Computer Aided Design of Experiments.
An Engineering Approach.

by
Luca Costantino Zullo

Thesis submitted for the degree of
Doctor of Philosophy of the University of London
and for the
Diploma of Imperial College

Imperial College of Science, Technology and Medicine
Department of Chemical Engineering and
Chemical Technology

February 1991
... The solution, it seems to me, lies in the cultivation of strong values that lie outside science. We must have more education for scientists in humanities, in history. We cannot afford to be ignorant technicians. We must have less rigidity of thought. We must avoid becoming a scientific priesthood.

... We must remember that though Mathematics may be the Queen of the Sciences, Science is not the only principle of life. “The tree of life is larger than the tree of thought,” said Immanuel Kant. I believe it.

Let us cultivate men of the mind who will also be men of the heart.

Philip J. Davis and Reuben Hersh
“Descartes’ Dream”

To my grandfather Bruno, a man of heart, in memoriam.
Building mechanistic models of reacting systems is a very important activity in chemical engineering research. This activity often involves an iterative process involving several steps such as experimental runs, parameter estimation, model adequacy tests and etc. Despite the importance of such an activity and the richness of the theoretical debate, there is still a lack of reliable and widely disseminated practical tools to help the experimenter during this process. In this work, an analysis of computational methods for the design of experiments is carried out and in assessing the results, it has been adopted the point of view of a chemical engineer developing a kinetic model, rather than that of a statistician.

In particular, the design of experiments to improve parameter estimates is examined. Although most of the theory used in this study dates back to the mid fifties, nonlinear experimental design poses numerical problems which only recently have been efficiently tackled. A computer code implementing a novel and efficient approach for the sequential design of experiments for models described by algebraic equations has been developed and tested. A series of numerical experiments has been carried out. The results show comparable or slightly better parameter estimates than those reported in the literature are obtained with the designed experiments. However, these have been obtained with an easy to use general purpose code rather than with some problem specific procedure.

The application of such design techniques to dynamic models has been examined. Dynamic models are important because many processes of industrial interest are not carried out at steady state. Moreover, dynamic experiments may give an intrinsically higher content of information. An algorithm enabling one to independently select optimal sampling points, final time, external controls and initial conditions of a dynamic experiment has been developed. Numerical problems in the computational sensitivity analysis of stiff systems of mixed algebraic and differential equations, have so far precluded an extensive testing. Despite such problems, the case studies carried out show promising results. A detailed analysis of the currently available numerical methods for sensitivity analysis is reported and available computer codes have been tested and compared.

The research on the dynamic systems has shown the inadequacy of the present algorithms for dynamic parameter estimation. A theoretical analysis of this problem is carried out and a novel algorithm is proposed. Moreover the use of dynamic sensitivity analysis and parametric identifiability techniques for the preliminary screening of a suggested model are investigated.
There are many people to whom I am deeply in debt and whose influence I felt during the years in which my research project has been carried out. Not all of them had a direct impact of the scientific substance of this work, however there is more to a Ph.D. project than what eventually goes into the thesis.

I am grateful to my supervisor, Sandro Macchietto, who put up with my frequent decisions to give up and always persuaded me that my work had some significance.

I am also grateful to Montedison SpA, their financial support, in the form of a scholarship, made this project and my stay in London possible. During my stay at the Montedipe research site in Mantova (Italy), I have always enjoyed an extremely friendly and helpful environment. In particular, my thanks are for Anna Rossi, Leonardo Castellani, Aldo Longo, Anna Maria Marconi, Edgardo Malaguti and the director of the centre Claudio Bonerba. I sadly regret that the work carried out there did not find adequate description in this thesis.

I can't express all my appreciation to “Mr.” Chen, who introduced me to the secrets of numerical optimisation. His suggestions have been of foremost importance for the development of the algorithms presented in this thesis. Many thanks also to Richard Jarvis for making his code available and for prompt answers to many questions on sensitivity analysis.

I want to remember and thank all the persons who made my years at Imperial a mostly enjoyable time in my life. Yishu Nanda, Symeon Kassianides, Joao Keller and Alice Vrielink, have been good friends of mine. Without them I would have probably finished one year before, but my life would have been rather miserable.

My thanks also to some outstanding members of the Greek community: Kostas Pipilis, for many pleasant discussions about life, universe and everything, Kostas Krallis and Vassili Papadopulos for being my personal UNIX gurus, Vassili Kouyonas for prompt support and the common enthusiasm in Italian sport cars and Elena Kyriakis for giving me shelter during my last weeks in London.

There are however persons for whom I feel it impossible to adequately express my gratefulness. My parents, my grandmother and my aunt Lella supported me in any possible manner, financially, morally and psychologically. The last, but not the least, my fiancee' Val Blanchard. If this thesis is written in good English, she deserves the praise. However, my legacy to her goes well beyond that and, without her, I would never have finished my Ph.D. This thesis and this degree are also hers.
### Abstract

2

### Acknowledgements

3

### Table of Contents

4

### CHAPTER 1

Introduction and background.

1.1 Sequential experimental design. ........................................... 10

1.1.1 Design to discriminate among models. .......................... 12

1.1.2 Design to improve parameter estimates. ..................... 14

1.2 Application of sequential design in chemical engineering. ....... 15

1.3 Aims and description of the project. ............................... 18

### CHAPTER 2

Sequential design of experiments to improve parameter estimates in non linear algebraic models. A computationally efficient approach.

2.1 Bayesian approach to the sequential design of experiments. ..... 21

2.2 Extension to multivariate models. ................................. 23

2.3 Design criteria. ...................................................... 25

2.4 Computer codes for sequential design of experiments. .......... 28

2.5 Numerical aspects of the sequential design of experiments. .... 30

2.5.1 Evaluation of the objective function. .......................... 30

2.5.2 Evaluation of the gradients. ................................. 31

2.6 Description of PRODES. A computer code for the
sequential design of nonlinear experiments. 33

2.6.1 Monovariate models. ........................................... 33
2.6.2 Multivariate models. .......................................... 34
2.6.3 General remarks. .............................................. 35

2.7 Case study: nitric oxide reduction. .......................... 35
2.7.1 Original model. .................................................. 36
2.7.2 Transformed model: Ratkowski's transformation. ...... 46
2.7.3 Transformed model: Lumping of multiplicative parameters. 56

2.8 Case study: Ethylene hydrogenation. .......................... 61

2.9 Case study: multivariate model. .............................. 66
2.9.1 Nonlinearly constrained experimental space ........... 69

2.10 Comments on the results and conclusion. .................. 70

CHAPTER 3
Optimal design of experiments with dynamic systems. A deterministic approach. 74

3.1 Statement of the problem. ...................................... 75
3.2 The optimal control approach. ............................... 77
3.3 A new formulation of the design of dynamic experiments. 81
3.4 An algorithm for the design of dynamic experiments. .... 85
3.5 Case study: a simple kinetic model. ......................... 87
3.5.1 Use of initial conditions and constant inputs as design variables. 95
3.6 Case study: optimal selection of sampling points in a fed batch fermentation reactor. 98
3.7 Conclusion and discussion. ................................... 101

CHAPTER 4
Sensitivity analysis of dynamic systems. Computational methods and use as an aid for model building. 104

4.1 Sensitivity analysis of dynamic systems: computational methods. 104
4.1.1 Comparison of sensitivity analysis codes. ............... 108
4.1.2 Comments on sensitivity analysis. ........................................... 114

4.2 The use of sensitivity analysis as

an aid for model building. ......................................................... 115

4.2.1 Structural identifiability analysis. ........................................ 115

4.2.2 The power expansion method for identifiability analysis. ....... 116

4.2.3 Example: preliminary investigation of a complex kinetic model. 118

4.2.4 Comments ................................................................. 124

CHAPTER 5

Discussion of results and future work. ........................................ 126

5.1 Practical problems in dynamic multivariate parameter estimation. . 126

5.2 Multiresponse parameter estimation in dynamic models:

an application of the Box-Draper criterion. ............................... 131

5.2.1 A new algorithm for parameter estimation

in nonlinear dynamic models. ................................................. 134

5.3 Critical remarks on the design of experiments to improve

parameter estimates. ................................................................ 136

5.3.1 Future work on sequential design of dynamic experiments. .... 138

5.4 Future areas of work in experimental design

and mechanistic model building. .............................................. 138

References ................................................................. 141

Appendices

A – On the use of the optimal control approach to

design experiments for model discrimination. ............................ 150

B – The need for an objective function analytical in

state variables in the optimal control problems. ....................... 151

C – Calling statement and user instruction of the

routine PRODES. .................................................................. 153

D - Note on the numerical derivatives scheme used for PRODES........... 158

E - Further notes on the sensitivity analysis................................. 159

F - Nomenclature .................................................................... 160
Mechanistic models of processes are needed in virtually every aspect of a chemical engineer's activity. They constitute the conceptual base on which the rational pursuing of activities as different as process development, equipment design and economical optimisation of operations is built upon.

Despite such a prominent practical importance, the building and the testing of such models is often not seen as a mainstream engineering activity and it is too common that models are accepted either from literature or empirical experience without adequate assessment. The shortcomings of this approach appear when the process drifts from normal conditions or if new operating possibilities are investigated. This is particularly evident with unsteady state operations. The increased importance of dynamic process has however stimulated the need for more sophisticated modelling tools. Moreover, it is the growing concern about the safety of chemical operation which requires the development of dynamic models also for processes usually carried out at the steady state.

Model building is a typical multidisciplinary task as it requires physical understanding of the process, mastering of mathematical and computational tools and the analytical ability to correctly set the problem. Clear definition of the objectives is a very important part of a correctly tackled modelling problem. As an example, the same chemical reaction may require different modelling approaches whether the task is to optimise the yield, rather than the investigation of basic molecular phenomena.

Figure 1.1 illustrates such concepts. In a preliminary phase the objective of the modelling is defined. This should be done taking into account the allocated budget (cost of the experiments) and possible deadline for project completion (length and number of experiments). In a second step qualitative analysis on the system under investigation is carried out. Tentative sets of equations may be written. During this phase a number of important assumptions may be made. These
Real system

PRELIMINARY: Problem definition
Objective of modelling
Allocated budget
Deadlines

QUALITATIVE ANALYSIS: Conceptual model
Available input data
Assumption on the system
Assumption on the parameters

QUANTITATIVE ANALYSIS: Mathematical modelling

Operational model

Fig 1.1 Conceptual diagram of the model building procedure
concern the available input data and the possible values of the parameter of interest. Moreover, during this phase the system should be decomposed to avoid unnecessary complexity. As an example, in a study concerning the dynamic behaviour of a reactor system, kinetic modelling of the reaction and fluodynamic modelling of the reactor could be separated.

It is during the development of such a conceptual model that some preliminary experiments are carried out. These are mainly intended to tune the experimental apparatuses in order to achieve a good level of reproducibility in the experimental data and to get a "feel" for the system responses.

In the third phase the actual mathematical modelling should be carried out. This involves the proposal of one model (or of a set of possible models), the choice of the final model (if more than one have been proposed) and the final refinement of parameter estimates until the desired level of accuracy is reached. This thesis mainly deals with some of the issues involved in this third phase. Particular interest is paid to the strategies and methods for efficient design of experiments in both stationary and dynamic conditions. However, since no clear cut division is possible from other aspects of the model building procedure some attention is also paid to other areas. The intimate correlation between parameter estimation and experimental design is highlighted.

Although chemical engineering deals with a rather wide variety of models, this work is mainly targeted towards kinetic models of reacting systems. Modelling of chemical reactions is still a rather difficult activity as it requires a great deal of physical insight into the system and, when dynamic cases are considered, the final models are often stiff systems of mixed algebraic and differential equations. However, in order to successfully pursue the mathematical modelling of a reacting system, its conceptual model should be clear and well defined. The analysis of input data requires to know whether the available set of measured variables (a subset of the state variables) would enable one to uniquely estimate all the parameters present in a model. Moreover, when dealing with nonlinear models it is very important to have reasonable initial guesses for the parameter values. Both these topics are briefly tackled and possible formal methods are examined as areas of future research.

The approach used in this study is from the point of view of the user, the chemical engineer who wants to improve his/her modelling. At the present moment, it is
difficult to do so without being distracted from the main task by algorithmic details or by the theoretical complexity of the tools being used. An efficient implementation of existing methods is required. This aim, in turn, requires critical analysis of existing statistical and mathematical tools rather than development of new ones.

In the next section the idea of sequential experimental design is introduced. In sequential design of experiments, the results of a set of previous experiments are used to design the following one(s). In this area a very rich theoretical debate has developed. Object of such debate have been mainly the statistical properties of the design criteria, rather than the development of working tools for the experimenter. Therefore, unlike parameter estimation, such a wide debate has not generated an equivalently wide variety of applications by experimenters. This is also due to the lack of a library of robust computer routines, available to engineers and experimenters who, although knowledgeable, do not have statistics as their main professional interest. This work aims to contribute to fulfill such a gap.

1.1 Sequential experimental design.

Figure 1.2 illustrate the stages involved in a sequential experimentation. Two distinct areas are identified: model discrimination and precise parameter estimation. The model discrimination stage is sometimes optional as often there is only one available model. The two stages differ because different criteria and methods are used to design and assess the experiment. In both cases the results and the settings of a set of experiments supply information for the design of the following ones.

In this context, factorial design, orthogonal design and similar designs [John and Quenoville, 1977, and McLean and Anderson, 1984] do not qualify as sequential design criteria as such prior information is not taken into account. For this reason they are also referred to as static or a priori designs [Titterington and al. 1989]. Static design methods are intended to improve the fitting of empirical regression models to experimental data. In the case of the factorial design such model are also linear. Static designs achieve such a task by choosing experimental setting which minimize the correlation level of the independent variables. Static design may be useful as a starting point for a sequential procedure when only highly uncertain first guesses are available. Nonlinear sequential design criteria depend on such guesses
Discrimination parameter estimation

Fig 1.2 Conceptual diagram of the sequential procedure of experimentation
and very uncertain parameter values may give unreliable designs. An assessment of the influence of such uncertainties on the results of the design is among the objective of this study.

Static methods are easy to use and, for a limited number of experiments and/or independent variables, do not even require the use of computers. They are largely used and trusted. Any results achieved with a sequential design should be, therefore, compared with those obtained using static methods.

The previous considerations are mainly valid for steady state models only. The basic concept of sequential experimentation can be applied to dynamic systems as well. However, despite the importance of dynamic systems, the design of dynamic experiments to assess mechanistic models is a rather new field in which very few definite results are available.

1.1.1 Design to discriminate among models.

Discrimination among alternative models is usually carried out by means of tests such as Barlett’s $\chi^2$ which measure the lack of fit [Himmelblau, 1970]. Acceptable values of such measure are tabulated as a function of the probability level and the numbers of degrees of freedom of the estimation problem.

Design criteria for model discrimination tend to find settings of the independent variables such that any possible lack of fit or discrepancy among competitive models is enhanced. Hunter and Reiner [1965] proposed to find settings of the independent variables such that the distance between the predictions of the competing models is maximized. For two monovariate models this reduces to:

$$\max [y_1 - y_2]^2$$ (1.1)

where $y_1$ and $y_2$ are the two responses. Criterion 1.1 can be easily extended to multivariate models and to cases where more than two models are considered.

Box and Hill [1967] objected that criterion 1.1 does not take into account the uncertainties in model predictions associated with inaccurate parameter values rather than with true inadequacy of the model. They proposed a model which is function also of the variance of the observations. This criterion is however of more difficult use as it requires information on the prior probability as well.

This problem has lead to the proposal of modified criteria which, although retaining information on the variance of the model’s predictions, may enable an easier
handling of the prior probability functions [Hosten and Froment, 1976]. Analogous criteria were proposed by Buzzi-Ferraris and Forzatti [1983] with subsequent extensions to the multivariate case [Buzzi-Ferraris and al., 1984].

The extension of the method to dynamic models has been carried out by Espie [1986] and Espie and Macchietto [1989]. The criterion chosen was the original Hunter-Reiner criterion because it is the only one compatible with the optimal control approach used (Appendix A). Despite the theoretical limitation of the design criterion such an approach performed satisfactorily.

From a practical point of view the Hunter-Reiner criterion, thanks to its simplicity, is the most used one. Moreover, Atkinson [1981] has shown there are conditions, common in most of the practical cases, where the Hunter-Reiner criterion is fully equivalent to the Box-Hill. Moreover, the same author shows that real world examples allow such an equivalence to be stretched even beyond its theoretical limits.

All the above criteria carry out some sort of comparison between the predicted responses. In the multivariate case, this would require the model having the same number of responses, otherwise, any comparison of the measured lack of fit is scarcely significative. Moreover, for obvious reasons, the independent variables need to be the same for all the models. In other terms the competitive models need to be structurally equivalent. This happens when, say, modelling a catalytic reaction, two different Hougen-Watson isotherms are proposed. In dynamic studies, however, very often alternative models contemplate different sets of state variables and different sets of independent variables. This makes the use of any of these criteria unreliable. If the dynamic models considered are structurally equivalent the optimal control approach used by Espie [1986] is still the most appealing method proposed so far.

In the following of this work the design for model discrimination is not considered in any further detail, as this problem, although important, is considered to be less relevant than the problem of improving parameter estimates. In many complex real world cases, only one approximate model can be formulated and, when alternatives are present, they are often not structurally equivalent.
1.1.2 Design to improve parameter estimates.

The theoretical aspects of design to improve parameter estimates are considered in some detail in the following chapter 2. The objective of this section is to introduce some general concepts.

Given the results and the setting of a number of experiments, it is necessary to define some measure of the statistical level of information associated with a new set of subsequent experiments. Probabilistic considerations allow one to define the information matrix $M$ as a function of the statistical significance associated with a certain set of experiments. However, a matrix is an entity which can be measured in different ways. Depending on which property of the matrix (determinant, condition number, etc.) one accepts as measurement, different design criteria are postulated. The information matrix is usually a function of the experimental settings (of both past and future experiments) and, for non linear models only, of the current value of the parameter estimates.

Sequential experimental design defines a non linear optimisation problem, regardless of the model being linear. Efficient implementation of a criterion and the type of the criterion are not separate issues. Some criteria, no matter if statistically sound, lead to objective functions incompatible with efficient optimisation algorithms. In nonlinear optimisation problems the solution is affected by the initial guesses of the design variables. Moreover, in sequential design, the shape itself of the objective function depends on the values of the parameter estimates. The influence of both initial guess of the design variables and uncertainty on parameter estimates is investigated in the following of this work.

Titterington and al. [1989] introduced the concept of fully sequential design and batch sequential design. In the former, experiments are designed one at a time, while in the latter batches of several experiments are considered. Batch sequential design is quite appealing as it would enable one to reduce the number of iterations needed before reaching satisfactory results. In this work the effect of the number of experiments in a single batch are investigated in order to establish whether an optimal size exists for the batch. (i.e the number of experiments in a single batch.)

It has been noted that parameter estimation techniques tend asymptotically toward the true value of parameter estimates as the number of samples increases. It should be verified if the sequential design has any asymptotic property. In this
case, there is a point after which the effect of the design, in terms of improvement of the parameter estimates, cannot be discriminated from those effects due to the mere addition of extra samples. The characterisation of such a behaviour supplies a termination criterion for the sequential procedure, which is irrespective of the quality of the parameters.

Table 1.1 summarises some of the practical problems to be investigated. The models used in sequential design are considered to be correct, compatibly with the assumptions made during the development of the conceptual model.

<table>
<thead>
<tr>
<th></th>
<th>Adequate number of experiments in a single batch</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Influence of initial guesses and uncertain parameter estimates on the design</td>
</tr>
<tr>
<td>3</td>
<td>Termination criteria for the sequential procedure</td>
</tr>
</tbody>
</table>

Table 1.1. — Practical problems in nonlinear sequential experimental design.

The topics of table 1.1 will be investigated, mainly through the study of selected case studies, in the following of this work.

The issue of robust sequential design [Steinberg and Hunter, 1984] is not considered in this work. Robust design is a design intended to accommodate modelling errors (lack of fit) and intervals of confidence of the parameter estimates. Although conceptually interesting, robust design can be applied, so far, only to linear models.

