Pakistan Journal of Statistics*, 3(1), 11–16.

Huda, S., Shafiq, M., (1987), "On D_s-efficiency of D-optimal fourth order rotatable designs", *Pakistan Journal of Statistics*, 3(3), 33–37.

Hunter, J.S., (1958), "Determination of optimum operating conditions by experimental methods, Part II-1: Models and Methods", *Indus. Qual. Control.*, 15(6), 16–24.

Hunter, J.S., (1959a), "Determination of optimum operating conditions by experimental methods, Part II-2: Models and Methods", *Indus. Qual. Control.*, 15(7), 7–15.

Hunter, J.S., (1959b), "Determination of optimum operating conditions by experimental methods, Part II-3: Models and Methods", *Indus. Qual. Control.*, 15(8), 6–14.

Iman, R.L., Helton, J.C., (1988), "An investigation of uncertainty and sensitivity analysis techniques for computer models", *Risk Analysis*, 8, 71–90.

Kanemasu, H., (1979), "A Statistical Approach to Efficient Use and Analysis of Simulation Models", *Bulletin of the International Statistics Institute*, 48, 573–604.

Khuri, A.I., (1985), "A Measure of Rotatability for Response Surface Designs", Technical Report No. 232, Dept. of Statistics, Univ. of Florida, Gainesville, FL 32611.

Khuri, A.I., Conlon, M., (1981), "Simultaneous optimization of multiple responses represented by polynomial regression functions", *Technometrics*, 23, 363–375.

Khuri, A. I., Cornell, J. A., (1987), *Response Surfaces: Designs and Analyses* , Dekker, New York.

Kiefer, J., (1958), "On the nonrandomized optimality and the randomized nonoptimality of symmetrical designs", *Ann. Math. Statist.*, 29, 675–699.

Kiefer, J., (1959), "Optimal experimental designs", *Journal of the Royal Statistical Society*, B21, 272–304.

Kiefer, J., (1960), "Optimal experimental designs II, with applications to systematic and rotatable designs", *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, 1, 381–405.

Kiefer, J., (1962a), "Two more criteria equivalent to D-optimality of designs", *Ann. Math. Statist.*, 29, 792–796.

Kiefer, J., (1962b), "An extremum result", *Canad. J. Math.*, 14, 597–601.

Kiefer, J., Wolfowitz, J., (1959), "Optimum designs in regression problems", *Ann. Math. Statist.*, 30, 271–294.

Kiefer, J., Wolfowitz, J., (1960), "The equivalence of two extremum problems", *Canad. J. Math.*, 12, 363–366.

Koshal, R.S., (1933), "Application of the Method of Maximum Likelihood to the Improvement of Curves Fitted by the Method of Moments", *Journal of the Royal Statistical Society*, A96, 303–313.

Lalor, N., (1979), "Computer optimised design of engine structures for low noise", *SAE 790364*.

Lalor, N., Petyt, M., (1982), "Noise Assessment of Engine Structure Designs by Finite Element Techniques", *Engine noise: Excitation, vibration and radiation*, 211–244, New York: Plenum Press.

Lucas, J.M., (1974), "Optimum Composite Designs", *Technometrics*, 16, 561–567.

Lucas, J.M., (1976), "Which Response Surface Design is Best", *Technometrics*, 18, 411–417.

Lucas, J.M., (1977), "Design efficiencies for varying numbers of centre points", *Biometrika*, 64, 145–147.

Martin, I.T., Law, B., (1989), "Prediction of Crankshaft and Flywheel Dynamics", *EAEC 2nd Conference on Powertrain and Chassis Engineering*, Strasbourg, Paper C382/046, 487–498.

Matheron, G., (1963), "Principles of Geostatics", *Economic Geology*, 58, 1246–1266.

McKay, M.D., Conover, W.J., Beckman, R.J., (1979), "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code", *Technometrics*, 21, 239–245.

Mead, R., Pike, D.J., (1975), "A Review of Response Surface Methodology from a Biometric Viewpoint", *Biometrics*, 31, 803–851.

Milsted, M.G., Zhang, T., Hall, R.A., (1990), "A numerical procedure for engine noise optimization". *VDI-Gesellschaft Fahrzeugtechnik. 5th International Congress: Numerical analysis in automotive engineering.*Würzburg, Germany.

Milsted, M.G., Zhang, T., Hall, R.A., (1991), "A numerical method for noise optimization of engine structures". *Second international conference: Computers in engine technology*, Cambridge, U.K..

Milsted, M.G., Zhang, T., Hall, R.A., (1993), "A numerical method for noise optimization of engine structures". *Proc. I.Mech.E.*, (D) 207, 135–143.

Mitchell, T.J., (1974a), "An Algorithm for the Construction of "D-Optimal" Experimental Designs", *Technometrics*, 16, 203–210.

Mitchell, T.J., (1974b), "Computer Construction of "D-Optimal" First-Order Designs", *Technometrics*, 16, 211–220.

Mitscherlich, E.A., (1930), *Die Bestimmung des Dungerbedurfnisses des Bodens*, Paul Parey, Berlin.

Montgomery, D.C., (1976), *Design and Analysis of Experiments*, Second Edition, Wiley, New York.

Morris, M.D., (1987), "Two-stage factor screening procedures using multiple grouping assignments", *Communications in Statistics — Theory and Methods*, A16, 3051–3067.

Myers, R.H., (1971), *Response Surface Methodology* , Allyn and Bacon, Boston.

Myers, R.H., Carter, W.H., Jr., (1989), "Response surface techniques for dual response systems", *Technometrics*, 15, 301–317.

Myers, R.H., Khuri, A.I., Carter, W.H., Jr., (1989), "Response surface methodology: 1966–1988", *Technometrics*, 31, 137–157.