1.2 Application of sequential design in chemical engineering.

The experiment design literature shows that mostly kinetic problems have been tackled. All of these works, but a few remarkable exceptions, come from researchers with a strong background and main research interest in statistics rather than in chemical engineering.
<table>
<thead>
<tr>
<th>Model</th>
<th>System</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1 [1 - \exp(\theta_2 x)]$</td>
<td>Monomolecular decay law</td>
<td>Box and Lucas [1959]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Behnken [1964]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hohman and Jung [1975]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Katz and al. [1981]</td>
</tr>
<tr>
<td>$\theta_1 \exp(\theta_2 x)$</td>
<td>Growth or decay law</td>
<td>Box and Lucas [1959]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jeanrich [1969]</td>
</tr>
<tr>
<td>$\exp[-x_1 \exp(\theta_1 - \theta_2 x_2)]$</td>
<td>Reaction A $\rightarrow$ B</td>
<td>Hill and al. [1968]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A measured</td>
</tr>
<tr>
<td>$[1 + \alpha x_1 \exp(\theta_1 - \theta_2 x_2)]^1/n$</td>
<td>Reaction A $\rightarrow$ B</td>
<td>Hill and al. [1968]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A measured</td>
</tr>
<tr>
<td>$\exp{-\theta_1 x_1 \exp[\theta_2 (x_2^{-1} - x_0^{-1})]}$</td>
<td>First order decay law</td>
<td>Box [1968]</td>
</tr>
<tr>
<td>$\theta_1 \exp(\theta_2 x) - \exp(-\theta_1 x)/ (\theta_1 - \theta_2)$</td>
<td>Reaction A $\rightarrow$ B $\rightarrow$ C</td>
<td>Atkinson [1966]</td>
</tr>
<tr>
<td></td>
<td>B measured, x time</td>
<td>Box and Lucas [1959]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Box and Hunter [1965]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Draper and Hunter [1966],[1967a,b]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hunter and al. [1969]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hill and Hunter [1974]</td>
</tr>
<tr>
<td>$1 - \theta_2 [\exp(\theta_1 x) - \theta_1 \exp(-\theta_2 x) / (\theta_2 - \theta_1)]$</td>
<td>Reaction A $\rightarrow$ B $\rightarrow$ C</td>
<td>Draper and Hunter [1966]</td>
</tr>
<tr>
<td></td>
<td>C measured</td>
<td>Katz and al. [1981]</td>
</tr>
<tr>
<td>$\theta_1 x_1 / [1 + \theta_1 x_1 + \theta_2 x_2]$</td>
<td>Reaction A $\rightarrow$ P_1 + P_2</td>
<td>Draper and Hunter [1966]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fedorov [1972]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Box [1968]</td>
</tr>
<tr>
<td>$\frac{\theta_1 x_x (1 - x)}{(\theta_2 + x)(\theta_2 - x)}$</td>
<td>Reaction A $\rightarrow$ B $\rightarrow$ C $\rightarrow$ D</td>
<td>Hill and Hunter [1974]</td>
</tr>
<tr>
<td></td>
<td>B measured</td>
<td></td>
</tr>
<tr>
<td>$\theta_1 x_2 [(1 - x) / [1 + (\theta_2 - 1)x]]^{-1}$</td>
<td>Brunauer-Emmett-Teller isotherm</td>
<td>Hanson and Hunter [1966]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hill and Hunter [1974]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kuhn [1984]</td>
</tr>
<tr>
<td>$\frac{\theta_1 \exp(-\theta_2 x_3) \theta_2 x_3 (1.632)}{1 + \theta_1 x_1 + \theta_3 x_2 + \theta_4 x_3}$</td>
<td>Isomerisation of n-pentane</td>
<td>Pritchard and Bacon [1977]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pritchard and al. [1977]</td>
</tr>
<tr>
<td>$\frac{\theta_1 \exp(-\theta_2 x_3) \theta_2 x_3 (1.632)}{1 + \theta_1 x_1 + \theta_3 x_2 + \theta_4 x_3}$</td>
<td>Oxidation of Benzene</td>
<td>Pritchard and Bacon [1977]</td>
</tr>
<tr>
<td>$x(\theta_1 x + 1)/ (\theta_2 + x)$</td>
<td>Copolymer reactivity ratio</td>
<td>Behnken [1974]</td>
</tr>
<tr>
<td>$\theta_1 x / (\theta_2 + x)$</td>
<td>Michoels Menten rate equation</td>
<td>Currie [1982]</td>
</tr>
</tbody>
</table>

Table 1.2. — Examples used to study sequential design from chemical kinetics and related fields (from Titterington and al. [1989]).
Hahn [1984] reviews, from a general standpoint, the problem of real word experimental design with particular attention to the integration with the other stages of the model building procedure. However, the author considers the situation where experimenters are backed by professional statisticians who are in charge of experimental design and data analysis. This situation, although recommendable, is not very common.

Review of specific interest for chemical engineering applications are given by Froment [1975] and Reilly and Blau [1974]. In both cases the problems of model discrimination and precise parameter estimation are considered. The first paper is particularly concerned with application to steady state modelling of heterogeneous catalytic reactions and the part on model discrimination receives a much higher attention. More recently, in a review paper Rippin [1988] considered more general type of models and introduced the idea of keeping explicitly into account the cost of the experiment.

Reilly and al. [1977] used sequential design to improve the parameter estimates of a first order kinetic model describing the uptake of pollutants in biorganisms. A vast number of kinetic problems have been used for theoretical study on design methods. These studies date back to the pioneering work by Box and Lucas [1959] who laid the basics for much of the following developments. Table 1.2 reports a list of some of these works.

Sutton and McGregor [1977] used sequential design to determine activity coefficients in vapour liquid equilibrium experiments. Sharma and Srivasta [1981 and 1982] used optimal design techniques to study the oxidation of various organic compounds over vanadium pentoxide catalyst. A similar study was carried out by Juusola and al. [1982] who used sequential design to model the catalytic oxidation of ortho-xylene.

Rippin and al. [1980] used nonlinear experimental design with an approximate model to study a complex reaction taking place in a batch reactor. The model described the economical performance of the reactor and correlated the commercial value of the batch, the cost of running the equipment (heat transfer, pumping, etc.) and the number of batches per year. Independent variables were reaction duration, temperature and initial concentration of the reagents.
Kilpinen and Westerlund [1988] used nonlinear experimental design to study the oxidation of iron ore in the sinter blast furnace. Independent variables were the stationary concentration profiles of the various species in the furnace.

1.3 Aims and description of the project.

Despite the interesting number of applications reviewed in the previous section, the use of sequential design is still, largely, a matter of specialistic interest. This is not the case for the related problem of parameter estimation. The reason for this mainly lies in the lack of a readily available and general purpose library of computer routines. At the moment any experimenter interested in the field is scared away by the time consuming task to develop his/her own specific routines. Moreover, any efficient implementation requires expertise in many areas, such as nonlinear statistics and optimisations, which are not necessarily of interest for most experimenters and engineers. Therefore, the approach followed in this work tends to stress those aspects related to the definition of general purpose tools for modeling of reacting systems. The problems of practical implementation and use of techniques for sequential design of dynamic experiments for kinetic modelling are considered with more attention than in similar works where the stress was upon the statistical properties of a selected design criterion.

An analysis of the available criteria to design experiments in order to improve parameter estimates and of their relative merits is carried out in chapter 2. Along with such analysis the literature is referenced and commented. Chapter 2 will particularly concentrate on issues such as robustness and efficiency of a general purpose computer implementation of sequential design for systems described by algebraic models. Issues such as termination criteria for the sequential procedure, influence of initial guesses and size of an experimental batch, which were pointed out in the previous section 1.1.2 (table 1.1), are dealt with and examined through the solution of a number of significative case studies.

Chapter 3 will extend the above concepts and results to the case of dynamic models described by systems of mixed algebraic and differential equations. A new algorithm for the optimal selection of sampling points and of possible external inputs
into the system is discussed and commented. Algorithmic and numerical problems are pointed out by the case studies presented.

The algorithm described in chapter 3 uses dynamic sensitivity analysis, the details of which are also examined. Chapter 4 critically examines the numerical techniques for dynamic sensitivity analysis. In particular existing computer codes are compared and their relative merits discussed. The use of sensitivity analysis as a modelling tool on its own is also discussed through an example.

The case studies of chapter 3 have shown the inadequacy of commonly used techniques to carry out efficiently parameter estimation in dynamic systems. A novel approach is proposed in chapter 5. This chapter also reports the guidelines for future work aiming to further improve the sequential design of experiments.
Sequential design of experiments to improve parameter estimates in non linear algebraic models. A computationally efficient approach.

In this chapter the theoretical background necessary to define a design criterion is outlined. In particular, the application of Bayes' theorem to the sequential design to improve the quality of parameter estimates is considered. All design criteria commonly used are offsprings of such a theorem.

Some of the most commonly proposed criteria are considered and analysed. In analysing the criteria, unlike in other studies, the statistical properties are not the only elements on which judgment is based. The problems of implementation of each criterion in an efficient and reliable computer code are also considered. This aspect is of foremost importance when the experimental design problem is tackled from a practical point of view. Very often an approach, mainly statistically oriented, has almost exclusively stressed the statistical properties of the design criteria and other problems of relevant importance for the experimenter have been neglected.

The following chapter discusses the complexity of the optimisation problems posed by the sequential design of experiments. Such complexity has often limited a more widespread use of the sequential design of experiments, no matter how theoretically appealing the criteria. Moreover, it is still object of debate whether the results of the sequential design are better than, say, those obtained by some static design, and justify the extra burden of the sequential procedure.

The objective of this chapter is to address these issues through the development of a general purpose algorithm to design sequential experiments for nonlinear monovariate and multivariate algebraic models. Attention is also paid to the implementation details that might affect the solution. The discussion of the results of the cases presented will also allow us to assess both the particular design criterion chosen and the applicability of such a general purpose code. Moreover, some of the results of this chapter are essential for the development of the following chapter, where a novel extension applicable to dynamic models is presented.
2.1 Bayesian approach to the sequential design of experiments.

Let us consider a monovariate algebraic model:

\[ g(y, x, \theta) = 0 \]  

(2.1)

where \( y \) is a single dependent variable (the response of the model), \( x \) is a \( m \) dimensional real vector of independent variables and \( \theta \) is a \( p \) dimensional real vector of invariant parameters. The vector \( \theta \) is unknown or only partially known. The model 2.1 is generally nonlinear in both the independent variables and the parameters; moreover, we assume, it has been already assessed as adequate to describe the physical process under investigation.

Given a vector of parameter estimates \( \hat{\theta}_n \) usually arising from a previous set of \( n \) experiments, the scope of the design is to determine the settings for the independent variables \( x \) in one subsequent experiment. The settings are chosen so that the statistical significance of \( \hat{\theta}_{n+1} \) is maximised. In this work, such a significance is defined by the bayesian statistic of maximum posterior probability. This applies also when the settings for a subsequent set of \( n_e \) experiments are sought. In such a case the scope is to maximise the posterior probability of \( \hat{\theta}_{n+n_e} \). Let us consider the case when only one single experiment is designed, the extension being limited to a change in the indices relative to the number of experiments which would become \( n+n_e \) rather than \( n+1 \).

When dealing with monovariate models, the terms observation and experiment are usually used interchangeably. Such habit, common in the statistical literature, is however misleading when the design for multivariate models is considered. In multivariate models, a single experiment generates several observations. To avoid confusion, in this work, the word observation is not used in the context of monovariate models.

Bayes' theorem states that the posterior probability function of \( \hat{\theta} \) given \( n+1 \) experiments is:

\[
P(\hat{\theta}|y_{n+1}) = \frac{L(\hat{\theta}|y_{n+1}) P(\hat{\theta})}{\int_{-\infty}^{+\infty} L(\hat{\theta}|y_{n+1}) P(\hat{\theta}) d\theta}
\]  

(2.2)

where \( L(\hat{\theta}|y_{n+1}) \) is the likelihood of \( \hat{\theta} \) after \( n+1 \) experiments and \( P(\hat{\theta}) \) is the prior probability function of the parameter estimates. The prior probability
P(\theta) is unknown, however the posterior probability P(\theta|y_n) can supply an estimate of it. This allows equation 2.2 to be transformed as follows:

$$P(\theta|y_{n+1}) = \frac{L(\theta|y_{n+1})}{\int_{-\infty}^{+\infty} L(\theta|y_{n+1}) P(\theta|y_n) d\theta}$$  \hspace{1cm} (2.3)

The likelihood function has the form:

$$L(\theta|y_{n+1}) = \frac{1}{\left(2\pi\right)^{n+1/2} \sigma_y} \exp\left\{ -\frac{1}{2} \left[ \frac{\sum_{i=1}^{n+1} [g(y_i, x_i, \theta)]^2}{\sigma_y^2} \right] \right\}$$  \hspace{1cm} (2.4)

where $y_i$ is the response of the $i^{th}$ experiment and $x_i$ the correspondent vector of independent variables. The posterior probability $P(\theta|y_n)$ is:

$$P(\theta|y_n) = \frac{1}{\left(2\pi\right)^{n/2} |W(\hat{\theta})|^{1/2}} \exp\left\{ -\frac{1}{2} \left[ (\theta - \hat{\theta})^T W(\hat{\theta})^{-1} (\theta - \hat{\theta}) \right] \right\}$$  \hspace{1cm} (2.5)

where $\hat{\theta}$ is the vector of parameter estimates after $n$ experiments and $W(\hat{\theta})$ is the correspondent matrix of variance-covariance.

The substitution of equations 2.4 and 2.5 in equation 2.3 gives the following expression for the posterior probability function:

$$P(\theta|y_{n+1}) = \delta_{n+1} \exp\left\{ -\frac{1}{2} \left[ (\theta - \hat{\theta})^T W(\hat{\theta})^{-1} (\theta - \hat{\theta}) \right] \right\}$$  \hspace{1cm} (2.6)

The symbols of equation 2.6 keep the previous meaning, $\delta_{n+1}$ is a normalising factor. Its significance is discussed below.

A bayesian design of the $n+1^{th}$ experiment requires the posterior probability, as expressed by equation 2.6, to be maximised. Observing equation 2.6, it can be noted that the two exponentials have already maximum values, as the arguments are proportional to maximum likelihood estimators. The only way to maximise 2.6 is to maximise the normalising factor $\delta_{n+1}$.

A suitable expression of $\delta_{n+1}$ is found by linearising the model 2.1 about the parameter estimate $\hat{\theta}$. Using a Taylor’s series expansion about the value $\hat{\theta}$ we have:

$$g(y, x, \hat{\theta}) = g(y, x, \hat{\theta}) + \sum_{j=1}^{p} \left( \hat{\theta}_j - \hat{\theta}_j \right) \frac{\partial g(y, x, \hat{\theta})}{\partial \hat{\theta}_j} + O(\hat{\theta})^2$$  \hspace{1cm} (2.7)

where $\hat{\theta}'$ is a small perturbation of $\hat{\theta}$.
A matrix of partial derivatives of the model \(2.1\) with respect to the parameters, calculated at the \(n + 1\) experimental points, is defined as:

\[
S = \{ s_{ij} \} = \left\{ \frac{\partial g(\hat{y}_i, x_i, \hat{\theta})}{\partial \hat{\theta}_j} \right\} \quad i = 1 \ldots n + 1 \quad j = 1 \ldots p
\]  

(2.8)

where \(\hat{y}_i\) is the vector of independent variables relative to the \(i^{th}\) experiments as predicted by the model. Such a matrix is called the design matrix. The linearisation 2.8 can now be introduced in equation 2.6. After some rearrangements equation 2.6 yields:

\[
\sum_{i=1}^{n+1} \{ \hat{y}_i - y_i \}^T S \{ \hat{y}_i - y_i \} = \text{min}
\]

(2.9)

\[
P(\hat{\theta}|y_{n+1}) = \delta_{n+1} \exp \left\{ -\frac{\sum_{i=1}^{n+1} \epsilon_i^T \epsilon_i}{2\sigma_y^2} \right\} \exp \left\{ -\frac{1}{2} [ (\theta - \hat{\theta})^T S T S (\theta - \hat{\theta}) + (\theta - \hat{\theta})^T \sigma_y^2 W (\hat{\theta})^{-1} (\theta - \hat{\theta}) ] \right\}
\]

where \(\epsilon\) is the vector of residuals over the \(n + 1\) experiments.

Himmelblau, [1970], shows that equation 2.9 is the combination of two multivariate normal probability densities, and therefore the following proportionality can be postulated:

\[
\delta_{n+1} \propto |S^T S + \sigma_y W (\hat{\theta})^{-1}|
\]

(2.10)

The maximisation of the posterior probability function of \(\hat{\theta}\) is achieved by maximisation of the lefthand side of equation 2.10. The argument of the determinant in equation 2.10 is a \(p \times p\) matrix defined as the information matrix. Thus maximising this determinant is equivalent to maximising \(\delta_{n+1}\) and this analysis provides the required fundamental theoretical support to more simplified optimisation criteria.

All the criteria for the sequential design of experiments derive from equation 2.10. The practical use of 2.10 for the sequential experimental design will be discussed in some detail later.

2.2 Extension to multivariate models.

Let us consider a multivariate algebraic model:

\[
f(y, x, \theta) = 0
\]

(2.11)

where \(y\) is a \(v\) dimensional vector of dependent variables, \(x\) is a \(m\) dimensional vector of independent variables and \(\theta\) is a \(p\) dimensional vector of invariant parameters.
Conceptually Bayes’ theorem 2.2 is applied exactly as explained in the previous section, however, the expressions of the likelihood and probability functions are different. These lead to an expression equivalent to, although formally rather different from equation 2.10.

In the monovariate case the information on the experimental error is supplied by the variance \( \sigma_y^2 \); in the multivariate case the variance-covariance matrix of the experimental error should be considered. Such a matrix is defined as:

\[
\Sigma = \{ \sigma_{rs} \} = \{ \text{cov}(y_r, y_s) \} \quad (2.12)
\]

The matrix \( \Sigma \) is not known, but an estimate can be computed as:

\[
\hat{\sigma}_{rs} = \frac{\sum_{i=1}^{n} (y_{ri} - \hat{y}_{ri}(\hat{\theta})) (y_{si} - \hat{y}_{si}(\hat{\theta}))}{(n-1)} \quad (2.13)
\]

where \( y_{ri} \) is the \( i \)th observation of the \( r \)th response and \( y_{ri}(\hat{\theta}) \) is the value of the same variable as calculated from the model with the current values of parameter estimates. It should be noted that such an estimate is computed over the \( n \) experiments already carried out and it is formally different from the variance-covariance matrix used, in its inverse form, in the following formulae 2.14 and 2.15. The symbol \( \sigma_{rs} \) in 2.14 and 2.15 indicates the \( rs \)th element of the inverse of the matrix of variance covariance \( \Sigma^{-1} \). The likelihood function is given by:

\[
L(\hat{\theta}|y_{n+1}) = \frac{|\Sigma^{-1}|^{(n+1)/2}}{(2\pi)^{(n+1)/2}} \exp \left\{ -\frac{1}{2} \sum_{r=1}^{v} \sum_{s=1}^{v} \sigma_{rs} \sum_{i=1}^{n+1} f_r(\hat{y}_i, x_i, \hat{\theta})^2 \right\} \quad (2.14)
\]

where \( \hat{y}_i \) is the vector of dependent variables as computed by model 2.11 with the values of the independent variables set as during the \( i \)th experiment and with the vector of parameter estimates \( \hat{\theta} \). The posterior probability function after \( n \) experiments is defined as:

\[
P(\hat{\theta}|y_n) = \frac{|\Sigma^{-1}|^{n/2}}{(2\pi)^{(n+n)v/2}|W(\hat{\theta})|^{1/2}} \exp \left\{ -\frac{1}{2} \sum_{r=1}^{v} \sum_{s=1}^{v} \sigma_{rs} \sum_{i=1}^{n} f_r(\hat{y}_i, x_i, \hat{\theta})^2 \right\} \exp \left\{ -\frac{1}{2} [ (\theta - \hat{\theta})^T W(\hat{\theta})^{-1} (\theta - \hat{\theta}) ] \right\} \quad (2.15)
\]

Equations 2.14 and 2.15 are substituted in equation 2.3 to generate a multivariate expression of the posterior probability function \( P(\hat{\theta}|y_{n+1}) \). For the sake of brevity, the detailed description of the intermediate steps is omitted, as they are conceptually equivalent to those discussed in the previous sections. It suffices to say that model 2.11 is linearised about the parameter estimates \( \hat{\theta} \) and the substitution of such a linearisation in the expression of the posterior probability function yields an expression equivalent to equation 2.9. Similarly to the previous section, it
is found that the maximisation of the probability function corresponds to the maximisation of a normalising factor $\Delta_{n+1}$, which satisfies the following proportionality:

$$\Delta_{n+1} \propto \left| \sum_{r=1}^{v} \sum_{s=1}^{y} \tilde{\sigma}_{rs}^{T} S_{r} S_{s} + W(\hat{\theta})^{-1} \right|$$  \hspace{1cm} (2.16)

where the $\tilde{\sigma}_{rs}$'s have been computed using the estimate $\Sigma$ of the matrix of variance covariance $\Sigma$ as from equation 2.13.

The matrix $S_{r}$ is the matrix of the partial derivatives of the $r^{th}$ equation in the model with respect to the parameters $\hat{\theta}$ calculated at the $n+1$ experimental points:

$$S_{r} = \{ s_{r,ij} \} = \left\{ \frac{\partial f_{r}(\tilde{y}_{i}, x_{i}, \hat{\theta})}{\partial \hat{\theta}_{j}} \right\} \quad i = 1 \ldots n + 1 \quad j = 1 \ldots p \quad r = 1 \ldots v$$ \hspace{1cm} (2.17)

The argument of the determinant in equation 2.16 defines the $p \times p$ multivariate information matrix. Maximisation of the lefthand side of equations 2.16 guarantees the maximum value of the posterior probability function for the parameter estimates $\hat{\theta}$.

### 2.3 Design criteria.

The two normalising factors 2.10 and 2.16 are both functions of the matrix of variance-covariance of the parameter estimates $W(\hat{\theta})$. When designing the $n + 1^{th}$ experiment, such a matrix is known as it was evaluated after the $n^{th}$ experiment. This value is a constant and there is no functional relationship allowing one to evaluate $W(\hat{\theta})$ as a function of the $n + 1^{th}$ experiment.

Box and Lucas [1959] had suggested that if the design of the $n + 1^{th}$ experiment is approached ignoring the prior information on $\hat{\theta}$, as when the very first experiment is designed, then the matrix $W(\hat{\theta})$ in equations 2.10 and 2.16 can be neglected. This is done by assuming that the covariance terms in $W(\hat{\theta})$ tend to infinity. The inverse matrix is therefore zero and under such an assumption equations 2.10 and 2.16 reduce to:

$$\delta_{n+1} \propto |S^{T}S|$$ \hspace{1cm} (2.18)

$$\Delta_{n+1} \propto \left| \sum_{r=1}^{v} \sum_{s=1}^{y} \tilde{\sigma}_{rs}^{T} S_{r} S_{s} \right|$$ \hspace{1cm} (2.19)
The maximisation of the determinant of the information matrices, as defined in equations 2.18 and 2.19, supplies a design criterion for the sequential design of nonlinear experiments.

This result is important not only because of the practical implications, but also because it links the nonlinear experimental design theory with the more widely assessed field of the linear experimental design. Although the previous theoretical treatment is relatively known, this fact is scarcely stressed in most of the literature on experimental design. In the case of linear models, an experiment is designed so that its level of statistical inference is maximised. The most common way to measure the inference level is in terms of the variance-covariance matrix of the parameter estimates. A small variance-covariance implies a high level of statistical inference.

For a monovariate linear model with uncorrelated error and constant error variance $\sigma^2_y$, a least squares estimator of the variance covariance matrix is given by:

$$W(\hat{\theta}) = \sigma^2_y (X^T X)^{-1}$$  

(2.20)

where the matrix $X$ is defined analogously to the matrix $S$ of equation 2.18. The determinant of a matrix can be used as a measure of its size. Therefore an optimal experiment can be designed by minimising $|W(\hat{\theta})|$, which requires minimising $|X^T X|^{-1}$. This is equivalent to the maximisation of $|S^T S|$ required by equation 2.18.

This important formal analogy should not hide the essential difference that while equation 2.20 is not dependent on the value of parameter estimates, such a dependence exists in equation 2.18. This is the reason why nonlinear models require a probabilistic approach in order to derive a design criterion. Similar considerations can be applied to the multivariate case.

An experiment designed by maximisation of equation 2.18 or 2.19 is called, by analogy with the theory of linear design, $D$-optimal. This criterion is also known for its "geometrical interpretation". The minimisation of $|W(\hat{\theta})|$, as in equation 2.20, causes the minimisation of the joint confidence region of parameter estimates, which is an ellipsoid in the parameter space. This behavior is preserved for a nonlinear model, although the joint confidence region is usually some less regularly shaped area. This causes the criterion being widely known as the volume criterion.
Other design criteria have been derived from the D-optimal criterion. The \textit{A-optimal} design criterion minimise the trace of the matrix \((X^T X)^{-1}\), which is equivalent to maximise the trace of \(S^T S\). This design has been used mainly when the off-diagonal elements of the variance-covariance matrix are much smaller than the diagonal ones and no appreciable loss of information is caused by ignoring them. Moreover the A-optimal criterion allows handling ill-conditioned matrices, whose determinant is not easily computed. D-optimal design and A-optimal design are equivalent when the variance covariance matrix is exactly diagonal. This criterion is scarcely suitable to be used when relatively noisy experimental data are likely to generate large cross-correlations in the parameter estimates.

It has been noticed that a by-product of the D-optimal design is often the growth of one of the variance terms and of the correspondent cross-correlation terms, even though an overall decrease of \(|W(\hat{\theta})|\) takes place. The D-optimal criterion weighs more the parameters with the highest sensitivities and the decrease in interval of confidence is mostly concentrated on these. Using the geometrical analogy, such a situation can be visualised by the ellipsoid reducing its volume, but simultaneously stretching along one or more axis. The relatively highly correlated parameters which arise are sometimes undesirable. Hosten [1974] suggested to minimise the largest eigenvalue of \((X^T X)^{-1}\). This is equivalent to maximise the smallest eigenvalue of \(S^T S\). This criterion is called \textit{E-optimal}. The eigenvalues of this matrix are proportional to the variance terms of the parameter estimates (Rosenbrock and Storey 1966).

Other design criteria, G-optimal and F-optimal [Fedorov 1972] well known for the linear case, are not suitable for the nonlinear design. They use a measure of statistical inference which does not exhibit any formal analogy to the results of the bayesian approach.

Pritchard and Bacon [1978], proposed a criterion aimed at reducing the level of correlation among parameters. This is achieved by concentrating the design on the reduction of the off-diagonal elements of the correlation matrix. The correlation matrix is a normalisation of \(W(\hat{\theta})\) having unity diagonal elements and off-diagonal elements ranging from zero to one, proportionally to the level of cross-correlation among the relevant parameters. A reduction of the level of correlation would be highly welcome when a parameter representative of a certain physical phenomenon
is to be used in a context different from the experimental one. However, Agarwal and Brisk [1984b] have shown that such a criterion does not give appreciably better results than those obtained with a simple D-optimal design.

2.4 Computer codes for sequential design of experiments.

In implementing a code for the sequential design of experiments, two goals had to be achieved: efficiency and generality of use. In turn these goals influenced the choice of a design criterion.

The A-optimal design can be discarded for the limitation discussed above. The E-optimal design criterion is scarcely suitable for such designs where the level of uncertainty on parameters is high and/or there is considerable level of experimental error. The D-optimal design appears as the most likely choice and it has been advocated in early algorithms for the sequential design of experiments.

Mitchell [1974a,b] implemented a D-optimal design strategy in the program DETMAX. The number of experimental runs to be designed is chosen by the experimenter and the program uses a random search optimisation strategy to maximise $|X^TX|$ in the feasible experimental space. The program was strictly intended for linear monovariate models and the parameter values are not taken in any account. The implementation of DETMAX was improved by Galil and Kiefer [1980] with gains in speed and memory requirements.

Beside the obvious lack of generality, DETMAX required the user to specify the experimental space as a set of discrete possible experimental points. A similar description of the independent variables space has been used also by Welch, [1982]. He used a branch and bound optimisation technique with the bounded experimental space approximated by a grid of discrete points. The distance between adjacent points had to be specified by the user, however computer time and memory requirements proved to be a severe limiting factor in the density of the grid.

The previous authors defended the choice of discretisation of the design space claiming that no experimenter can realistically manipulate the independent variables in a totally continuous fashion. This is only partially true. The typical independent variables in kinetic experiments carried out in a laboratory, such as
pressure, temperature, concentration and weight, can be set with such a precision that a continuous design space is a more likely representation.

Cook and Nachtstheim [1980], used a conjugate direction algorithm with a continuous description of the search space, and a later refinement of the same code [Johnson and Nachtstein, 1983] allowed an improvement in the code’s performance. A different approach was used by Evans [1979] who used a Nelder-Mead simplex algorithm to augment an existing design by a fixed, user chosen, number of experimental points so that the final design is D-optimal.

Although the previous authors dealt only with linear models, they all point out the difficulties of the optimisation problem and the utmost necessity of a sensible starting point. They advocate the use of repeated design before accepting a solution as the optimal one. This is a rather typical clue of highly nonlinear optimisation problems. Such problems are not efficiently and reliably tackled by direct optimisation methods such as those described.

The use of an indirect optimisation method would improve the efficiency of the code and allow the handling of nonlinear equality/inequality constraints in the design space, such as those arising from stoichiometric and other physical relations. However an indirect optimisation method requires the gradient of the objective functions 2.18 and 2.19 and the complex form of such functions does not allow a reliable computation by finite differences. Bates [1983] introduced an analytical expression of the gradient of the objective function 2.18. This and an extension to functions such as equations 2.19 are discussed in some detail in the next section.

The optimisation method chosen here is a robust successive quadratic programming algorithm as implemented by Chen [1988] in the code SQPD. This code, as other quadratic programming routines, uses an infeasible path approach. However it separates simple bounds on the design variables from user defined nonlinear constraints. Other codes, such as VF20AD, Powell [1977], force the user to define simple bounds as linearised constraints. Such constraints, unlike the simple bounds, are not required to be satisfied during any iteration. This is undesirable because overshooting a bound might cause the code to reach regions where the models 2.1 and 2.11 are not defined, with subsequent failure of the optimisation.
2.5 Numerical aspects of the sequential design of experiments.

The objective is to find the settings for \( m \) independent experimental variables \( x \) in \( n_e \geq 1 \) experiments so that:

\[
\max |M(\phi, \hat{\theta})| \tag{2.21}
\]

is obtained. \( \phi \) is the vector of design variables defined as:

\[
\phi = [x_{n+1}^T, x_{n+2}^T, \ldots, x_{n+n_e}^T]^T \tag{2.22}
\]

\( \hat{\theta} \) is the vector of \( p \) parameter estimates as calculated in the previous \( n \) experiments.

The information matrix \( M \) is a \( pxp \) matrix defined as the argument of the determinants in the right-hand sides of equations 2.18 (monovariate case) and 2.19 (multivariate case). Such determinants depend on the design matrices \( S \) (monovariate case) or \( S_r \) (\( r=1 \ldots v \)) (multivariate case). In both cases the first \( n \) rows of the design matrices depend only on the \( n \) prior experiments and are therefore kept constant.

Constraints are allowed, such as simple bounds on the independent variables:

\[
x_j^L \leq x_{ji} \leq x_j^U \quad \begin{array}{ll}
i = n + 1 \ldots n + n_e \\
j = 1 \ldots m
\end{array} \tag{2.23}
\]

and general nonlinear equality/inequality constraints such as:

\[
h(x) \geq 0 \quad \begin{array}{ll}
i = n + 1 \ldots n + n_e
\end{array} \tag{2.24}
\]

2.5.1 Evaluation of the objective function.

Bates [1983], suggests some algebraic manipulations on the design matrix for the monovariate case, which allow a sensible gain in computational efficiency. The matrix \( S \) is decomposed as follows:

\[
S = QR = Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1 \tag{2.25}
\]

where \( S \) is the design matrix defined as from equation 2.8. \( S \) has dimensions \((n+n_e)x_p\) where \( n_e \geq 1 \) is the number of experiments to be designed. \( Q \) is a \((n_e+n)x(n_e+n)\) orthogonal matrix and \( R \) a \((n_e+n)x_p\) upper triangular matrix. \( Q_1 \) and \( R_1 \) are respectively the matrices defined by the first \( p \) columns of \( Q \) and the first \( p \) rows of \( R \). Because of the orthogonality of \( Q \), it holds:

\[
S^T S = R_1^T Q_1^T Q_1 R_1 = R_1^T R_1 \tag{2.26}
\]

30
and the objective function 2.21 can be computed as:

\[ \delta = |S^T S| = |R_1^T R_1| = |R_1|^2 = \left( \prod_{i=1}^{p} r_{ii} \right)^2 \]  

(2.27)

where the \( r_{ii} \)'s are the diagonal elements of \( R_1 \).

As noted, the design matrix \( S \) is naturally partitioned in two parts as the first \( n \) rows are kept fixed during the optimisation. Making this partition explicit the information matrix can be rewritten as:

\[
S^T S = \begin{bmatrix} S_n^T & S_{n_e}^T \end{bmatrix} \begin{bmatrix} S_n & S_e \end{bmatrix} = S_n^T S_n + S_{n_e}^T S_{n_e} 
\]

(2.28)

where \( S_n \) is the matrix obtained by the first \( n \) rows of \( S \) and \( S_{n_e} \) the matrix obtained by the following \( n_e \) rows. Applying to the first term of the sum a QR decomposition as described in equations 2.25 and 2.26 one obtains:

\[
S^T S = R_{1n}^T R_{1n} + S_{n_e}^T S_{n_e} 
\]

(2.29)

From equation 2.29 it emerges the reduced design matrix:

\[
S_R = \begin{bmatrix} R_{1n} \\ S_{n_e} \end{bmatrix}
\]

(2.30)

The matrix \( S_R \) can be used to compute the objective function 2.27 instead of the full matrix \( S \). This is computationally convenient since \( S_R \) has dimensions \((p+n_e) \times p \) rather than \((n+n_e) \times p \). The computations of the objective function using the reduced design matrix is carried out taking advantage of QR decomposition, as in equations 2.24, 2.26, 2.27.

Such manipulations cannot be carried out in the multivariate case and the information matrix has to be computed in its entirety at each iteration. However, as before, only the last \( n_e \) rows of the design matrices \( S_r \) must be updated. The determinant of \( M \) is calculated by its LU decomposition [Dahlquist and al., 1974].

2.5.2 Evaluation of the gradients.

Given a square matrix \( M \), Bard [1974] shows that the following is true:

\[
\frac{\partial |M|}{\partial M} = (M^{-1})^T |M| \]  

(2.31)

or equivalently

\[
\frac{\partial \log |M|}{\partial M} = (M^{-1})^T \]

(2.32)

In the monovariate case as \( M = S^T S \), it follows that \( dM = 2SdS \) and therefore equations 2.31 and 2.32 become:

\[
\frac{\partial \delta}{\partial S} = 2(S^+) |S^T S| \]

(2.33)
and

$$\frac{\partial \log \delta}{\partial \mathbf{S}} = 2(S^+)$$  \hspace{1cm} (2.34)

where $\delta$ is the objective function as defined in equation 2.27 and $S^+$ is the generalized pseudoinverse of $S$. Equation 2.34 is convenient if the objective function is transformed as follows:

$$\gamma = \log(\delta) = \frac{1}{2} \log\left(\prod_{i=1}^{p} r_{ii}\right)$$  \hspace{1cm} (2.35)

The elements of the gradient of the objective function 2.35 with respect to each element of the vector of design variables $\phi$ are given by:

$$\frac{\partial \gamma}{\partial \phi_k} = \sum_{i=n+1}^{n+n_{e}} \sum_{j=1}^{p} \frac{\partial s_{ij}}{\partial \phi_k} \left[(S^+)^T\right]_{ij}$$  \hspace{1cm} (2.36a)

$$\frac{\partial s_{ij}}{\partial \phi_k} = \frac{\partial}{\partial \phi_k} \left[\frac{\partial g(y_j, x_i, \delta)}{\partial \theta_j}\right]$$  \hspace{1cm} (2.36b)

where $\left[(S^+)^T\right]_{ij}$ indicates the $ij^{th}$ element of the transpose pseudoinverse of $S$.

The partial derivatives 2.36b can be computed either by finite differences or analytically.

When $M$ is for a multivariate model, the gradient of the objective function 2.21 is computed by equation 2.31. The elements of the gradient of the objective function with respect to the design variables vector are given by:

$$\frac{\partial \Delta}{\partial \phi_k} = |M| \sum_{i=1}^{p} \sum_{j=1}^{p} \frac{\partial m_{ij}}{\partial \phi_k} \left[(M^{-1})^T\right]_{ij}$$  \hspace{1cm} (2.37)

The evaluation of the pseudoinverse $S^+$ is carried out by singular value decomposition of the matrix $S$. Such a decomposition is defined as follows:

$$S = U \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} V^T$$  \hspace{1cm} (2.38)

where $U$ and $V$ are respectively $n \times n$ and $p \times p$ orthogonal matrices and $D$ is a diagonal matrix of rank $p$ having as elements the singular values of $S$. The pseudoinverse is computed by:

$$S^+ = V \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T$$  \hspace{1cm} (2.39)

where $D^{-1}$ is the inverse of $D$. 

32
When $S$ is either structurally or numerically ill-conditioned, some of its singular values are so small that the calculation of $D^{-1}$ could be impossible without incurring an overflow. A tolerance value can be set and any singular value below it can be ignored, setting the relevant element of $D^{-1}$ equal to zero. This allows the optimisation to continue with a less ill-conditioned approximation to $S$.

The use of equations 2.38 and 2.39 with a square matrix $M$ yields the inverse $M^{-1}$. This technique is used to compute the inverse needed in equation 2.37. Besides reducing the amount of code, this is advantageous as it enables one to control the inversion of ill-conditioned matrices.

2.6 Description of PRODES: a computer code for the sequential design of nonlinear experiments.

The numerical details discussed in the previous section allow one to outline a general purpose algorithm. For the sake of simplicity, the two cases, monovariate and multivariate, are discussed separately. However the code has been implemented to keep most of the differences hidden to the user. In both cases, the user has to supply an initial guess for the vector of design variables $\phi$, the settings of the previous $n$ experiments, an estimate of the parameters $\hat{\theta}$, the simple bounds on the design space and the partial derivatives required to evaluate the elements of the design matrices. Moreover, in the multivariate case an estimate of the inverse of the variance covariance matrix $\Sigma^{-1}$ should be supplied, however, if unavailable, a default matrix is used. If applicable, equations for the nonlinear constraints and their gradients should also be supplied.

2.6.1 Monovariate models.

The steps taken by PRODES when designing a set of $n_e \geq 1$ new experiments for a monovariate models are:

0 - User set any parameters, tolerances for the optimisation routine and value of the parameter vector $\theta$.

1 - Calculate the design matrix $S$ according to equation 2.8, but with $i=1...n_e$ ($n_e \geq 1$) using the vector of parameter estimates $\hat{\theta}$.

2 - Carry out a QR decomposition of the first $n$ rows of the design matrix $S$ as by equation 2.25.
3 - Compute the reduced design matrix $S_R$ of equation 2.30.

4 - Compute the objective function $\gamma$ (eq. 2.35) by applying the QR decomposition indicated in equations 2.25 and the calculation procedure for the determinant indicated in equations 2.26, 2.27 and 2.35 to the reduced design matrix $S_R$.

5 - Carry out a singular value decomposition of the reduced design matrix and compute its pseudoinverse.

6 - Calculate the gradient of the objective function by equation 2.36. The partial derivatives 2.36b are computed by finite differences\(^{(1)}\) if not available analytically.

7 - Compute any nonlinear inequality/equality constraints (eq. 2.24) and their gradients. These are computed $n_\phi$ times for each of the $x_i$.

8 - Call the optimisation routine.

9 - On exit from the optimiser, if no termination criteria have been met, update the variable part of the reduced design matrix and repeat from point 4.

2.6.2 Multivariate models.

The difference in the algorithm for multivariate models is mainly in the objective function calculation. Steps 1 to 6 are replaced by:

1 - Calculate the information matrix $M$ defined as the argument of the determinant in equation 2.19.

2 - Calculate the objective function $|M|$.

3 - Calculate, by singular value decomposition, the inverse of $M$.

4 - Calculate the gradient of the objective function, by equation 2.37. The partial derivatives $\partial m_{ij}/\partial \phi_k$ are computed by finite differences.\(^{(1)}\)

The subsequent steps are the same as step 7,8 and 9 of the monovariate case with the difference that the loop from step 9 should return to step 1 as no reduced design matrix is present.

\(^{(1)}\) Details on the finite differences scheme used are reported in Appendix D.
2.6.3 General remarks.

Maximum ease of use is achieved by having default values for all the tolerances needed and for the perturbation step used in the finite difference calculation of the partial derivatives. However such values can be changed by the user. A robust implementation for the algebraic manipulations required is supplied by the routines of the linear algebra library LINPACK, Dongarra and al. [1979]. Appendix C reports user instructions and calling statement of the routine PRODES as implemented in FORTRAN 77.

2.7 Case study: nitric oxide reduction.

In this section various optimal experimental design strategies are carried out, utilising the PRODES software, with reference to a model describing the reduction of nitric oxide over vanadium pentoxide catalyst. Ayen and Peters [1962], found that the reaction:

\[ \text{NO} + \text{H}_2 \rightleftharpoons \text{H}_2\text{O} + \frac{1}{2}\text{N}_2 \]

can be described by the following Hougen-Watson type rate equation:

\[
r = \frac{kK_{NO}K_{H_2}P_{NO}P_{H_2}}{(1 + K_{NO}P_{NO} + K_{H_2}P_{H_2})^2}
\]

(2.40)

where \( k \) is the kinetic constant for the forward reaction and \( K_{NO} \) and \( K_{H_2} \) are the equilibrium constant of nitric oxide and hydrogen, respectively. In a following study, Kittrel and al. [1966] used a D-optimal design to improve the parameter estimates. Such a work is widely quoted as an example of the application of D-optimal design [Himmelblau 1970, Froment and Bishoff 1979] and it provides a sound term of comparison for the results obtained with PRODES. In the model \( k, K_{NO} \) and \( K_{H_2} \) are the parameters to be estimated and the partial pressures \( p_{NO} \) and \( p_{H_2} \) are the independent variables. The reaction rate \( r \) is measured by means of the differential method of kinetic analysis.

<table>
<thead>
<tr>
<th>( k )</th>
<th>( K_{NO} )</th>
<th>( K_{H_2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 4.94 \times 10^{-4} )</td>
<td>14.64</td>
<td>19.00</td>
</tr>
</tbody>
</table>

Table 2.1. — True parameter values at 400°C for the nitric oxide reduction rate expression

35
Simulated experimental points have been generated by equation 2.40 with the “true” parameter values of table 2.1. A normally distributed error with variance $\sigma^2 = 9.5 \times 10^{-12}$ and zero mean was added to the computed reaction rates. The value of the error variance corresponds to the value estimated by Kittrel and coworkers. The partial pressures of nitric oxide, $p_{NO}$, and of hydrogen, $p_{H_2}$, should be higher than 0.01 atm and not exceeding 0.10 atm. The simulated experiments were carried out at the constant temperature of 400°C. In the following subsections three cases are considered: in the first one the model 2.40 is used, in the subsequent ones the effects of two different parameter transformations are considered. The parameter estimation for this and the other cases in this chapter was carried out by using the code GREG [Steward, 1987]. GREG is a generalized regression package for nonlinear algebraic models. It implements a least squares algorithm for monovariate models and an extended Box-Draper objective function [Box and Draper 1965] for multivariate models.

2.7.1 Original model.

The results of six different experimental strategies are compared. The effect of the number of experiments in each batch of a batch sequential design and the effect of the initial design are pointed out. Table 2.2 summarizes the experiments carried out.

<table>
<thead>
<tr>
<th>Run</th>
<th>Strategy</th>
<th>Number of batches</th>
<th>Initial design</th>
<th>Experiments in initial design</th>
<th>Total number of experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>BS</td>
<td>3</td>
<td>F</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>BS</td>
<td>6</td>
<td>F</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>C</td>
<td>BS</td>
<td>4</td>
<td>U</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>D</td>
<td>BS</td>
<td>6</td>
<td>U</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>E</td>
<td>U</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>17</td>
</tr>
<tr>
<td>F</td>
<td>FR</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 2.2. — Summary table of experiments carried out. BS: batch sequential. FR: repeated factorial. U: unplanned.
The initial design used for runs A and B was a $2^2$ orthogonal design (5 experiments). The same design, replicated three times, was used in run F. Run E was totally unplanned, points having been chosen manually trying to cover uniformly the available experimental space. The first six points of run E were used to give starting parameters for the designed runs C and D. Figure 2.1 shows the experimental grid used for run E. Table 2.3 shows the parameter estimates obtained after runs E and F.