NAG FORTRAN Library Manual Mark 11 (1983), Oxford : NAG Central Office.

National Bureau of Standards, (1957), "Fractional Factorial Experiment Designs For Factors at Two Levels", National Bureau of Standards, Applied Mathematics Series, No. 48, Washington, D.C..

National Bureau of Standards, (1959), "Fractional Factorial Experiment Designs For Factors at Three Levels", National Bureau of Standards, Applied Mathematics Series, No. 54, Washington, D.C..

Notz, W., (1982), "Minimal Point Second Order Designs", *Journal of Statistical Planning and Inference*, 6, 47–58.

Ogendo, J.E.W., Zhang, T., (1989) *A numerical procedure for engine noise optimization.* Department of Transport Technology Publication TT 89 R 09, Loughborough University of Technology.

Ott, W., Kaiser, H.-J., Meyer, J., (1990), "Finite Element Analysis of the Dynamic Behaviour of an Engine Block and Comparison with Modal Test Results", *MSC World Users Conference*, Los Angeles.

Pesotchinsky, L., (1975), "D-optimum and quasi-D-optimum second-order designs on a cube", *Biometrika*, 62, 335–340.

Plackett, R.L., Burman, J.P., (1946), "The design of optimum multifactorial experiments", *Biometrika*, 33, 305–325.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., (1986), *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press.

Priede, T., (1980), "In search of origins of engine noise - an historic review", *SAE 800534*, 33, 305–325.

Rechtschaffner, R., (1967), "Saturated Fractions of 2^n and 3^n Factorial designs", *Technometrics*, 9, 569–575.

Roquemore, K.G., (1976), "Hybrid Designs for Quadratic Response Surfaces", *Technometrics*, 18, 419–423.

Sacks, J., Schiller, S.B., Welch, W.J., (1989), "Designs for Computer Experiments", *Technometrics*, 31, 41–47.

Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., (1989), "Design and Analysis of Computer Experiments", *Statistical Science*, 4, 409–435.

Sehmi, N.S., (1985), "A Newtonian procedure for the solution of the Kron characteristic value problem", *Journal of Sound and Vibration*, 100, 409–421.

Sehmi, N.S., (1986), "The Lanczos algorithm applied to Kron's method", *International Journal for Numerical Methods in Engineering*, 23, 1857–1872.

Sehmi, N.S., (1989), *Large Order Structural Eigenanalysis Techniques*, Ellis Horwood, Chichester.

Silvey, S.D., (1980), *Optimal Design*, Chapman and Hall, London.

Somkhanay, N., (1982), *Computer Optimised Design of Engine Structures for Low Noise*, Ph.D. thesis, University of Southampton.

Srivastava, J.N., (1975), "Designs for searching non-negligible effects", in *A Survey of Statistical Designs and Linear Models*, J.N. Srivastava, ed., North Holland, Amsterdam, 507–519.

Turner, G.L., (1983), *Finite Element Modelling and Dynamic Substructuring for Prediction of Diesel Engine Vibration*, Ph.D. thesis, Loughborough University of Technology.

Turner, G.L., Milsted, M.G., Hanks, P., (1984), "Vibration Characteristics of an In-line Engine Structure", *Proc.I.Mech.E., C138/84*, 161–171.

Turner, G.L., Milsted, M.G., Hanks, P., (1986), "The Adaptation of Kron's Method for Use With Large Finite-Element Models", *Trans. ASME. Vol. 108, Journal of Vibration, Acoustics, Stress and Reliability in Design*, 405–410.

Vanderplaats, G.N., (1984), *Numerical Optimization Techniques for Engineering Design, with Applications*, McGraw-Hill, New York.

Vanderplaats, G.N., Blakely, K.D., (1989), "The Best and The Lightest", *Mechanical Engineering*, 111(2), 56–62.

Wardrop, D.M., Myers, R.H., (1990), "Some Response Surface Designs for Finding Optimal Conditions", *Journal of Statistical Planning and Inference*, 25, 7–28.

Watson, G.S., (1961), "A Study of the Group Screening Method", *Technometrics*, 3, 371–388.

Welch, W.J., Buck, J.S., Sacks, J., Wynn, H.P., Mitchell, T.J., Morris, M.D., (1989), *Screening a Large-Dimensional Input to a Computer Experiment*, Technical Report #31, Department of Statistics, University of Illinois.

Westlake, W.J., (1965), "Composite designs based on irregular fractions of factorials", *Biometrics*, 21, 324–336.

Wilcox, C., (1988), *Shape Optimisation of Engine Structures*, Ph.D. thesis, University of Southampton.

Winsor, C.P., (1932), "The Gompertz curve as a growth curve", *Proc. Natl. Acad. Sci.*, 18, 1–8.

Wishart, J., (1938), "Growth rate determination in nutrition studies with the bacon pig, and their analysis", *Biometrika*, 30, 16–28.

Wishart, J., (1939), "Statistical treatment of animal experiments", *Journal of the Royal Statistical Society*, B6, 1–22.

Wynn, H.P., (1970), "The sequential generation of D-optimal designs", *Ann. Math. Statist.*, 41, 1655–1664.

Yates, F., (1935), "Complex experiments", *Journal of the Royal Statistical Society*, B2, 181–223.

Zelen, M., Connor, W.S., (1959), "Multi-Factor Experiments", *Industrial Quality Control*, 15(9), 14–17.

Zhang, T., (1992), *Structural Optimization of Reciprocating Engines for Minimum Noise Radiation*, Ph.D. thesis, Loughborough University of Technology.

Ziegel, E.R., Gorman, J.W., (1980), "Kinetic modeling with multiresponse data", *Technometrics*, 22, 139–151.