<table>
<thead>
<tr>
<th>Run</th>
<th>$k \times 10^4$</th>
<th>$K_{NO}$</th>
<th>$K_{H_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>3.86 ± 1.04</td>
<td>20.98 ± 9.80</td>
<td>26.59 ± 12.28</td>
</tr>
<tr>
<td>F</td>
<td>4.82 ± 0.68</td>
<td>16.76 ± 4.71</td>
<td>17.49 ± 4.76</td>
</tr>
</tbody>
</table>

Table 2.3. — Replicated orthogonal design (Run F) and unplanned experiment (Run E). Parameter estimates and 95% interval of confidence.

The orthogonal design chooses the experimental points in a non intuitive manner, however it avoids the introduction of unwanted dependencies among independent variables. This is likely to happen if points are empirically chosen so that the
available experimental area is uniformly covered. Figure 2.1 shows a rather evi-
dent linear dependency among the partial pressures. Such a dependency is mainly
responsible for the poor quality of the estimates from run E (First row of table 2.3).

Runs A and B are illustrated in table 2.4 to 2.7. It can be observed that the
final value of the objective function is similar in both runs. However run B, with
smaller batches, performs better as the parameter estimates are closer to the true
values of table 2.1. In the runs A and B it is of particular interest to notice that,
although the value of the objective function around the 12th experiment is lower
than the one originally reported by Kittrel and al. [1966], table 2.8, the central
value of the parameter estimates at the same point are slightly better, even though
the intervals of confidence are slightly wider. However, unlike in our study, in the
original paper no further improvement was achieved by further experimentation.
This is likely due the more precise optimisation algorithm used here to carry out
the design.

| Experiment | Batch | $p_{NO}$ atm | $p_{H_2}$ atm | $r \times 10^5$ | $|M|$ |
|------------|-------|--------------|--------------|----------------|-----|
| 1          | 1     | 0.100        | 0.100        | 7.373          |     |
| 2          | 1     | 0.010        | 0.100        | 1.602          |     |
| 3          | 1     | 0.010        | 0.010        | 1.002          |     |
| 4          | 1     | 0.100        | 0.010        | 2.074          |     |
| 5          | 1     | 0.050        | 0.050        | 4.362          |     |
| 6          | 2     | 0.100        | 0.034        | 5.189          |     |
| 7          | 2     | 0.042        | 0.100        | 4.864          |     |
| 8          | 2     | 0.042        | 0.100        | 4.694          |     |
| 9          | 2     | 0.100        | 0.100        | 7.324          |     |
| 10         | 2     | 0.035        | 0.035        | 2.950          | 5.1416 $\times 10^{-25}$ |
| 11         | 3     | 0.100        | 0.100        | 7.738          |     |
| 12         | 3     | 0.100        | 0.100        | 7.233          |     |
| 13         | 3     | 0.100        | 0.100        | 7.769          |     |
| 14         | 3     | 0.100        | 0.100        | 7.296          |     |
| 15         | 3     | 0.100        | 0.100        | 7.484          | 1.5830 $\times 10^{-24}$ |

Table 2.4. — Run A. 15 experiments in 2 batches of
five experiments each following an initial $2^2$ orthogonal
design. Optimal experimental grid, measured reaction
rate and objective function evaluation.
Observing the experimental points of runs A and B of table 2.4 and 2.6, it appears that there is a tendency to create clusters of points within the same batch. This behaviour is common to all the situations and models examined in this and other cases.

<table>
<thead>
<tr>
<th>Batch</th>
<th>$k \times 10^4$</th>
<th>$K_{NO}$</th>
<th>$K_{H_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.45 ± 1.88</td>
<td>12.05 ± 7.91</td>
<td>16.75 ± 8.42</td>
</tr>
<tr>
<td>2</td>
<td>5.45 ± 0.81</td>
<td>12.37 ± 3.28</td>
<td>17.24 ± 3.72</td>
</tr>
<tr>
<td>3</td>
<td>5.58 ± 0.64</td>
<td>11.98 ± 2.77</td>
<td>16.81 ± 3.15</td>
</tr>
</tbody>
</table>

Table 2.5. — Run A. Evolution of parameter estimates and 95% interval of confidence.

| Experiment | Batch | $p_{NO}$ atm | $p_{H_2}$ atm | $r \times 10^5$ | $|M|$ |
|------------|-------|--------------|---------------|----------------|------|
| 1          | 1     | 0.100        | 0.100         | 7.373          |      |
| 2          | 1     | 0.010        | 0.100         | 1.602          |      |
| 3          | 1     | 0.010        | 0.010         | 1.002          |      |
| 4          | 1     | 0.100        | 0.010         | 2.074          |      |
| 5          | 1     | 0.050        | 0.050         | 4.362          |      |
| 6          | 2     | 0.100        | 0.100         | 6.769          |      |
| 7          | 2     | 0.100        | 0.100         | 6.910          | 5.5405 $\times 10^{-26}$ |
| 8          | 3     | 0.100        | 0.100         | 7.173          |      |
| 9          | 3     | 0.100        | 0.100         | 6.799          | 6.5585 $\times 10^{-26}$ |
| 10         | 4     | 0.036        | 0.100         | 3.820          |      |
| 11         | 4     | 0.036        | 0.100         | 4.514          | 2.1703 $\times 10^{-26}$ |
| 12         | 5     | 0.100        | 0.031         | 4.556          |      |
| 13         | 5     | 0.100        | 0.031         | 4.324          | 6.6278 $\times 10^{-25}$ |
| 14         | 6     | 0.031        | 0.028         | 3.007          |      |
| 15         | 6     | 0.031        | 0.028         | 3.183          | 1.5434 $\times 10^{-24}$ |

Table 2.6. — Run B. 15 experiments in 6 batches of two experiments each following an initial $2^3$ orthogonal design. Optimal experimental grid, measured reaction rate and objective function evaluation.
Table 2.7. — Run B. Evolution of parameter estimates and 95% interval of confidence.

<table>
<thead>
<tr>
<th>Batch</th>
<th>$k \times 10^4$</th>
<th>$K_{NO}$</th>
<th>$K_{H_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.45 ± 1.88</td>
<td>12.05 ± 7.91</td>
<td>16.75 ± 8.42</td>
</tr>
<tr>
<td>2</td>
<td>4.91 ± 1.19</td>
<td>13.77 ± 7.55</td>
<td>18.96 ± 7.90</td>
</tr>
<tr>
<td>3</td>
<td>4.88 ± 1.01</td>
<td>13.87 ± 6.64</td>
<td>19.10 ± 6.88</td>
</tr>
<tr>
<td>4</td>
<td>4.77 ± 0.86</td>
<td>14.92 ± 5.52</td>
<td>18.94 ± 7.61</td>
</tr>
<tr>
<td>5</td>
<td>4.76 ± 0.76</td>
<td>14.95 ± 4.99</td>
<td>19.11 ± 5.42</td>
</tr>
<tr>
<td>6</td>
<td>4.64 ± 0.49</td>
<td>15.80 ± 3.47</td>
<td>20.07 ± 3.74</td>
</tr>
</tbody>
</table>

Table 2.8. — Final parameter estimates and objective value reported by Kittrel and coworkers [1966] after 12 runs.

| $k \times 10^4$ | $K_{NO}$     | $K_{H_2}$     | $|M|$          |
|----------------|--------------|--------------|----------------|
| 4.70 ± 0.53    | 16.09 ± 4.00 | 20.20 ± 4.20 | 8.1360 x $10^{-25}$ |

Figure 2.2. — Untransformed model. Evolution of 95% interval of confidence of parameter $k$. 
Run C and D were carried out using as a first batch the first six points of run E. These are indicated by a star in figure 2.1. It is quite clear that such points are extremely poor for estimation purposes as a very strong linear dependency is present. The objective of these runs was to assess whether a sound design strategy can be used to improve upon an otherwise weak series of experiments. The results are shown in tables 2.9 to 2.12.

<table>
<thead>
<tr>
<th>Batch</th>
<th>( k \times 10^4 )</th>
<th>( K_{NO} )</th>
<th>( K_{H_2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.49 ± 0.49</td>
<td>17.15 ± ∞</td>
<td>20.14 ± 5.57</td>
</tr>
<tr>
<td>2</td>
<td>5.07 ± 0.98</td>
<td>14.35 ± 5.57</td>
<td>17.02 ± 5.96</td>
</tr>
<tr>
<td>3</td>
<td>4.95 ± 1.00</td>
<td>14.53 ± 6.52</td>
<td>18.68 ± 6.55</td>
</tr>
<tr>
<td>4</td>
<td>5.07 ± 0.95</td>
<td>14.10 ± 5.81</td>
<td>17.82 ± 5.49</td>
</tr>
</tbody>
</table>

Table 2.9. — Run C. Evolution of parameter estimates and 95% interval of confidence.

Figure 2.3. — Untransformed model. Evolution of 95% interval of confidence of parameter \( K_{NO} \).
Figure 2.4. — Untransformed model. Evolution of 95% interval of confidence of parameter $K_{H_2}$.

| Experiment | Batch | $p_{NO}$ atm | $p_{H_2}$ atm | $r \times 10^5$ gmol/min g-cat | $|M|$ |
|------------|-------|---------------|---------------|-------------------------------|------|
| 1          | 1     | 0.010         | 0.100         | 1.562                         |      |
| 2          | 1     | 0.100         | 0.010         | 2.219                         |      |
| 3          | 1     | 0.080         | 0.030         | 3.943                         |      |
| 4          | 1     | 0.060         | 0.030         | 4.936                         |      |
| 5          | 1     | 0.040         | 0.070         | 4.910                         |      |
| 6          | 1     | 0.020         | 0.020         | 2.546                         |      |
| 7          | 2     | 0.100         | 0.100         | 7.377                         |      |
| 8          | 2     | 0.100         | 0.100         | 7.448                         |      |
| 9          | 2     | 0.100         | 0.100         | 7.338                         |      |
| 10         | 2     | 0.100         | 0.100         | 6.801                         |      |
| 11         | 3     | 0.100         | 0.100         | 7.331                         |      |
| 12         | 3     | 0.100         | 0.100         | 7.562                         |      |
| 13         | 3     | 0.100         | 0.032         | 4.884                         |      |
| 14         | 3     | 0.100         | 0.100         | 6.482                         | 1.8500 $\times 10^{-25}$ |
| 15         | 4     | 0.100         | 0.030         | 4.320                         |      |
| 16         | 4     | 0.100         | 0.100         | 7.214                         |      |
| 17         | 4     | 0.100         | 0.100         | 7.710                         | 1.4117 $\times 10^{-24}$ |

Table 2.10. — Run C. 17 runs in 4 batches. The first batch has not been designed. Optimal experimental grid, measured reaction rate and objective function evaluation.
The comparison of the final results of runs C and D with those of run E, shows quite clearly that even an initially flawed design can be sensibly improved. However, as shown in figures 2.2 to 2.4, where the evolution of the intervals of confidence of the various parameters versus the number of experiments is illustrated, the results of these runs are still considerably worse than those of runs A and B. Only runs A and B, because of a sounder initial design, outperform run F, where a repeated orthogonal design was used. Moreover, it is apparent that the value of the interval of confidence in runs C and D tends towards an asymptotic value. It is likely that no further possibility of improvement is left, no matter how many designed experiments are carried out. It is likely that the design, given the poor initial guess, leads to a local optimum.

| Experiment | Batch | \( p_{NO} \) atm | \( p_{H_2} \) atm | \( r \times 10^5 \) gmol/min gr\(_{cat}\) | \(|M|\) |
|------------|-------|-----------------|-----------------|----------------------------------|-----|
| 1          | 1     | 0.010           | 0.100           | 1.562                            |     |
| 2          | 1     | 0.100           | 0.010           | 2.219                            |     |
| 3          | 1     | 0.080           | 0.030           | 3.943                            |     |
| 4          | 1     | 0.060           | 0.030           | 4.936                            |     |
| 5          | 1     | 0.040           | 0.070           | 4.910                            |     |
| 6          | 1     | 0.020           | 0.020           | 2.546                            |     |
| 7          | 2     | 0.100           | 0.100           | 7.574                            |     |
| 8          | 2     | 0.100           | 0.100           | 7.580 \(9.250 \times 10^{-26}\) |
| 9          | 3     | 0.100           | 0.100           | 7.006                            |     |
| 10         | 3     | 0.100           | 0.100           | 7.738 \(4.7760 \times 10^{-25}\) |
| 11         | 4     | 0.100           | 0.100           | 6.769                            |     |
| 12         | 4     | 0.100           | 0.032           | 4.443 \(8.9828 \times 10^{-25}\) |
| 13         | 5     | 0.100           | 0.100           | 6.823                            |     |
| 14         | 5     | 0.100           | 0.100           | 7.518 \(1.0140 \times 10^{-24}\) |
| 15         | 6     | 0.100           | 0.100           | 7.480                            |     |
| 16         | 6     | 0.100           | 0.100           | 6.794 \(1.2850 \times 10^{-24}\) |

Table 2.11. — Run D. 17 runs in 6 batches. The first batch has not been designed. Optimal experimental grid, measured reaction rate and objective function evaluation.
Table 2.12. — Run D. Evolution of parameter estimates and 95% interval of confidence.

An assessment of the level of correlation in the parameter estimates is done by evaluating the determinant and the condition number of the correlation matrix. The final value of the correlation matrix in the different runs is given in table 2.13. Ideally, unity determinant would indicate the total absence of any correlation, while complete correlation would be indicated by a determinant equal to zero.

![Graph](image)

Figure 2.5. — Untransformed model. Evolution of the determinant of the correlation matrix
Such a value is also caused by the presence of an undetermined parameter, such as $K_{NO}$ in the first batch of runs C and D. The condition number, defined as the ratio between the biggest and the smallest eigenvalue, of the correlation matrix, provides a measure of the relative level of cross correlation. Ideally this should be close to one. The evolution of these measures with the number of experiments is illustrated in figures 2.5 and 2.6.

![Figure 2.6. — Untransformed model. Evolution of the condition number of the correlation matrix](image)

It is observed that, while all the designed runs have a far better level of cross-correlation than the unplanned one, the static design still achieves a better final value. The condition number increases as the designed experimentation proceeds. This is predictable and corresponds to the stretching of the joint confidence region discussed previously. The decrease in the value of the determinant seems to be mostly concentrated in the first batches, as the last ones have more constant values.
Table 2.13. — Untransformed model. Correlation matrix at the end of different runs.

2.7.2 Transformed model: Ratkowski's transformation

The use of parametric transformations in nonlinear models in order to simplify the parameter estimation is a well known and established technique (Espie, 1986). This is convenient because any least square estimator relies on a local linearisation of the model about the parameter estimate \( \hat{\theta} \). Any manipulation of the model reducing the level of nonlinearity in the parameter space is bound to ease the estimation problem. Hougen–Watson type rate equations such as model 2.40 are common in catalysis studies. They are generally expressed as:

\[
r = \frac{\theta_1 g(p_1, p_2, \ldots, p_q)}{[f_1(p_1, p_2, \ldots, p_q) + \theta_2 f_2(p_1, p_2, \ldots, p_q) + \ldots + \theta_m f_m(p_1, p_2, \ldots, p_q)]^{\frac{1}{n}}} \tag{2.41}
\]
where \( p_1, p_2, \ldots, p_q \) are the partial pressures of the reagents and \( \theta_1, \theta_2, \ldots, \theta_m \) are the kinetic parameters. Ratkowski, [1985], shows that an adequate transformation of the rate equation 2.41 is given by:

\[
\frac{g(p_1, p_2, \ldots, p_q)}{[\gamma_1 f_1(p_1, p_2, \ldots, p_q) + \gamma_2 f_2(p_1, p_2, \ldots, p_q) + \ldots + \gamma_m f_m(p_1, p_2, \ldots, p_q)]^n}
\]

(2.42a)

where:

\[
\gamma_1 = \frac{1}{\theta_1^{1/n}}
\]

\[
\gamma_2 = (1/\theta_1^{1/n}) \theta_2
\]

\[
\ldots
\]

\[
\gamma_m = (1/\theta_1^{1/n}) \theta_m
\]

(2.42b)

Under transformation 2.42, model 2.40 becomes:

\[
r = \frac{PNOPH_2}{(k_1 + k_2 P_{NO} + k_3 P_{H_2})^2}
\]

(2.43)

Table 2.14 reports the expected values of the parameters following the transformation. Comparing tables 2.1 and 2.14, it appears that the values of the transformed parameters vary over a much narrower range of magnitudes. The advantages for parameter estimation are obvious.

\[
\begin{array}{ccc}
  & k_1 & k_2 & k_3 \\
2.697 & 39.494 & 51.256 \\
\end{array}
\]

Table 2.14. — New values of the "true" parameters of table 2.1 following the Ratkowski's transformation of equations 2.40.
A set of runs analogous to those of the previous section was carried out. The details are in table 2.15.

<table>
<thead>
<tr>
<th>Run</th>
<th>Strategy</th>
<th>Number of batches</th>
<th>Initial design</th>
<th>Experiments in initial design</th>
<th>Total number of experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>BS</td>
<td>3</td>
<td>F</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>B1</td>
<td>BS</td>
<td>4</td>
<td>F</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>C1</td>
<td>BS</td>
<td>4</td>
<td>U</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>D1</td>
<td>BS</td>
<td>6</td>
<td>U</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>E1</td>
<td>U</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>17</td>
</tr>
<tr>
<td>F1</td>
<td>FR</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 2.15. — Summary table of experiments carried out. Transformed model: Ratkowski's transformation. Symbols as in table 2.2.

The results are illustrated in figures 2.7 to 2.11 and tables 2.16 to 2.25.
Table 2.16. — Use of Ratkowski's transformation. Replicated orthogonal design (run F1) and unplanned experiment (run E1). Parameter estimates and 95% interval of confidence.

The positive effect of the parametric transformation on both interval of confidence and closeness of the parameters to the expected value is evident. However the parametric transformation affects any run, as also the unplanned run E1 results in better estimation. Runs C1 and D1, with an unplanned initial batch, yield eventual results much closer to those obtained by the designed ones, A1 and D1. Unlike in the previous section, runs A1 and D1 outperform in every respect, other than the overall level of correlation, run F1 designed by orthogonal design.

Table 2.17. — Run A1. Use of Ratkowski’s transformation. Optimal experimental grid, measured reaction rate and objective function evaluation.
The smaller condition numbers (figure 2.11) indicate a tendency towards more spherical joint confidence region. Moreover the growth of the condition number with the number of experiments is comparatively smaller.

The intervals of confidence displayed in figures 2.7 through 2.9 show a more marked asymptotic behaviour than in the previous section. Unlike with the untransformed model, the use of the Ratkowski's transformation does not clearly show that the design carried out with smaller batches is better. However, this is to be expected as for a linear model the design does not depend on the value of the parameters. Hence, two batches of two experiments each or one of four experiments should give the same results.

<table>
<thead>
<tr>
<th>Batch</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>$k_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.56 ± 0.531</td>
<td>34.99 ± 6.95</td>
<td>56.24 ± 9.93</td>
</tr>
<tr>
<td>2</td>
<td>2.60 ± 0.260</td>
<td>34.26 ± 3.20</td>
<td>56.08 ± 2.76</td>
</tr>
<tr>
<td>3</td>
<td>2.54 ± 0.271</td>
<td>35.62 ± 2.86</td>
<td>55.33 ± 2.80</td>
</tr>
</tbody>
</table>

Table 2.18. — Run A1. Use of Ratkowski's transformation. Evolution of parameter estimates and their 95% interval of confidence.
| Experiment | Batch | $p_{NO}$ atm | $p_{H_2}$ atm | $r \times 10^5$ | $|M|$ |
|------------|-------|--------------|--------------|---------------|--------|
| 1          | 1     | 0.100        | 0.010        | 2.310         |        |
| 2          | 1     | 0.010        | 0.010        | 0.963         |        |
| 3          | 1     | 0.050        | 0.050        | 4.798         |        |
| 4          | 1     | 0.010        | 0.100        | 1.589         |        |
| 5          | 1     | 0.100        | 0.100        | 7.000         |        |
| 6          | 1     | 0.050        | 0.050        | 5.298         |        |
| 7          | 2     | 0.100        | 0.100        | 7.599         |        |
| 8          | 2     | 0.100        | 0.100        | 7.018         |        |
| 9          | 2     | 0.100        | 0.100        | 7.393         |        |
| 10         | 2     | 0.190        | 0.100        | 2.473         | 3.3373 $\times 10^{-28}$ |
| 11         | 3     | 0.100        | 0.100        | 7.197         |        |
| 12         | 3     | 0.020        | 0.100        | 2.854         |        |
| 13         | 3     | 0.100        | 0.100        | 7.337         |        |
| 14         | 3     | 0.100        | 0.100        | 7.448         | 4.1172 $\times 10^{-28}$ |
| 15         | 4     | 0.100        | 0.100        | 7.233         |        |
| 16         | 4     | 0.028        | 0.050        | 4.008         |        |
| 17         | 4     | 0.020        | 0.100        | 2.777         | 9.8514 $\times 10^{-28}$ |

Table 2.19. — Run B1. Use of Ratkowski's transformation. Optimal experimental grid, measured reaction rate and objective function evaluation.

<table>
<thead>
<tr>
<th>Batch</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>$k_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.11 ± 0.48</td>
<td>39.71 ± 8.01</td>
<td>58.05 ± 7.85</td>
</tr>
<tr>
<td>2</td>
<td>2.33 ± 0.43</td>
<td>37.74 ± 6.47</td>
<td>56.68 ± 6.34</td>
</tr>
<tr>
<td>3</td>
<td>2.29 ± 0.35</td>
<td>38.78 ± 4.71</td>
<td>55.39 ± 4.94</td>
</tr>
<tr>
<td>4</td>
<td>2.19 ± 0.28</td>
<td>40.00 ± 3.70</td>
<td>55.32 ± 4.29</td>
</tr>
</tbody>
</table>

Table 2.20. — Run B1. Use of Ratkowski's transformation. Evolution of parameter estimates and their 95% interval of confidence.
Figure 2.8. — Ratkowski's transformation. Evolution of 95% interval of confidence of parameter $k_2$

Figure 2.9. — Ratkowski's transformation. Evolution of 95% interval of confidence of parameter $k_3$

It has to be remembered that the design obtained by a model and by some transformation of the model should not be compared on the basis of the objective function.
values. Such a comparison would be misleading as the structural differences between
the models lead to different design matrices.

From discussing the results one is able to state that a parametric transforma-
tion, intended to make the parameter estimation easier, affects also the sequential
design of experiments, making it more effective. Similar conclusions were reached
by Agarwal and Brisk [1984a].

| Experiment | Batch | $P_{NO}$ atm | $P_{H_2}$ atm | $r \times 10^5$ gmol/min $gr_{cat}$ | $|M|$ |
|------------|-------|--------------|--------------|---------------------------------|------|
| 1          | 1     | 0.010        | 0.100        | 1.562                           |      |
| 2          | 1     | 0.100        | 0.010        | 2.219                           |      |
| 3          | 1     | 0.080        | 0.030        | 3.943                           |      |
| 4          | 1     | 0.060        | 0.030        | 4.937                           |      |
| 5          | 1     | 0.040        | 0.070        | 4.910                           |      |
| 6          | 1     | 0.020        | 0.090        | 2.546                           |      |
| 7          | 2     | 0.100        | 0.100        | 7.233                           |      |
| 8          | 2     | 0.100        | 0.100        | 7.758                           |      |
| 9          | 2     | 0.100        | 0.100        | 7.295                           |      |
| 10         | 2     | 0.100        | 0.100        | 7.478                           | 7.0139 x 10^{-34} |
| 11         | 3     | 0.030        | 0.035        | 3.617                           |      |
| 12         | 3     | 0.100        | 0.033        | 4.939                           |      |
| 13         | 3     | 0.030        | 0.035        | 3.282                           |      |
| 14         | 3     | 0.100        | 0.033        | 4.858                           | 41.5910 x 10^{-34} |
| 15         | 4     | 0.100        | 0.032        | 4.448                           |      |
| 16         | 4     | 0.029        | 0.031        | 3.560                           |      |
| 17         | 4     | 0.010        | 0.032        | 4.677                           | 87.1730 x 10^{-34} |

Table 2.21. — Run C1. Use of Ratkowski's transformation. Optimal experimental grid, measured reaction rate and objective function evaluation.

<table>
<thead>
<tr>
<th>Batch</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>$k_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.91 ± 0.774</td>
<td>34.99 ± $\infty$</td>
<td>47.71 ± 13.36</td>
</tr>
<tr>
<td>2</td>
<td>3.03 ± 0.641</td>
<td>39.00 ± 7.22</td>
<td>46.58 ± 6.92</td>
</tr>
<tr>
<td>3</td>
<td>2.73 ± 0.390</td>
<td>35.42 ± 4.72</td>
<td>49.34 ± 4.58</td>
</tr>
<tr>
<td>4</td>
<td>2.60 ± 0.340</td>
<td>41.09 ± 4.01</td>
<td>49.25 ± 4.01</td>
</tr>
</tbody>
</table>

Table 2.22. — Run C1. Use of Ratkowski's transformation. Evolution of parameter estimates and their 95% interval of confidence.
| Experiment | Batch | $p_{NO}$ atm | $p_{H_2}$ atm | $r \times 10^5$ | $|M|$ |
|------------|-------|--------------|---------------|----------------|------|
| 1          | 1     | 0.010        | 0.100         | 1.562          |      |
| 2          | 1     | 0.100        | 0.010         | 2.219          |      |
| 3          | 1     | 0.080        | 0.030         | 3.943          |      |
| 4          | 1     | 0.060        | 0.030         | 4.937          |      |
| 5          | 1     | 0.040        | 0.070         | 4.910          |      |
| 6          | 1     | 0.020        | 0.090         | 2.546          |      |
| 7          | 2     | 0.100        | 0.100         | 7.574          |      |
| 8          | 2     | 0.100        | 0.100         | 7.408          | 3.8730 $\times 10^{-30}$ |
| 9          | 3     | 0.100        | 0.100         | 7.238          |      |
| 10         | 3     | 0.100        | 0.100         | 7.323          | 1.2081 $\times 10^{-30}$ |
| 11         | 4     | 0.024        | 0.100         | 3.137          |      |
| 12         | 4     | 0.024        | 0.100         | 3.673          | 1.7545 $\times 10^{-29}$ |
| 13         | 5     | 0.023        | 0.100         | 3.401          |      |
| 14         | 5     | 0.034        | 0.050         | 3.702          | 3.2211 $\times 10^{-29}$ |
| 15         | 6     | 0.100        | 0.100         | 7.393          |      |
| 16         | 6     | 0.033        | 0.050         | 3.718          | 5.1354 $\times 10^{-29}$ |

Table 2.23. — Run D1. Use of Ratkowski's transformation. Optimal experimental grid, measured reaction rate and objective function evaluation.

<table>
<thead>
<tr>
<th>Batch</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>$k_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.91 ± 0.774</td>
<td>34.99 ± $\infty$</td>
<td>47.71 ± 13.36</td>
</tr>
<tr>
<td>2</td>
<td>3.06 ± 0.710</td>
<td>38.74 ± 7.79</td>
<td>46.32 ± 7.51</td>
</tr>
<tr>
<td>3</td>
<td>2.97 ± 0.580</td>
<td>39.57 ± 6.49</td>
<td>47.15 ± 6.23</td>
</tr>
<tr>
<td>4</td>
<td>2.91 ± 0.570</td>
<td>41.02 ± 5.17</td>
<td>46.20 ± 5.89</td>
</tr>
<tr>
<td>5</td>
<td>2.89 ± 0.460</td>
<td>41.63 ± 4.23</td>
<td>45.77 ± 5.11</td>
</tr>
<tr>
<td>6</td>
<td>2.94 ± 0.370</td>
<td>41.42 ± 3.73</td>
<td>45.50 ± 4.58</td>
</tr>
</tbody>
</table>

Table 2.24. — Run D1. Use of Ratkowski's transformation. Evolution of parameter estimates and their 95% interval of confidence.

Table 2.25 reports the final values of the correlation matrix at the end of the different experimental runs.
Figure 2.10. — Ratkowski's transformation. Evolution of the determinant of the correlation matrix.

Figure 2.11. — Ratkowski's transformation. Evolution of the condition number of the correlation matrix.
2.7.3 Transformed model: Lumping of multiplicative parameters.

Although not particularly significant from a statistical point of view, a parametric transformation appealing to the user is the lumping of the multiplicative parameters present at the numerator of equation 2.40. Under such a transformation the model becomes:

\[
r = \frac{k_l p n o p h_2}{(1 + k_{n o p n o} + k_{h_2 p h_2})^2}
\]  

(2.44)

where \( k_l = k_{n o} k_{h_2} \). Such a transformation partially reduces the level of correlation among parameters.

In the present context, the main interest for such a transformation arises from the fact that model 2.44 is linear with respect to the parameter \( k_l \). Hill [1980] shows that when a parameter \( \theta_i \) appears linearly in an otherwise nonlinear model, the optimal value of \( |M| \) is independent from the particular value of \( \theta_i \). This is consistent with the design in the linear situation where the value of the objective function does on depend from the whole vector \( \hat{\theta} \).
One could expect this transformation to outperform, in terms of design of sequential experiments, the Ratkowski’s transformation, where all parameters occur nonlinearly. The experimental runs carried out are summarized in table 2.26. Only one size of experimental batch is considered.

<table>
<thead>
<tr>
<th>Run</th>
<th>Strategy</th>
<th>Number of batches</th>
<th>Initial design</th>
<th>Experiments in initial design</th>
<th>Total number of experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>BS</td>
<td>4</td>
<td>F</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>C2</td>
<td>BS</td>
<td>4</td>
<td>U</td>
<td>6</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 2.26. — Summary table of experiments carried out. Transformed model: lumping of multiplicative parameters. Symbols as in table 2.2.

The true values of the parameters in model 2.44 are those given in table 2.1, with the exclusion of $k_1$, the expected value of which is $0.137 \text{ gmol/(min gr}_{\text{cat}} \text{ atm)}$.

| Experiment | Batch | $p_{NO}$ atm | $p_{H_2}$ atm | $r \times 10^5$ (gmol/min gr$_{cat}$) | $|M|$ |
|------------|-------|--------------|---------------|--------------------------------------|------|
| 1          | 1     | 0.100        | 0.010         | 2.310                                |      |
| 2          | 1     | 0.010        | 0.010         | 0.962                                |      |
| 3          | 1     | 0.050        | 0.050         | 4.798                                |      |
| 4          | 1     | 0.010        | 0.100         | 1.589                                |      |
| 5          | 1     | 0.100        | 0.100         | 7.000                                |      |
| 6          | 1     | 0.050        | 0.050         | 5.298                                |      |
| 7          | 2     | 0.019        | 0.100         | 2.778                                |      |
| 8          | 2     | 0.100        | 0.100         | 7.238                                |      |
| 9          | 2     | 0.024        | 0.050         | 3.221                                |      |
| 10         | 2     | 0.019        | 0.100         | 2.370                                | 1.5595 $\times 10^{-25}$ |
| 11         | 3     | 0.100        | 0.100         | 7.738                                |      |
| 12         | 3     | 0.020        | 0.051         | 3.435                                |      |
| 13         | 3     | 0.100        | 0.100         | 7.769                                | 5.7673 $\times 10^{-25}$ |
| 14         | 4     | 0.034        | 0.050         | 3.604                                |      |
| 15         | 4     | 0.100        | 0.100         | 7.018                                |      |
| 16         | 4     | 0.033        | 0.050         | 7.393                                | 1.3299 $\times 10^{-24}$ |

Table 2.27. — Run A2. Optimal experimental grid, measured reaction rate and objective function evaluation.
From the results of runs A2 and C2, it appears that the parameter estimation is not improved by the lumping of the multiplicative parameter neither in the values of the intervals of confidence nor in the final value of the correlation.

### Table 2.28. — Experiment A2. Lumping of multiplicative parameters. Evolution of parameter estimates and their 95% interval of confidence.

<table>
<thead>
<tr>
<th>Batch</th>
<th>$k_1$</th>
<th>$K_{NO}$</th>
<th>$K_{H_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.217 ± 0.097</td>
<td>18.51 ± 8.66</td>
<td>27.04 ± 8.15</td>
</tr>
<tr>
<td>2</td>
<td>0.199 ± 0.060</td>
<td>17.47 ± 3.85</td>
<td>25.34 ± 5.43</td>
</tr>
<tr>
<td>3</td>
<td>0.159 ± 0.048</td>
<td>14.38 ± 3.31</td>
<td>21.65 ± 5.20</td>
</tr>
<tr>
<td>4</td>
<td>0.166 ± 0.055</td>
<td>15.67 ± 3.48</td>
<td>21.69 ± 5.56</td>
</tr>
</tbody>
</table>

### Table 2.29. — Run C2. Optimal experimental grid, measured reaction rate and objective function evaluation.

| Experiment | Batch | $P_{NO}$ atm | $P_{H_2}$ atm | $r \times 10^5$ gmol/min g$\text{cat}$ | $|M|$ |
|------------|-------|--------------|----------------|--------------------------------------|------|
| 1          | 1     | 0.010        | 0.100          | 1.562                                |      |
| 2          | 1     | 0.100        | 0.010          | 2.219                                |      |
| 3          | 1     | 0.080        | 0.030          | 3.943                                |      |
| 4          | 1     | 0.060        | 0.030          | 4.936                                |      |
| 5          | 1     | 0.040        | 0.070          | 4.9101                               |      |
| 6          | 1     | 0.020        | 0.090          | 2.546                                |      |
| 7          | 2     | 0.100        | 0.100          | 7.544                                |      |
| 8          | 2     | 0.100        | 0.100          | 7.408                                |      |
| 9          | 2     | 0.100        | 0.100          | 7.238                                |      |
| 10         | 2     | 0.026        | 0.026          | 7.438                                | 0.7243 $\times 10^{-30}$ |
| 11         | 3     | 0.070        | 0.040          | 4.746                                |      |
| 12         | 3     | 0.010        | 0.035          | 5.435                                |      |
| 13         | 3     | 0.040        | 0.100          | 5.542                                |      |
| 14         | 3     | 0.031        | 0.031          | 3.718                                | 1.4015 $\times 10^{-30}$ |
| 15         | 4     | 0.100        | 0.037          | 4.924                                |      |
| 16         | 4     | 0.029        | 0.029          | 3.068                                |      |
| 17         | 4     | 0.100        | 0.037          | 4.601                                | 2.4536 $\times 10^{-30}$ |
The lumping of multiplicative parameters does not affect the design as strongly as the Ratkowski’s transformation does although it yields a model linear in one parameter. The design appears to be more strongly affected by the overall level of nonlinearity of the model, rather than by its possible partial linearity. Such level is higher in model 2.42 than in model 2.44. A better assessment of such an hypothesis should require further theoretical investigation.

Figures 2.12 and 2.13 show the evolution of the determinant and the condition number of the correlation matrix respectively.
Figure 2.12. — Lumping of multiplicative parameters. Evolution of the determinant of the correlation matrix

Figure 2.13. — Lumping of multiplicative parameters. Evolution of the condition number of the correlation matrix
2.8 Case study: Ethylene hydrogenation.

The following rate equation to describe the hydrogenation of ethylene was introduced by Barton [1976]:

\[ r = k_0 \exp \left(-\frac{E}{RT}\right) x_{C_2H_4} x_{H_2}^\alpha \]  

(2.45)

The "true" parameter values are given in table 2.32 and \( x_{C_2H_4} \) and \( x_{H_2} \) are the molar fraction of ethylene and hydrogen respectively. The molar fractions and the temperature \( T \) are the design variables.

<table>
<thead>
<tr>
<th>Rate constant</th>
<th>Activation energy</th>
<th>Pre-exponential factor</th>
<th>Temperature</th>
<th>Molar fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_0 ) sec (^{-1} )</td>
<td>( E ) J/mol (^{-1} )</td>
<td>( \alpha )</td>
<td>( \beta )</td>
<td></td>
</tr>
<tr>
<td>( 4.62 \times 10^{-4} )</td>
<td>( 1.160 \times 10^4 )</td>
<td>0.330</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.32. — Parameter value for the ethylene hydrogenation model.

Model 2.45 is transformed by substituting into the Arrhenius expression its reparameterisation about a reference temperature \( T^* \), as follows:

\[ r = k'_0 \exp \left[-\frac{E}{RT^*} \right] x_{C_2H_4} x_{H_2}^\beta \]  

(2.46)

Letting \( T' = (1/T - 1/T^*) \), \( \exp(\gamma) = k'_0 \) and \( \exp(\delta) = E/R \), equation 2.46 yields:

\[ r = \exp(\gamma) \exp(-T' \exp(\delta)) x_{C_2H_4} x_{H_2}^\beta \]  

(2.47)

The "true" value of the transformed parameters are given in table 2.33.

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>( \delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-8.8426</td>
<td>8.6726</td>
</tr>
</tbody>
</table>

Table 2.33. — Values of the transformed parameters in the ethylene hydrogenation model.

By comparing of table 2.33 and 2.32, it appears evident that this transformation reduces the differences in order of magnitude among the parameters. In the following only the transformed model 2.47 is considered with reference temperature \( T^* = 338 \) K. The experimental space is bound by the following constraints:

\[ 0.1 \leq x_{C_2H_4} \leq 0.4 \]  

(2.48a)

\[ 0.4 \leq x_{H_2} \leq 0.8 \]  

\[ 313K \leq T \leq 363K \]
and, since the concentrations are molar fractions, it also holds that:

\[ x_{\text{C}_2\text{H}_4} + x_{\text{H}_2} \leq 1.0 \] (2.48b)

This sum of molar fractions is allowed to be less than one, corresponding to the use of an inert carrier. This model was used by Agarwal and Brisk [1984a] to test a D-optimal sequential design algorithm. The simulated experiments used here have been generated using the same value \( \sigma^2 = 1.5 \times 10^{-6} \) for the variance of the measurement error as in the original paper. The settings for the independent variables in the first six experiments are also the same as in the original paper, table 2.34. This first batch of unplanned experiments yield the parameter estimate of table 2.35.

<table>
<thead>
<tr>
<th>( x_{\text{C}_2\text{H}_4} )</th>
<th>( x_{\text{H}_2} )</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.35</td>
<td>0.65</td>
<td>318.16</td>
</tr>
<tr>
<td>0.30</td>
<td>0.70</td>
<td>326.16</td>
</tr>
<tr>
<td>0.28</td>
<td>0.72</td>
<td>318.16</td>
</tr>
<tr>
<td>0.38</td>
<td>0.45</td>
<td>318.16</td>
</tr>
<tr>
<td>0.25</td>
<td>0.55</td>
<td>318.16</td>
</tr>
<tr>
<td>0.15</td>
<td>0.60</td>
<td>318.16</td>
</tr>
</tbody>
</table>

Table 2.34. — Ethylene hydrogenation. Settings for the first batch of experiments.

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>( \delta )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-10.341 \pm 2.325)</td>
<td>(8.586 \pm 0.788)</td>
<td>(0.101 \pm 0.573)</td>
<td>(0.973 \pm 0.920)</td>
</tr>
</tbody>
</table>

Table 2.35. — Estimated parameter after the first batch of six experiments.

PRODES has been used to design the seventh experiment, which follows those of table 2.34, with initial parameter estimates of table 2.35. The solution found corresponds with the one reported by Agarwal and Brisk (Table 2.36).

<table>
<thead>
<tr>
<th>( x_{\text{C}_2\text{H}_4} )</th>
<th>( x_{\text{H}_2} )</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.40</td>
<td>363</td>
</tr>
</tbody>
</table>

Table 2.36. — Ethylene hydrogenation. Settings for the seventh experiment.
Unlike in the original paper, the values of table 2.36 do not change when the error variance in the simulated experiments is increased fourfold to $\sigma^2 = 6.5 \times 10^{-6}$. In such a situation, our study shows that the parameter estimates arising from the six experiments of table 2.34 do not differ appreciably from the parameter estimates of table 2.35. The differences are mainly concentrated in the intervals of confidence rather than in the value of the estimates. PRODES does not take into account the uncertainty on parameter estimates, but only the central values. In this case, the central values do not differ enough to yield any appreciably different design.

The effect on the optimisation of the initial values of the design variables is summarized in table 2.37, which shows the results of the design of a batch of two experiments, with different initial guesses, and initial estimates of the parameters given in table 2.35.

| Initial guess | Solution | $|M|$  |
|--------------|----------|------|
| $x_{C_2H_4}$ $x_H$ T | $x_{C_2H_4}$ $x_H$ T |       |
| 0.28 0.72 333 0.10 0.80 363 | 7.74 $\times 10^{-37}$ |
| 0.10 0.40 326 0.10 0.40 363 | |
| 0.20 0.80 313 0.10 0.80 363 | 3.65 $\times 10^{-36}$ |
| 0.40 0.60 333 0.33 0.66 363 | |
| 0.40 0.60 313 0.10 0.80 363 | 7.74 $\times 10^{-37}$ |
| 0.25 0.45 329 0.10 0.40 363 | |

Table 2.37. — Ethylene hydrogenation. Effects of initial guess in designed experiments.

It is remarkable that the temperature is always at the maximum allowable value and the differences among the solutions are only given by the molar fractions. The model has much greater sensitivity to variations in the temperature than to variations in the molar fractions. This is also confirmed by inspection of the values of the Lagrange multipliers at the optimum solutions.

This observation leads to the conclusions that the suboptimal designs of table 2.37 are not as bad as the value of the objective function might suggest. Moreover, it also shows that, in this particular case, the molar fractions in model 2.45 make rather poor design variables.

Table 2.38 and 2.39 summarise the experiment carried out and the evolution of the parameter estimates and their 95% interval of confidence, respectively, for an
experiment carried out following the best batch of table 2.37. It can be noted that at the sixth batch, the number of function evaluations necessary to converge towards the solution increases considerably. The following batch cannot be designed unless the tolerance value for the optimiser is released. Similar behaviour takes place at the ninth batch. However, further release of the tolerance is not possible as the value would be too high and no reliable separation between numerical noise and real change of the objective value could be made. The design is therefore stopped.

| Experiment | Batch | $x_{C_2H_4}$ | $x_{H_2}$ | $T$ | $|M| \times 10^{35}$ | Function eval. | Tolerance |
|------------|-------|-------------|-----------|-----|----------------|----------------|------------|
| 9          | 3     | 0.2         | 0.8       | 363 |                |                |            |
| 10         | 3     | 0.2         | 0.8       | 334 | 1.23           | 15             | $10^{-4}$  |
| 11         | 4     | 0.2         | 0.8       | 363 |                |                |            |
| 12         | 4     | 0.2         | 0.8       | 342 | 1.55           | 14             | $10^{-4}$  |
| 13         | 5     | 0.2         | 0.8       | 333 |                |                |            |
| 14         | 5     | 0.2         | 0.8       | 313 | 6.02           | 14             | $10^{-4}$  |
| 15         | 6     | 0.1         | 0.4       | 363 |                |                |            |
| 16         | 6     | 0.4         | 0.4       | 363 | 14.21          | 34             | $10^{-4}$  |
| 17         | 7     | 0.4         | 0.6       | 363 |                |                |            |
| 18         | 7     | 0.2         | 0.8       | 333 | 22.53          | 12             | $10^{-3}$  |
| 19         | 8     | 0.4         | 0.6       | 361 |                |                |            |
| 20         | 8     | 0.2         | 0.8       | 331 | 60.01          | 12             | $10^{-3}$  |
| 21         | 9     | 0.4         | 0.6       | 363 |                |                |            |
| 22         | 9     | 0.4         | 0.6       | 363 | 322.59         | 29             | $10^{-3}$  |

*Table 2.38. — Ethylene hydrogenation. Experimental sequence and objective function evaluation.*

<table>
<thead>
<tr>
<th>Batch</th>
<th>$\gamma$</th>
<th>$\delta$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$-10.31 \pm 0.37$</td>
<td>$8.59 \pm 0.12$</td>
<td>$0.17 \pm 0.057$</td>
<td>$0.81 \pm 0.24$</td>
</tr>
<tr>
<td>3</td>
<td>$-10.44 \pm 0.27$</td>
<td>$8.59 \pm 0.10$</td>
<td>$0.15 \pm 0.047$</td>
<td>$0.72 \pm 0.16$</td>
</tr>
<tr>
<td>4</td>
<td>$-10.52 \pm 0.22$</td>
<td>$8.61 \pm 0.08$</td>
<td>$0.15 \pm 0.045$</td>
<td>$0.67 \pm 0.14$</td>
</tr>
<tr>
<td>5</td>
<td>$-10.48 \pm 0.21$</td>
<td>$8.59 \pm 0.07$</td>
<td>$0.15 \pm 0.045$</td>
<td>$0.69 \pm 0.14$</td>
</tr>
<tr>
<td>6</td>
<td>$-10.65 \pm 0.20$</td>
<td>$8.63 \pm 0.10$</td>
<td>$0.14 \pm 0.049$</td>
<td>$0.50 \pm 0.09$</td>
</tr>
<tr>
<td>7</td>
<td>$-10.62 \pm 0.20$</td>
<td>$8.66 \pm 0.09$</td>
<td>$0.16 \pm 0.044$</td>
<td>$0.50 \pm 0.09$</td>
</tr>
<tr>
<td>8</td>
<td>$-10.59 \pm 0.18$</td>
<td>$8.67 \pm 0.08$</td>
<td>$0.16 \pm 0.040$</td>
<td>$0.50 \pm 0.09$</td>
</tr>
<tr>
<td>9</td>
<td>$-10.61 \pm 0.16$</td>
<td>$8.66 \pm 0.07$</td>
<td>$0.16 \pm 0.035$</td>
<td>$0.49 \pm 0.07$</td>
</tr>
</tbody>
</table>

*Table 2.39. — Ethylene hydrogenation. Parameter estimates and interval of confidence evolution.*
Comparing these results with those of the original paper, two points can be remarked. The number of the objective function evaluations can be considered as a termination criterion for the sequential design procedure. However the use of an indirect optimisation algorithm allows PRODES to converge towards the solution with a number of function evaluations an order of magnitude smaller than reported by Agarwal and Brisk. Moreover, the use of a batch sequential design procedure appears to be more convenient than a purely sequential design. The batch sequential design is able to achieve similar or better results with a reduced number of optimisation runs.

The evolutions of the determinant and of the condition number of the correlation matrix are shown in figures 2.14 and 2.15. These show the expected behaviour. As in the corresponding plots of the previous section, the largest effects are concentrated in the initial runs.

![Figure 2.14. — Ethylene hydrogenation. Evolution of determinant of correlation matrix](image)
2.9 Case study: multivariate model.

The following multivariate model has been considered:

\[
\begin{align*}
  y_1 & = \frac{\theta_1 x_1}{1 + (\theta_1 - 1)x_1 + (\theta_2 - 1)x_2} & (2.49a) \\
  y_2 & = \frac{\theta_2 x_2}{1 + (\theta_1 - 1)x_1 + (\theta_2 - 1)x_2} & (2.49b)
\end{align*}
\]

The range for the independent variable is:

\[
0.1 \leq x_i \leq 1.0, \quad i = 1, 2
\]  

(2.50)

Parameter values and variances used in the simulated experiments are reported in table 2.40.

<table>
<thead>
<tr>
<th>( \theta_1 )</th>
<th>( \sigma_1^2 )</th>
<th>( \theta_1 )</th>
<th>( \sigma_1^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.740</td>
<td>0.0144</td>
<td>1.210</td>
<td>0.0625</td>
</tr>
</tbody>
</table>

Table 2.40. — Multivariate model. Expected parameter values and error variances.
A batch sequential design with batches of two experiments each is compared with a replicated $2^2$ orthogonal design. The initial batch of the sequential design is a $2^2$ orthogonal experiment. The results are shown in table 2.40 and 2.41. It is pointed out that, as already seen, the experimental points concentrate on the upper boundary of the feasible region. Moreover, at the sixth batch, there is a sudden increase of the number of function evaluations necessary to reach a solution. Even bigger is the increase in number of gradient evaluations, showing that the optimiser spends most of its time performing line searches. This indicates a rather flat objective function. Table 2.43 reports the final values of the correlation matrix, where the sequential experiments show a better result in terms of smaller interval of confidence. This problem is numerically more intensive than the previous two, however the required CPU time, on a VAXstation 3100 running VMS, is very low. This indicates the good computational efficiency of the algorithm and that the algorithm is realistically portable to much more modest hardware. The evolution of the determinant and of the condition number of the correlation matrix is displayed in figures 2.16 and 2.17.

| Experiment | Batch | $x_1$ | $x_2$ | $|M|\ |$ | Obj. eval. | Grd. eval. | CPU seconds |
|------------|-------|-------|-------|---------|-----------|------------|-------------|
| 1          | 1     | 1.00  | 1.00  | 1.00    |           |            |             |
| 2          | 1     | 1.00  | 0.10  | 0.10    |           |            |             |
| 3          | 1     | 0.10  | 0.10  | 0.10    |           |            |             |
| 4          | 1     | 0.10  | 1.00  |         |           |            |             |
| 5          | 1     | 0.50  | 0.50  |         |           |            |             |
| 6          | 2     | 1.00  | 1.00  |         |           |            |             |
| 7          | 2     | 1.00  | 1.00  | 88.014  | 3         | 3          | 0.49        |
| 8          | 3     | 1.00  | 1.00  |         |           |            |             |
| 9          | 3     | 1.00  | 1.00  | 90.369  | 4         | 4          | 0.60        |
| 10         | 4     | 1.00  | 1.00  |         |           |            |             |
| 11         | 4     | 1.00  | 1.00  | 361.990 | 4         | 4          | 0.69        |
| 12         | 5     | 1.00  | 1.00  |         |           |            |             |
| 13         | 5     | 1.00  | 1.00  | 932.070 | 4         | 4          | 0.75        |
| 14         | 6     | 0.56  | 1.00  |         |           |            |             |
| 15         | 6     | 0.56  | 1.00  | 1526.830| 13        | 45         | 3.60        |

Table 2.41. — Multivariate model. Experimental sequence and objective function evaluation.
Figure 2.16. — Multivariate model. Evolution of determinant of correlation matrix

<table>
<thead>
<tr>
<th>Batch</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$0.866 \pm 0.477$</td>
<td>$1.135 \pm 0.294$</td>
</tr>
<tr>
<td>2</td>
<td>$0.827 \pm 0.255$</td>
<td>$1.154 \pm 0.173$</td>
</tr>
<tr>
<td>3</td>
<td>$0.855 \pm 0.212$</td>
<td>$1.233 \pm 0.171$</td>
</tr>
<tr>
<td>4</td>
<td>$0.812 \pm 0.163$</td>
<td>$1.242 \pm 0.126$</td>
</tr>
<tr>
<td>5</td>
<td>$0.772 \pm 0.124$</td>
<td>$1.220 \pm 0.096$</td>
</tr>
<tr>
<td>6</td>
<td>$0.745 \pm 0.112$</td>
<td>$1.218 \pm 0.088$</td>
</tr>
<tr>
<td></td>
<td>$0.738 \pm 0.170$</td>
<td>$1.206 \pm 0.144$</td>
</tr>
</tbody>
</table>

Table 2.42. — Multivariate model. Evolution of parameter estimates and their 95% interval of confidence. The last row refers to static design.

<table>
<thead>
<tr>
<th>Run</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential</td>
<td>1.0000 1.0000</td>
</tr>
<tr>
<td></td>
<td>0.5215 1.0000</td>
</tr>
<tr>
<td>Static</td>
<td>1.0000 1.0000</td>
</tr>
<tr>
<td></td>
<td>0.5624 1.0000</td>
</tr>
</tbody>
</table>

Table 2.43. — Multivariate model. Correlation matrix at the end of different runs.
2.9.1 Nonlinearly constrained experimental space

In addition to the bounds 2.50, the following arbitrary nonlinear constraint has been imposed on the design space:

\[ x_1^2 + x_2^2 - 3x_1 - 3x_2 + 3.5 \geq 0 \]  

(2.51)

Constraint 2.51 makes the point [1.0,1.0], where many of the previous designed points were located, infeasible. The results of a sequential run comparable to the previous one are summarized in the tables 2.44 and 2.45.

<table>
<thead>
<tr>
<th>Batch</th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.866 ± 0.824</td>
<td>1.465 ± 0.329</td>
</tr>
<tr>
<td>2</td>
<td>0.827 ± 0.621</td>
<td>1.480 ± 0.429</td>
</tr>
<tr>
<td>3</td>
<td>0.855 ± 0.625</td>
<td>1.480 ± 0.420</td>
</tr>
<tr>
<td>4</td>
<td>0.812 ± 0.476</td>
<td>1.348 ± 0.313</td>
</tr>
<tr>
<td>-</td>
<td>0.772 ± 0.653</td>
<td>1.514 ± 0.481</td>
</tr>
</tbody>
</table>

Table 2.44. — Multivariate model. Constrained design space. Evolution of parameter estimates and their 95% interval of confidence. The last row refers to static design.

Figure 2.17. — Multivariate model. Evolution of condition number of correlation matrix
Table 2.45. — Multivariate model. Constrained design space. The star [*] indicates active constraints. Experimental sequence and objective function evaluation.

| Experiment | Batch | $x_1$ | $x_2$ | $|M|$ |
|------------|-------|-------|-------|------|
| 1          | 1     | 0.10  | 0.10  |      |
| 2          | 1     | 1.00  | 0.10  |      |
| 3          | 1     | 0.10  | 1.00  |      |
| 4          | 1     | 0.50  | 0.50  |      |
| 5          | 1     | 1.00  | 0.50  |      |
| 6*         | 2     | 0.63  | 1.00  |      |
| 7          | 2     | 1.00  | 1.00  | 5.34 |
| 8*         | 3     | 0.64  | 1.00  |      |
| 9          | 3     | 1.00  | 0.53  | 36.56|
| 10*        | 4     | 0.64  | 1.00  |      |
| 11         | 4     | 1.00  | 0.57  | 98.27|
| 12*        | 5     | 0.64  | 1.00  |      |
| 13         | 5     | 1.00  | 0.51  | 277.59|

It should be noted that the oddly shaped experimental region (square with a circular cut-off at the upper right) prevents the use of a factorial design for the first batch of experiments. As expected, the constraints are active in several experiments, the solution is however very far from the optimum previously chosen. The quality of the parameter estimates is reflected by the comparatively low values of the objective function. This example, although not completely satisfactory from the point of view of the parameter estimates, shows the feasibility of designing optimal experiments with a nonlinearly constrained design space.

2.10 Comments on the results and conclusion.

The results of the previous cases show the computational efficiency of the approach used. Particularly, the CPU time required by the multivariate case study, numerically the most intensive, prove that there is no need of sophisticated hardware to use this algorithm. As a test, some of the problems described in the previous section have been run on a personal computer with no observed loss of numerical precision. The computational efficiency is also shown by the much lower number
of function evaluations required in comparison to similar algorithms developed by other authors.

More important is to assess whether the design strategy proposed is efficient, reliable and suitable for real world problems. Also it needs to be established if any substantial advantage is going to be gained by its use. Figure 2.18 compares the rate constants as predicted by model 2.40 calculated with the expected values of the parameters and with the parameters arising from different experimental sequences.

![Graph of Partial pressure NO vs. Reaction rate for different parameter estimates](image)

**Figure 2.18.** Comparison among different parameter estimates for the nitric oxide reduction model. 1 - Expected. 2 - Sequential design, run B. 3 - Repeated factorial design, run F. 4 - Unplanned design, run E.

Although the sequentially designed experiments appear to give better results in terms of prediction of the true rate, such an advantage might not appear large enough to justify the extra costs of going through the sequential procedure. However, the benefits of the sequential design should be assessed mainly from a statistical point of view. If one wants good estimates, close to the true values, with small intervals of confidence, this approach is mostly beneficial. On the other hand, if the statistical property desired is a low overall level of correlation, this approach is relatively weaker.
Generally, it is observed that the biggest effect of the design is concentrated in the early stages of the experimentation. At the end it is difficult to judge whether any further improvement achieved is due to the mere addition of extra points or to the intrinsic higher information content of such points.

The number of function evaluations required for the optimum design gives an useful termination criteria for the use of this type of sequential design. One could advocate to switch to another criterion, such as the E-optimal, in order to continue the design. However, the present lack of an efficient algorithm for non-linear E-optimal design makes this idea scarcely appealing from a practical point of view.

The design depends on the value of parameter estimates, therefore is advisable to use for the very first batch a static design. As a D-optimal design with very approximate initial parameter values can be rather misleading, this has been done in most of the cases discussed.

The positive effect of parametric transformations in reducing the level of non linearity has been shown. This can be qualitatively explained by remembering that a nonlinear transformation reduces the level of non linearity in the parameter space. Such a level can be measured by a radius of curvature, as defined by Bates and Watt [1980]. However the linearisation of equation 2.7 is carried out in the parameter space as well and, by introducing a transformation, the level of approximation of equation 2.7 is reduced.

The approach presented in this work has the advantage of enabling one to handle easily any arbitrarily bounded experimental space. As shown by the example of section 2.8, this is a real requirement of many design problems. So far such constraints were mainly handled by mere rejection of inadequate points.

It has been noted how the D-optimal design criterion creates clusters of repeated experiments, usually on the upper boundary of the experimental region. This repetition of measurements at the same point is responsible for the good quality of the estimates values and for their relatively high correlation.

As predictable, the design is influenced by the independent variables with highest sensitivity. When the sensitivities have large gaps among their values, the variables with lower sensitivities make poor design variables. As a matter of future research, the effect on design of transformations acting on the sensitivities with
respect to the variables could be investigated. Such transformations are unlike the non linear ones which affect the sensitivities with respect to the parameters.

To summarise and conclude, this chapter has presented the following novel results:

1) - An original interpretation of the experiment design problem in the probabilistic Bayesian framework, which stresses the presence of formal analogies linking linear and nonlinear design theories and allow the use of equations 2.18 and 2.19 as objective functions. A critical analysis of the criteria for experiment design is also given.

2) - The first application of an efficient method, due to Bard [1974] and extended by Bates [1983] for calculating the gradients of a determinant to the case of monovariate experiment design, and a novel extension of that gradient formula to the case of multivariate models (eq. 2.37). These developments have enabled the application of an efficient indirect optimisation method to the experiment design. This, in turn, enables one to impose arbitrary nonlinear constraints on the experimental space. This particular aspect had not been previously achieved. A novel algorithm has been proposed which embeds the above developments.

3) - The development of a robust and computationally efficient implementation of the new algorithm in a general purpose computer code for the optimal design of sequential or batch sequential steady state experiments (PRODES). PRODES does not, at the moment have a match in any of the generally available routines of scientific and technical software. The use of PRODES was demonstrated on case studies with typical mono and multivariate problems.
Optimal design of dynamic experiments. A deterministic approach.

The aim of this chapter is to discuss a general extension of the sequential design techniques discussed in the previous chapter to dynamic models described by systems of mixed algebraic and differential equations. Previous work done on this area is critically analysed and discussed. In particular, alternatives to overcome some of the limitations of the methods proposed thus far are examined.

The title of the chapter refers to the approach used in this study where dynamic systems are described by nonlinear deterministic models. This differentiates the substance of this research from similar works where the models considered were stochastic and mainly linear.

Sensitivity analysis techniques play a major role in the design of dynamic experiments, and more generally in deterministic model building. Some of the results presented in this section need, in order to be properly assessed, some evaluation of the algorithm used for dynamic sensitivity analysis. Moreover, numerical aspects of the sensitivity analysis should be considered in some detail as dynamic systems may pose numerically very intensive problems. However, as sensitivity analysis is a rather large topic and it can be used as a modeling tool on its own, these issue are going to be dealt with in a subsequent independent chapter.

The design of dynamic experiments may help the experimenter in developing his/her physical insight into the system. Points where the information content is higher are pointed out to the attention of the experimenter. These points are often where important physical phenomena take place.

Parameter estimation in dynamic models is still a difficult problem which strains most the capability of most estimation algorithms. This may cause the inability to statistically assess the improvement in the parameter estimates achieved by an optimal experimental strategy. The need for a more robust and general technique for dynamic parameter estimation is pointed out.
3.1 Statement of the problem.

Let us consider a dynamic system described by a system of mixed algebraic and differential equations (DAE's):

\[ f(y'(t), y(t), x(t), u(t), w, t, \theta) = 0 \]
\[ y(t_0) = y_0 \]

where \( y(t) \) is a \( nd \) dimensional vector of differential variables, \( x(t) \) is a \( na \) dimensional vector of algebraic variables, \( u(t) \) is a \( nc \) dimensional vector of time dependent controls and \( w \) is a \( ni \) dimensional vector of constant controls.

The problems related to the integration of DAE's systems are well known and not every DAE's system is solvable [Petzold, 1982b]. Here, only systems of index one, or less are considered. These are the only ones reliably solved with the current techniques. A complete treatment of the numerical solution of DAE's systems is given by Petzold and Gear [1982] and, with reference to systems of higher order, by Gritsis [1990].

The differential variables describe time dependent states of the system such as concentration, pressure and temperature. The algebraic variables are quantities which vary instantaneously as the time dependent states evolve. One such quantity is, for example, the apparent kinetic coefficient in a polymerisation reaction, the value of which is an algebraic function of the fractional conversion. Time dependent controls are usually inputs into the system. A typical example is a feed stream into a reactor where both the flowrate and the concentration of reagents vary with time. Constant controls are similar, however their initial level at time \( t_0 \) is kept fixed during the system evolution.

The aim of a dynamic experiment is the estimation of the parameter vector \( \hat{\theta} \) by measuring the system variables, or a suitable subset, given certain initial conditions \( y_0 \) and control policy \([u(t), w]\). The measurements are taken at discrete sampling points whose number \( nsp \) and sometimes location \( t_{sp_i} \) are planned in advance by the experimenter.

As a general definition, the object of designing a dynamic experiment is to determine the control policy \([u(t), w]\), the initial conditions \( y_0 \), and the location of the sampling points \( t_{sp_i} \), so that a statistical design criterion is optimised. The
number of sampling points and maximum allowable length, $t^{\text{max}}_f$ of the experiment are chosen a priori by the experimenter.

The design criteria for dynamic experiments are defined in the probabilistic terms discussed in the first sections of chapter 2, however the different mathematical structure of model 3.1 leads to substantial differences in the implementation and usage of the criterion.

System 3.1 is a deterministic, generally nonlinear, dynamic model. The parameters $\theta$ are usually representative of specific physical phenomena and their values are therefore constrained by boundary values. As an example, kinetic coefficients are bounded to non negative values. If any relaxation of these bounds is necessary in order to obtain a reasonable reproduction of the experimental data, the model is likely to be inadequate to describe the model in mechanistic terms.

Such a qualification is important as a great deal of work has thus far been devoted to the design of experiments when the dynamic system is described by a stochastic model. In stochastic models, there is much more freedom on the values of the parameters and many, physically different, processes can be described by models having the same formal mathematical structure [Ljung, 1987]. The models considered by this approach are classified in two basic categories. The first one comprises input-output representations such as:

$$y(t) = \sum_{i=1}^{\infty} a_i y(t - i) + b_1 u(t) + e(t)$$  \hspace{1cm} (3.2)

where $y(t)$ is the vector of measured variables at time $t$, $u(t)$ is a vector of controls and $e(t)$ is an uncorrelated sequence of random variables with zero mean and $\sigma^2$ variance also called white noise. The second approach, called state space representation, describes a dynamic model as:

$$x(t + 1) = Gx(t) + Hu(t) + Fe(t)$$
$$y(t) = Bx(t) + Cu(t) + D\eta(t)$$  \hspace{1cm} (3.3)

where $x(t)$ is a vector of state variables, $y(t)$ a vector of measured variables and $e(t)$ and $\eta(t)$ are the process and the observation noises. $G,H,F,B,C$ and $D$ are matrices of appropriate dimensions. The other symbols are matrices of adequate dimensions. Equations 3.2 and 3.3 are in the time domain, however similar representations are
also used in the frequency domain. A general treatment of the design of experiments for systems described by such stochastic models is given by Titterington [1980] and Goodwin and Payne [1977]. It is evident that the description of the dynamic system such as those of equations 3.2 and 3.3 differs greatly from that of equation 3.1. The deterministic description of dynamic models requires a separate approach to the design of experiments, although there is some common terminology and some formal analogies are present.

3.2 The optimal control approach.

The problem of experiment design in steady state systems, discussed in the previous chapter, was defined as an optimisation problem. It is quite consequential to define an optimal control problem by applying the same techniques to dynamic systems such as 3.1. An optimal control problem seeks a control policy \([u^*(t), w^*]\) and initial conditions \(y_0^*\) for the system 3.1 so that an objective function:

\[
\Phi = \Psi(y(t_f), x(t_f), u(t_f), w, \theta, t_f) + \int_{t_0}^{t_f} L(y(t), x(t), u(t), w, \theta, t)dt
\]

is maximized. In optimal control terminology, the objective function \(\Phi\) is also called performance index.

The application of the design method discussed in chapter 2 requires an expression of a design criterion compatible with equation 3.4. Given model 3.1 and a set of \(n_{sp}\) sampling points, Rosenbrook and Storey [1966] have shown that an estimate of the matrix of variance covariance \(W(\hat{\theta})\) is given by:

\[
W(\hat{\theta})^{-1} = \sum_{k=1}^{n_{sp}} Q(t_k)^T G(t_k)^T \Sigma^{-1}(t_k) Q(t_k) G(t_k)
\]

\(Q(t_k)\) in equation 3.5 is the dynamic sensitivity matrix computed at time \(t_k\). Its elements are defined by:

\[
q_{ij}(t_k) = \frac{\partial y_i(t_k)}{\partial \theta_j} \quad k = 1 \ldots n_{sp} \quad i = 1 \ldots neq \quad j = 1 \ldots p
\]
where \( neq \) indicates the total number of equations in the model, either differential or algebraic. Numerical methods to compute 3.6 are described in section 4.1. When the state variables \( y \) and \( x \) of 3.1 do not correspond with the measured variables, a functional relation is also defined:

\[
z(t) = g(y(t), x(t), t, \theta)
\]

(3.7)

\( z \) being a \( na + nd \) dimensional vector. The matrix \( G(t_k) \) contains the sensitivities of the measurement \( z \) with respect to the state variables:

\[
g_{ij}(t_k) = \frac{\partial g(y, x, t_k, \theta)}{\partial y_j} \quad k = 1 \ldots nsp
\]

\[
g_{iv}(t_k) = \frac{\partial g(y, x, t_k, \theta)}{\partial y_j} \quad j = 1 \ldots nd
\]

\[
g_{iv}(t_k) = \frac{\partial g(y, x, t_k, \theta)}{\partial y_j} \quad v = nd + 1 \ldots nd + na
\]

(3.8a)

(3.8b)

If the measured variables \( z \) correspond to the vector state variables \([y^T, x^T]^T \) then \( G \) is constantly equal to an identity matrix of order \( na + nd \) and can be ignored. If any state is not measured, the relevant diagonal element of \( G \) is zero.

If the state variables coincide with the measured variables and the variance covariance matrix \( \Sigma \) is constant, equation 3.5 reduces to:

\[
W(\hat{\theta})^{-1} = C \sum_{k=1}^{nsp} Q(t_k)^T Q(t_k)
\]

(3.9)

where \( C \) is a constant matrix. Having \( \Sigma \) constant implies constant measurement error during the experiment. Under such hypotheses, equation 3.9 defines a \( D \)-optimal design criterion as follows:

\[
\max \left| \sum_{k=1}^{nsp} Q(t_k)^T Q(t_k) \right|
\]

(3.10)

Derivation of the \( E \)-optimal or \( A \)-optimal criteria is immediate. Equation 3.10 allowed Espie [1986] and Espie and Macchietto [1989] to define the design of dynamic experiments as an optimal control problem. The optimal control algorithm used here and by Espie is based on the adjoint system approach as developed and implemented in the code VOPCON by Morison [1984]. However the assumptions leading
to the criterion 3.10 are not sufficient to define an objective function suitable for VOPCON. Some extra hypotheses are necessary. These are explained below.

The objective function 3.4 is a function of state variables $y$ and $x$, however equation 3.10 is a function of the dynamic sensitivity coefficients $q_{ij}$, which are not state variables. This problem is overcome by redefining the model 3.1 as a supersystem including the state equations and the explicit sensitivity equations. These are discussed later. Now, it suffices to say that in the $(na + nd) \times (p + 1)$ dimensional supersystem, the sensitivity coefficients appear as state variables.

In order to make criterion 3.10 a suitable objective function for the optimal control problem, the summation of equation 3.10 is substituted by an integral, as follows:

$$|\sum_{k=1}^{nap} Q(t_k)^T Q(t_k)| \simeq |\int_{t_0}^{t_f} Q(t_k)^T Q(t_k) dt|$$

VOPCON does not handle directly objective functions such as 3.4 and the integral part can be dealt with only by defining a slack variable $y_s = L(y, x, u, w, \theta, t)$. Espie [1986] assumed $W(\dot{\theta})$ to be diagonal or with off-diagonal elements small enough to be negligible. The matrix $Q^T Q$ is therefore diagonal as well with elements $w_{jj} = \sum_{i=1}^{na+nd} q_{ij} q_{ij}$, $j = 1 \ldots p$. The determinant of a diagonal matrix is computed by multiplying its elements. The objective function is therefore written as:

$$\max_{y} \prod_{j=1}^{p} \int_{t_0}^{t_f} \sum_{i=1}^{na+nd} q_{ij}(t) dt$$

Since all the operators in equation 3.12 are linear, the objective function can be rewritten as:

$$\max_{y} \int_{t_0}^{t_f} \prod_{j=1}^{p} \sum_{i=1}^{na+nd} q_{ij}^2(t) dt$$

The objective function 3.13 defines a D-optimal criterion and is compatible with the requirements of VOPCON. A further simplification can be introduced by noticing that the elements of the summation are all non negative and equation 3.13 corresponds to:

$$\max_{y} \int_{t_0}^{t_f} \sum_{j=1}^{p} \sum_{i=1}^{na+nd} q_{ij}^2(t) dt$$

Due to diagonality hypothesis equations 3.13 and 3.14 define also an E-optimal criterion.
In order to ease the optimisation and to partially compensate for the lack of information due to the loss of the offdiagonal elements of $Q^TQ$, Espie suggested the use of weighting factors which yield the objective function below:

$$\max \int_{t_0}^{t_f} \sum_{j=1}^{p} \sum_{i=1}^{na+nd} w_{ij} q_{ij}(t) dt$$

(3.15)

The A-optimal criterion is defined by the minimization of the smallest of the inner summation terms and can be similarly used. In equations 3.13, 3.14 and 3.15, the argument of the integral can be easily defined as a slack variable function of the state variables $q_{ij}$ and integrated with the other equations. It should be noted that only the hypothesis of diagonality allowed Espie [1986] to swap the operators as in equation 3.12 and 3.13 and to make the objective function compatible with the requirements of the optimal control algorithm. Moreover only by accepting this hypothesis, the determinant can be easily calculated analytically. In any other case, an analytical expression for the determinant would be impossible.

Vopcon uses the control vector parameterization technique, CVP [Himmelblau and Edgar, 1988], in which time dependent controls $u(t)$ are parameterized as piecewise constant inputs. The locations of the switching points between different constant control levels are design variables. Such a representation of the controls is realistic when these are concentrations and/or flowrate of input streams. However controls such as temperature are more realistically described by piecewise linear controls. VOPCON does not allow piecewise linear control unless a slack variable is defined [Morison, 1984]. Common to other optimal control algorithms is the infeasibility to define path constraints on the controls. Presently this is a limit of nonlinear optimal control algorithms and no practical solutions are available yet. Appendix A reports a short mathematical explanation of the necessity for $Q^TQ$ to be diagonal in order to generate a suitable objective function.

The admission of diagonal or close to diagonal information matrix is a limiting factor as such an event is not common and, when it occurs, from the point of view of the parameter estimation, hardly any more significant improvement can be achieved. Another limiting factor of this approach is the substitution of the summation with an integral in equation 3.11. However necessary, such a substitution leads to the loss of any information associated with the sampling points locations. The optimal
control algorithm searches the solution in terms of final time, $t_f$, which is possibly seen as a design variable, but no consideration is made for the intermediate sampling points, unless the system is discretized in time [Bryson and Ho, 1975]. This solution, theoretically feasible, is however very cumbersome to be implemented.

The introduction of the integral in place of the summation is equivalent to considering an infinite number of sampling points. Such a condition is closely approximated by systems where the concentrations are measured by on line analytical instruments which supply virtually continuous readings, such as mass spectrometers and pH-meters. A vast class of reacting systems, however, require off line analytical techniques which, for time and economical reasons, limit the samplings to a finite, often small, number of samplings. Moreover there are systems where the location of the sampling points are the main, if not the only, available design variable.

The optimal control algorithm has the positive characteristic to maximize the absolute value of all parametric sensitivities during the system evolution. This is a welcomed and always positive fact from an experimental point of view. However, if the diagonality hypothesis does not hold, it does not suffice to define the design as optimal in the bayesians terms of chapter 2.

Recently other optimal control algorithms have been introduced in order to cope with the complexity of the nonlinear systems arising in chemical engineering. In particular Biegler [1984] and Renfro and al. [1987] have proposed algorithms which simultaneously integrate and optimise the dynamic system under investigation by means of collocation techniques [Villasden and Michelsen, 1978] which approximate the dynamic system with an appropriate system of algebraic equations. Although these approaches overcome some of the limitations of the adjoints method, they do not appear thus far able to tackle the design of dynamic experiments in any substantially better way.

3.3 A new formulation of design of dynamic experiments.

The approach developed in the previous chapter for multivariate algebraic models is extended to dynamic systems described by systems of mixed algebraic and differential equations such as 3.1. The design criterion requires the maximization of
the determinant of the information matrix. The information matrix is defined as in equation 2.19:

\[ M(\hat{\theta}, \phi) = \sum_{r=1}^{nd+na} \sum_{s=1}^{nd+na} \sigma^{rs} Q^T_r Q_s \]  (3.16)

As in the previous discussion \( \hat{\theta} \) is a vector of parameter estimates and \( \phi \) is a vector of design variables. The matrix \( Q_r \) is the matrix of the sensitivity coefficients of the \( r^{th} \) equation in the model computed at each of the \( n_{sp} \) sampling points. It is defined as follows:

\[ Q_r = \{ q_{r,ij} \} = \left\{ \frac{\partial y_r(t_{sp})}{\partial \theta_j} \right\} \]

\[ i = 1 \ldots n_{sp} \]
\[ j = 1 \ldots p \]
\[ r = 1 \ldots nd + nd \]  (3.17)

By analogy with equation 2.19 one might expect the elements of \( Q_r \) to be defined as \( \partial f_r / \partial \hat{\theta} \). However, relevant to the information matrix are the sensitivities of the measured variables, not the partial derivatives of the model equation with respect to the parameters. These are coincident in algebraic models, but not in differential models.

The elements of \( Q_r \) are computed by a sensitivity analysis. The \( Q_r \) matrices should not be confused with the dynamic sensitivity matrix, \( Q(t) \), although they have elements in common. Given a certain sampling point \( t_{sp} \), the matrix \( Q(t_{sp}) \) contains all the sensitivities of model 3.1 computed at that time while \( Q_r \) contains the sensitivities of the \( r^{th} \) state variable in the model computed at any time. However, the \( r^{th} \) row of \( Q(t_{sp}) \) coincides with the \( sp^{th} \) row of \( Q_r \). The \( \sigma^{rs} \)'s in equation 3.16 indicate the elements of the inverse variance covariance matrix \( \Sigma^{-1} \) as defined in equations 2.12 and 2.13.

Having defined the information matrix \( M \) as in equation 3.16, the design of dynamic experiments is also posed as a steady state optimisation problem. The objective function is formally the same as in equation 2.21 and its gradient can be computed by the use of equation 2.37.

The control vector \( u(t) \) is parameterised using a piecewise constant parameterisation similar to the CVP representation adopted in VOPCON and other optimal control algorithms:

\[ u_i(t) = z_{i,j} \]
\[ i = 1 \ldots nc \]
\[ t_{sw_i,j-1} \leq t \leq t_{sw_i,j} \]
\[ j = 1 \ldots nsw_i \]  (3.18)

82
where nc is the number of controls and nswi is the number of switching points associated to the i\textsuperscript{th} control. Both the zi,j's and the t\textsubscript{swi, j} are design variables. The most general vector of design variables φ is:

\[\phi^T = [t_{T}^{T}, t_{T}^{T}, \ldots, t_{T}^{T}, t_{nc}^{T}, z_{1}^{T}, \ldots, z_{nc}^{T}, y_{0}^{T}, w^{T}]\] (3.19)

where w is a vector of constant inputs and y\textsubscript{0} are the initial conditions. Each control variable may be discretised over a different number of intervals and different control levels may be switched at different times. The sampling points are however the same for all the state variables. Constraints are allowed on the design variables. There are simple bounds and optional nonlinear inequality/equality constraints on the initial conditions:

\[y_{0i}^{L} \leq y_{0i} \leq y_{0i}^{U} \quad i = 1 \ldots nd\] (3.20a)

\[h(y_{0}) \geq 0\] (3.20b)

The nonlinear constraints 3.20b may be necessary to account for stoichiometric relations among reagents and other similar physical bounds. Simple bounds are imposed on the constant inputs:

\[w_{i}^{L} \leq w_{i} \leq w_{i}^{U} \quad i = 1 \ldots ni\] (3.21)

The maximum allowable final time \(t_{f}^{maz}\) is chosen by the experimenter. Linear constraints are imposed on the switching points and the sampling points as it follows:

\[\delta t_{sp}^{max} \leq t_{sp} - t_{sp-1} \leq \delta t_{sp}^{min} \quad i = 1 \ldots nsp\] (3.22)

where \(\delta t_{sp}^{max}\) and \(\delta t_{sp}^{min}\) are the maximum and the minimum allowable intervals between consecutive sampling points. In order to allow the maximum flexibility independent limits are allowed for each interval between consecutive sampling points. Proper setting of such independent limits would also allow one to keep a certain sampling point fixed at a desired values. This may be convenient when planning a long experimental run in an industrial environment in order to take into account operator's availability, change of shifts, etc.
Analogous considerations lead to the following linear constraints on the interval between adjacent switching points:

\[
\beta t_{swi,j}^{max} \leq t_{swi,j} - t_{swi,j-1} \leq \beta t_{swi,j}^{min} \quad i = 1 \ldots nc \\
\beta \quad j = 1 \ldots nsw
\]  

(3.23)

Simple bounds are also imposed on the level of the constant controls inside each control interval:

\[
z_{i,j}^L \leq z_{i,j} \leq z_{i,j}^U \quad i = 1 \ldots nc \\
\beta \quad j = 1 \ldots nsw
\]  

(3.24)

Finally, the time of the last sampling point should be less or equal to the maximum allowable final time and the last switching point of each control should coincide with the last sampling point and the end of the experiment.

\[
t_{sp_{nsp}} \leq t_f^{max} \quad i = 1 \ldots nc \\
t_{nsw_i} = t_{sp_{nsp}}
\]  

(3.25)

It is to be noticed that unlike in the CVP representation, a switching point is associated with the preceding control interval. In such a way the initial time \( t_0 \) is not considered, but, in order to have a consistent description of the system, the
The final switching point of each control interval has to coincide with the last sampling point. This, in turn, marks the end of the experiment.

The above formulation defines, very comprehensively and flexibly, a large range of dynamic experimental conditions, and is therefore very attractive. However, even from a superficial analysis, two major drawbacks emerge in this approach. The optimisation problems are quite large also for small dynamic systems. Let us consider a dynamic system with two state variables, ten sampling points, variable initial conditions and one constant input. To design the experiments in this system, one should consider thirteen design variables subject to eleven nonlinear constraints plus the simple bounds. The possible introduction of variable inputs with the relevant switching points may generate large scale optimisation problems, even for small dynamic models. However, the problem related to the size of the optimisation problems are correctly tackled using a code such as SRQP [Chen, 1988], which has been designed for the large scale problems arising in flowsheeting.

More serious is the fact that the computation of the derivatives $\partial m_{ij}/\partial \phi_k$ of equation 2.37 by finite differences requires a complete integration of the dynamic system for each of the design variables. This may generate serious problems of numerical accuracy and numerical noise. A more detailed discussion of these aspects is carried out in the next chapter.

3.4 An algorithm for the design of dynamic experiments.

From the discussion in the previous chapters, it is possible to formalise an algorithm for the design of dynamic experiments. The minimum set of design variables is made up by the sampling points, $t_{sp_i}$, the number of which, $n_{sp}$, is chosen by the experimenter. It is mandatory to supply a value $t_f^{\text{max}}$ for the maximum allowable final time. Such a time is tied to the last sampling point by either an equality or inequality constraints (equation 3.25). The steps in the algorithm are:

1 – Supply an initial guess for the design variables: these are the sampling points and optionally initial conditions and parameterised controls.

2 – Supply consistent values for the intervals $\delta t_{sp_i}^{\text{max}}$ and $\delta t_{sp_i}^{\text{min}}$. If applicable, values for $\delta t_{sw_i}^{\text{max}}$, $\delta t_{sw_i}^{\text{min}}$, and any other upper and lower bounds of design variables should also be supplied.
3 - If necessary, any constraints such as 3.20b should be supplied along with the relevant gradients.

4 - Supply an estimate of the variance covariance matrix \( \Sigma^{-1} \). If not available, a default value should be available in the code.

5 - Integrate the system from \( t_0 \) to \( t_f \), storing at each sampling points the sensitivity coefficients. If switching points for variable controls are considered, then the integration is interrupted at the switching points and restarted after re-initialisation with the appropriate values of the controls.

6 - Retrieve the sensitivity coefficients to compute the \( Q_i \)'s (equation 3.17).

7 - Compute the information matrix as in equation 3.16.

8 - Compute the objective function (equation 2.21) and the inverse, by singular value decomposition, of the information matrix.

9 - Compute, where applicable, the constraints 3.20b, 3.22, 3.23, 3.24 and 3.25 with the relevant gradients.

10 - Compute the gradients of the objective function, equation 2.37. The computation of the \( \partial m_{ij} / \partial \theta_j \) is carried out by finite differences over the elements of the information matrix. This involves the perturbation, one at a time, of the design variables and the repetition of steps 5, 6 and 7. Neither the determinant nor the inverse of the information matrix need to be recomputed. Details on the numerical derivatives scheme used are reported in Appendix D.

11 - Call the optimiser.

12 - On exit from the optimiser if no termination criteria are satisfied, then repeat from step 5 onwards.

Step 10 is the most critical one for the design. Most of the computer time required is spent in the calculation of the \( \partial m_{ij} / \partial \theta_j \).

The algorithm described was implemented in a computer code and the results are presented in section 3.6 and following ones.
3.5 Case study: a simple kinetic model.

The following dynamic model describing the reaction \( A \rightarrow B \rightarrow C \) was used to test the dynamic design procedure:

\[
\begin{align*}
\dot{c}_A &= -\exp(\theta_1) c_A \\
\dot{c}_B &= \exp(\theta_1) c_A - \exp(\theta_2) c_B \\
\dot{c}_C &= \exp(\theta_2) c_B
\end{align*}
\]  
(3.26)

In this case study only the optimisation of the sampling points is considered with the objective of improving the estimates of the two parameters \( \theta_1 \) and \( \theta_2 \). Simulated experiments have been carried out with two levels of experimental error. The cases have been considered where all the state variables or only the concentrations \( c_A \) and \( c_B \) are measured. Each run consists of four successive experiments. Initial conditions for each experiment were \( c_A = 1.0 \) and \( c_B = c_C = 0.0 \) (in this example arbitrary units are used).

<table>
<thead>
<tr>
<th>Run</th>
<th>( \sigma^2 )</th>
<th>Variable</th>
<th>Measured variable</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.015</td>
<td>identity</td>
<td>( c_A, c_B, c_C )</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.015</td>
<td>estimated</td>
<td>( c_A, c_B, c_C )</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.015</td>
<td>n/a</td>
<td>( c_A, c_B, c_C )</td>
<td>repeated unplanned</td>
</tr>
<tr>
<td>D</td>
<td>0.015</td>
<td>identity</td>
<td>( c_A, c_B )</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>0.015</td>
<td>estimated</td>
<td>( c_A, c_B )</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.015</td>
<td>n/a</td>
<td>( c_A, c_B )</td>
<td>repeated unplanned</td>
</tr>
<tr>
<td>G</td>
<td>0.030</td>
<td>identity</td>
<td>( c_A, c_B, c_C )</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>0.030</td>
<td>estimated</td>
<td>( c_A, c_B, c_C )</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>0.030</td>
<td>n/a</td>
<td>( c_A, c_B, c_C )</td>
<td>repeated unplanned</td>
</tr>
<tr>
<td>L</td>
<td>0.030</td>
<td>identity</td>
<td>( c_A, c_B )</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0.030</td>
<td>estimated</td>
<td>( c_A, c_B )</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>0.030</td>
<td>n/a</td>
<td>( c_A, c_B )</td>
<td>repeated unplanned</td>
</tr>
</tbody>
</table>

Table 3.1. — Summary of experiments for the dynamic model of the two consecutive reaction

Experimental points were simulated by adding to the exact state trajectories a normally distributed error term with variance as reported in table 3.1 which summarises the experiments. The integration of the system was carried out using the exact parameter values \( \theta_1 = -1.572 \) and \( \theta_2 = -0.702 \). The results of each planned
run are compared with those of a similar unplanned run. The first experiment in
the planned runs has also been designed. Initial guesses for the parameter esti-
mates were $\theta_1 = -1.00$ and $\theta_2 = -0.50$. Fourteen sampling points were set with
$\delta tsp^{max} = 2.00$ hours and $\delta tsp^{min} = 0.25$ hours. These were kept constant for each
sampling interval. Maximum final time is 7.0 hrs. Unplanned experiments were
carried out by taking a measurement every half an hour. Runs have also been re-
peated using the prior information $\Sigma$ or not. In this second case the identity matrix
has been used as a default. This introduces an approximation as the information
matrix becomes
\[ M = \sum_{r=1}^{na+nd} Q_r^T Q_r \]  
(3.27)
The effects of such an approximation are discussed below. The two levels of ex-
perimental error generate rather scattered measurements as shown by figures 3.2
and 3.3, where typical simulated experimental points are plotted along with the
expected state trajectories (continuous lines).

![Figure 3.2. — Typical experimental points with $\sigma^2 = 0.015$. Equally spaced sampling points. Expected state
trajectories shown as continuous lines.](image)

88
Figure 3.3. — Typical experimental points with $\sigma^2 = 0.030$. Same as figure 3.2.

<table>
<thead>
<tr>
<th>Time (hours)</th>
<th>run G</th>
<th>run H</th>
<th>run L</th>
<th>run M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2/3/4*</td>
<td>1/2/3/4*</td>
<td>1/2/3/4*</td>
<td>1/2/3/4*</td>
<td>1/2/3/4*</td>
</tr>
<tr>
<td>2.00</td>
<td>2.00</td>
<td>0.15</td>
<td>1.02</td>
<td>0.90</td>
</tr>
<tr>
<td>4.00</td>
<td>4.00</td>
<td>1.40</td>
<td>1.27</td>
<td>1.22</td>
</tr>
<tr>
<td>4.25</td>
<td>4.25</td>
<td>1.65</td>
<td>1.52</td>
<td>1.47</td>
</tr>
<tr>
<td>4.50</td>
<td>4.50</td>
<td>1.90</td>
<td>2.75</td>
<td>2.75</td>
</tr>
<tr>
<td>4.75</td>
<td>4.75</td>
<td>2.15</td>
<td>4.75</td>
<td>4.75</td>
</tr>
<tr>
<td>5.00</td>
<td>5.00</td>
<td>3.25</td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>5.25</td>
<td>5.25</td>
<td>4.30</td>
<td>5.25</td>
<td>5.25</td>
</tr>
<tr>
<td>5.50</td>
<td>5.50</td>
<td>5.50</td>
<td>5.50</td>
<td>5.50</td>
</tr>
<tr>
<td>5.75</td>
<td>5.75</td>
<td>5.75</td>
<td>5.75</td>
<td>5.75</td>
</tr>
<tr>
<td>6.00</td>
<td>6.00</td>
<td>6.00</td>
<td>6.00</td>
<td>6.00</td>
</tr>
<tr>
<td>6.50</td>
<td>6.50</td>
<td>6.50</td>
<td>6.50</td>
<td>6.50</td>
</tr>
<tr>
<td>6.75</td>
<td>6.75</td>
<td>6.75</td>
<td>6.75</td>
<td>6.75</td>
</tr>
<tr>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
</tr>
</tbody>
</table>

Table 3.2. — Optimal sampling points for the two consecutive reactions case. The star (*) indicates that there is no appreciable difference among the different designed experiments.
Table 3.2 reports the values of the sampling points in runs G, H and L, M. These are representative also of the runs with lower error level, as the error level does not affect the optimum. The selected sampling points tend to be reproduced quite similarly among analogous runs, although slight variations in the position of the optimum exist. It is remarkable to note that the length of the experiments with two measured variables is limited to 5.30 hrs. The design procedure has moved the final time and suggested a much shorter experiment. In comparing the results it should, therefore, be borne in mind that the planned experiments, although comprising the same number of samplings, are shorter.

The computational load for the optimal designs of the runs of table 3.2 is summarized in table 3.3. There is no clear difference between the run using the prior information $\Sigma$ and those where default values (an identity matrix) are used. It is clear that most of the time is spent in the repeated integrations of the dynamic system necessary to compute the gradients of the objective function.

<table>
<thead>
<tr>
<th>Run</th>
<th>Experiment</th>
<th>CPU sec.</th>
<th>Objective eval.</th>
<th>Gradient eval.</th>
<th>Integrations</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>1</td>
<td>39.46</td>
<td>5</td>
<td>5</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>2/3/4</td>
<td>58</td>
<td>9</td>
<td>9</td>
<td>153</td>
</tr>
<tr>
<td>H</td>
<td>1/2</td>
<td>32</td>
<td>4</td>
<td>4</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>3/4</td>
<td>64</td>
<td>10</td>
<td>10</td>
<td>170</td>
</tr>
<tr>
<td>L</td>
<td>1</td>
<td>52.72</td>
<td>8</td>
<td>8</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>2/3/4</td>
<td>58</td>
<td>10</td>
<td>10</td>
<td>170</td>
</tr>
<tr>
<td>M</td>
<td>1/2/3/4</td>
<td>38</td>
<td>6</td>
<td>6</td>
<td>102</td>
</tr>
</tbody>
</table>

Table 3.3. — Computational load of the design of table 3.5. The CPU times refer to a VAXstation 3100 running VMS.

The matrix $\Sigma^{-1}$ has a scaling effect on the optimisation and the location of the optimum does not change when $\Sigma$ is multiplied by a scalar. This is most convenient as it very easily prevents the numerical noise from overriding the changes in a relatively low and flat objective function. As an example for the second experiment of run M the estimate for $\Sigma^{-1}$ was:

$$
\Sigma^{-1} = \begin{pmatrix}
12.49 \\
14.72 & 31.89 \\
9.37 & 4.28 & 7.83
\end{pmatrix} \times 10^2
$$
The use of such a value caused overflow. Scaling $\Sigma^{-1}$ by $10^{-2}$ fixed the problem, but other scaling factors have been effective as well.

The results of all the experimental runs are reported in the following tables. The terms in the offdiagonal columns of the tables refer to the only offdiagonal element of the variance covariance matrix of parameter estimates.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.581 \pm 0.028$</td>
<td>$-0.695 \pm 0.035$</td>
<td>$-0.1228$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.576 \pm 0.014$</td>
<td>$-0.697 \pm 0.021$</td>
<td>$-0.2249$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.575 \pm 0.013$</td>
<td>$-0.703 \pm 0.017$</td>
<td>$-0.2627$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.577 \pm 0.009$</td>
<td>$-0.705 \pm 0.014$</td>
<td>$-0.2495$</td>
</tr>
</tbody>
</table>

Table 3.4. — Run A. Low error level, 3 measured variables, no prior information, evolution of parameter estimates and correlation

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.581 \pm 0.028$</td>
<td>$-0.695 \pm 0.035$</td>
<td>$-0.1228$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.584 \pm 0.012$</td>
<td>$-0.688 \pm 0.017$</td>
<td>$-0.3916$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.583 \pm 0.012$</td>
<td>$-0.685 \pm 0.020$</td>
<td>$-0.3996$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.580 \pm 0.010$</td>
<td>$-0.688 \pm 0.019$</td>
<td>$-0.3276$</td>
</tr>
</tbody>
</table>

Table 3.5. — Run B. Low error level, 3 measured variables, prior information, evolution of parameter estimates and correlation

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.582 \pm 0.016$</td>
<td>$-0.699 \pm 0.032$</td>
<td>$-0.4547$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.587 \pm 0.014$</td>
<td>$-0.699 \pm 0.027$</td>
<td>$-0.1855$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.583 \pm 0.010$</td>
<td>$-0.689 \pm 0.024$</td>
<td>$-0.1042$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.577 \pm 0.009$</td>
<td>$-0.696 \pm 0.019$</td>
<td>$-0.1735$</td>
</tr>
</tbody>
</table>

Table 3.6. — Run C. Low error level, 3 measured variables, unplanned runs, evolution of parameter estimates and correlation
<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.587 \pm 0.022$</td>
<td>$-0.702 \pm 0.040$</td>
<td>$-0.1194$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.581 \pm 0.016$</td>
<td>$-0.715 \pm 0.028$</td>
<td>$-0.1853$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.580 \pm 0.013$</td>
<td>$-0.718 \pm 0.023$</td>
<td>$-0.2157$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.577 \pm 0.012$</td>
<td>$-0.716 \pm 0.023$</td>
<td>$-0.3967$</td>
</tr>
</tbody>
</table>

Table 3.7. — Run D. Low error level, 2 measured variables, no prior information, evolution of parameter estimates and correlation

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.583 \pm 0.022$</td>
<td>$-0.703 \pm 0.040$</td>
<td>$-0.1194$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.580 \pm 0.015$</td>
<td>$-0.708 \pm 0.032$</td>
<td>$-0.2190$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.583 \pm 0.011$</td>
<td>$-0.705 \pm 0.026$</td>
<td>$-0.2293$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.585 \pm 0.016$</td>
<td>$-0.706 \pm 0.025$</td>
<td>$-0.2193$</td>
</tr>
</tbody>
</table>

Table 3.8. — Run E. Low error level, 2 measured variables, prior information, evolution of parameter estimates and correlation

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.582 \pm 0.027$</td>
<td>$-0.733 \pm 0.040$</td>
<td>$-0.3963$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.579 \pm 0.016$</td>
<td>$-0.734 \pm 0.035$</td>
<td>$-0.2918$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.576 \pm 0.015$</td>
<td>$-0.718 \pm 0.029$</td>
<td>$-0.4220$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.576 \pm 0.012$</td>
<td>$-0.712 \pm 0.021$</td>
<td>$-0.3872$</td>
</tr>
</tbody>
</table>

Table 3.9. — Run F. Low error level, 2 measured variables, unplanned experiments, evolution of parameter estimates and correlation

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.563 \pm 0.042$</td>
<td>$-0.707 \pm 0.064$</td>
<td>$-0.0245$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.572 \pm 0.028$</td>
<td>$-0.717 \pm 0.043$</td>
<td>$-0.1651$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.567 \pm 0.021$</td>
<td>$-0.713 \pm 0.034$</td>
<td>$-0.1839$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.572 \pm 0.017$</td>
<td>$-0.702 \pm 0.029$</td>
<td>$-0.2384$</td>
</tr>
</tbody>
</table>

Table 3.10. — Run G. High error level, 3 measured variables, no prior information, evolution of parameter estimates and correlation
<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.563 \pm 0.042$</td>
<td>$-0.707 \pm 0.064$</td>
<td>$-0.0245$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.572 \pm 0.035$</td>
<td>$-0.685 \pm 0.051$</td>
<td>$-0.2885$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.569 \pm 0.020$</td>
<td>$-0.696 \pm 0.038$</td>
<td>$-0.2218$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.569 \pm 0.020$</td>
<td>$-0.696 \pm 0.038$</td>
<td>$-0.2286$</td>
</tr>
</tbody>
</table>

Table 3.11. — Run H. High error level, 3 measured variables, prior information, evolution of parameter estimates and correlation

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.592 \pm 0.032$</td>
<td>$-0.696 \pm 0.064$</td>
<td>$-0.4549$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.602 \pm 0.027$</td>
<td>$-0.688 \pm 0.056$</td>
<td>$-0.1874$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.595 \pm 0.021$</td>
<td>$-0.670 \pm 0.050$</td>
<td>$-0.1060$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.582 \pm 0.019$</td>
<td>$-0.690 \pm 0.039$</td>
<td>$-0.1743$</td>
</tr>
</tbody>
</table>

Table 3.12. — Run I. High error level, 3 measured variables, unplanned runs, evolution of parameter estimates and correlation

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.599 \pm 0.042$</td>
<td>$-0.704 \pm 0.074$</td>
<td>$-0.0647$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.591 \pm 0.032$</td>
<td>$-0.727 \pm 0.058$</td>
<td>$-0.2832$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.588 \pm 0.026$</td>
<td>$-0.733 \pm 0.048$</td>
<td>$-0.2821$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.583 \pm 0.024$</td>
<td>$-0.702 \pm 0.046$</td>
<td>$-0.4415$</td>
</tr>
</tbody>
</table>

Table 3.13. — Run L. High error level, 2 measured variables, no prior information, evolution of parameter estimates and correlation

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.599 \pm 0.042$</td>
<td>$-0.704 \pm 0.074$</td>
<td>$-0.0647$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.592 \pm 0.029$</td>
<td>$-0.710 \pm 0.064$</td>
<td>$-0.3007$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.571 \pm 0.031$</td>
<td>$-0.704 \pm 0.048$</td>
<td>$-0.3129$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.571 \pm 0.025$</td>
<td>$-0.701 \pm 0.047$</td>
<td>$-0.3049$</td>
</tr>
</tbody>
</table>

Table 3.14. — Run M. High error level, 2 measured variables, prior information, evolution of parameter estimates and correlation
<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>Offdiagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-1.538 \pm 0.054$</td>
<td>$-0.677 \pm 0.084$</td>
<td>$-0.4057$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.587 \pm 0.031$</td>
<td>$-0.767 \pm 0.061$</td>
<td>$-0.2999$</td>
</tr>
<tr>
<td>3</td>
<td>$-1.580 \pm 0.029$</td>
<td>$-0.734 \pm 0.059$</td>
<td>$-0.4248$</td>
</tr>
<tr>
<td>4</td>
<td>$-1.579 \pm 0.025$</td>
<td>$-0.725 \pm 0.050$</td>
<td>$-0.3773$</td>
</tr>
</tbody>
</table>

Table 3.15. — Run N. High error level, 2 measured variables, unplanned experiments, evolution of parameter estimates and correlation

From the results above, it appears the situation which benefits most from the design is when the error level is highest and only two variables are measured (runs L and M). The gains in information, due to designed experiments, are inversely proportional to the intrinsic content of information of the experiment itself. The results of these runs, compared with similar unplanned runs, are better both in terms of smaller interval of confidence and lower correlation. They are comparable with the results of the runs with lower error level. However, the situation where all the variables are measured and the error level is lower, does not show any decisive improvement when the experiments are designed.

Figure 3.4. — Simple kinetic model. Evolution of the 95% interval of confidence of the parameter $\theta_1$ in different runs.
The optimisation runs necessary to design these experiments have been carried out quite easily. Occasionally, however, slow convergence due to flat objective function has been noted. This has been solved by relaxing the tolerance for the optimiser which in each of the runs mentioned above had been less or equal to $10^{-3}$. The evolution of the interval of confidence of the parameter estimates and of determinant and correlation matrix during various runs is illustrated in figures 3.4 to 3.6.

3.5.1 Use of initial conditions and constant inputs as design variables.

Using the same model of section 3.5, the use of variable initial conditions has been investigated. Two cases have been considered. In the first one, only the initial concentration of the reagent A has been allowed to vary. It has been observed that, whatever the upper bound for such an initial condition, this is reached in the first call to the optimiser. No further changes are recorded in the following calls to the optimiser. The selection of the sampling points was not appreciably affected and the optimal solutions found were comparable with those of the previous section. As a
consequence, the results of the parameter estimation were not remarkably different from those of the previous section.

Similar results are also obtained when all initial conditions are allowed to vary within fixed boundaries. A non zero concentration of products at the initial time is equivalent to shifting the initial time. This is consistent with the results of table 3.5 where the first sampling point is always quite distant from the initial time. The values of the initial conditions were always set to the upper bound after the first call to the optimiser. This, in turn, leads to shorter experiments.

From the results above, it may appear quite pointless to include the initial conditions as design variables, since they seem invariably to be set to the highest possible level. This could be done by the experimenter, avoiding the expensive burden of adding independent variables in the optimisation. However, although such a conclusion is true for this kind of simple kinetic systems, it cannot be inferred as a rule for other dynamic models. The provision of variable initial condition should be kept as a feature of any algorithm for the design of dynamic experiments.

In kinetic experiments it is quite natural to advocate the use of temperature as a design variable. Although for practical reasons non isothermal experiments should be ruled out, if not strictly necessary, one would like to select an optimal
value of the temperature for an isothermal experiment. This has been investigated by rewriting model 3.26 as:

\[ c'_A = -k_1 \exp(-E_1/T)c_A \]
\[ c'_B = k_1 \exp(-E_1/T)c_A - k_2 \exp(-E_2/T)c_B \]
\[ c'_C = k_2 \exp(-E_2/T)c_B \]  
(3.28)

with the parameter values reported in the following table 3.16.

<table>
<thead>
<tr>
<th>( E_1 )</th>
<th>( k_1 )</th>
<th>( E_2 )</th>
<th>( k_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2875.14</td>
<td>1162.011</td>
<td>2344.16</td>
<td>563.464</td>
</tr>
</tbody>
</table>

Table 3.16. — Simple kinetic model. Parameter values used in model 3.28. Units are sec\(^{-1}\) for \( k_1 \) and \( k_2 \), J/K for \( E_1 \) and \( E_2 \).

The parameter values of the previous section correspond to the temperature 60\(^{0}\)C. Bounds were fixed on the feasible temperature as:

\[ 50^0 \leq T \leq 120^0 \]  
(3.29)

Prior information was used to design the experiments. The normally distributed error for the simulated measurements had variance \( \sigma^2 = 0.030 \). Other constraints and experimental conditions were equal to those used in the previous section. Depending on the number of measured variables, three \((c_A, c_B, c_C)\) or two \((c_A, c_B)\), the designed optimal sampling points are reported in the following table 3.17:

<table>
<thead>
<tr>
<th>Measured variables</th>
<th>Optimal sampling times</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_A, c_B, c_C )</td>
<td>0.42 1.15 1.40 1.83 2.51 2.90</td>
</tr>
<tr>
<td></td>
<td>3.46 3.97 4.49 5.00 5.49 5.59</td>
</tr>
<tr>
<td></td>
<td>6.20 7.00</td>
</tr>
<tr>
<td>( c_A, c_B )</td>
<td>2.00 2.35 2.60 2.85 3.10 3.35</td>
</tr>
<tr>
<td></td>
<td>3.60 3.85 4.10 4.35 4.60 4.85</td>
</tr>
<tr>
<td></td>
<td>5.10 5.65</td>
</tr>
</tbody>
</table>

Table 3.17. — Simple kinetic model: optimal sampling points with variable temperature.

In both cases the optimal temperature was about 85\(^{0}\)C. Such an increase in temperature, compared to the cases of the previous section, was expected. The
temperature is however quite distant from the upper boundary. This is intuitively explained since a very high temperature would cause the system to reach steady state conditions very quickly, and there would not even be the time for all the fourteen samples to be taken.

Although the numerical effort necessary for these designs was comparable with those reported in the previous section, these have been characterised by low and relatively flat objective function. A scaling factor $\lambda = 10^3$ was used in order to overcome the numerical noise generated during the gradient calculation.

It has been impossible to assess the effectiveness of the design procedure on the parameter estimates as the available estimation code proved to be inadequate to carry out the necessary statistical analysis on the parameter estimates.

3.6 Case study: optimal selection of sampling points in a fed-batch fermentation reactor.

A model describing a fermentation reaction in a fed-batch bioreactor has been considered:

$$x_1' = (x_3 - u_1)x_1 - \theta_4 x_1$$
$$x_2' = -\frac{x_3x_1}{\theta_3} + u_1(u_1 - x_2)$$
$$x_3 = \frac{\theta_1 x_1}{\theta_2 + x_2}$$

(3.30)

where $x_1$ and $x_2$ are the biomass and substrate concentrations (gr/lt), $x_3$ is the Michaelis-Menten rate constant, $u_1$ is the dilution factor in the feed, $u_2$ is the substrate concentration in the feed and the $\theta$'s are kinetic parameters. The dilution factor is defined as the ratio between the reacting volume in the reactor and the inlet flowrate. The model was developed by Nikila and Virkunen [1977]. In the following example the dilution factor has been set to 0.20 hours$^{-1}$ and the substrate concentration in the feed is 20.00 gr/lt.

This model was used by Espie [1986] to design experiments using the optimal control approach. Unlike Espie's, the present implementation of the algorithm proposed in the previous sections is not able to handle variable inputs $u_1$ and $u_2$ as severe problems of numerical noise prevented convergence during the optimisation.

The computation of the partial derivatives of the elements of the information matrix with respect to the design variables is quite easy if the design variables are...
the sampling points. However such a calculation is more difficult and it generates more noise if the variables of interest are the control levels. The sensitivities are rather less affected by small variations in the controls than by similar variations in the sampling points. Such behaviour is the cause of scaling problems in computing, by finite differencing, the partial derivatives of the elements of the information matrix.

Tests of the optimal selection of sampling points have been carried out using the vector of parameter estimates $\hat{\theta} = [0.25, 0.20, 0.50, 0.03]^T$. This value is relatively close to the expected one $\theta = [0.31, 0.18, 0.55, 0.05]^T$. Two optimal experiments were designed with twenty sampling points each. The maximum distance between consecutive sampling points was set to 10 hours while the minimum was set to 0.25 hours. The maximum final time was set at 50 hours. Two cases, with fixed or variable final time, were examined. The results are in the following table 3.18.

<table>
<thead>
<tr>
<th>Final time</th>
<th>Optimal</th>
<th>sampling</th>
<th>times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>1.870</td>
<td>11.87</td>
<td>20.57</td>
</tr>
<tr>
<td></td>
<td>36.16</td>
<td>36.53</td>
<td>41.73</td>
</tr>
<tr>
<td></td>
<td>47.19</td>
<td>48.02</td>
<td>48.15</td>
</tr>
<tr>
<td></td>
<td>47.75</td>
<td>50.00</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>1.300</td>
<td>10.50</td>
<td>20.50</td>
</tr>
<tr>
<td></td>
<td>32.19</td>
<td>32.53</td>
<td>37.73</td>
</tr>
<tr>
<td></td>
<td>43.19</td>
<td>44.02</td>
<td>44.75</td>
</tr>
<tr>
<td></td>
<td>45.75</td>
<td>46.00</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.18. — Fed-batch fermentation reactor. Optimal sampling points with fixed and variable final time.

In the second case (variable final time) there is an evident shift of about 4 hrs. for the last points. Figure 3.7 shows the above sampling points and the state trajectories.
In the case with variable final time, the sampling points were in good agreement with the results by Kalogerakis and Luus [1984]. However, these authors do not use a statistical criterion to generate an objective function to be used in some design algorithm. They define a sensitivity index as:

\[ s(t) = \theta^T Q(t)^T \Sigma^{-1} Q(t) \dot{\theta} \]  

(3.43)

\( s(t) \) is a time dependent \( p \) dimensional vector, the \( j^{th} \) element of which gives a measure of the overall sensitivity of the state variables with respect to the parameter \( \dot{\theta}_j \). Their sensitivity index is used to identify intervals of higher sensitivity in which the sampling points should be located. The actual discrete sampling points are found by overimposing on such intervals some distribution of sampling points which has been established a priori.

The two designs have proved to be computationally extremely expensive as shown in table 3.19, where the CPU times refers to a VAXstation 3100 running VMS.
Experiment | CPU hours | system integ. | objective eval. | gradients eval.
--- | --- | --- | --- | ---
Fixed final time | 4.67 | 795 | 33 | 102
Variable final time | 2.02 | 331 | -- | --

Table 3.19. — Selection of optimal sampling points in a fermentation reactor. Computational requirements.

The fixed time case is particularly expensive. The comparatively high number of gradient evaluations suggest that most of the time is spent in line searches while the optimiser is struggling with relatively flat objective functions.

3.7 Conclusion and discussion.

The cases presented in the previous sections allow an assessment of the proposed algorithm for the design of dynamic experiments. The case studies of section 3.5 show that the approach is able to improve parameter estimation in those particular experimental situations where the intrinsic level of information is limited.

Two elements of novelty should be pointed out in the approach presented. The first one is the independent selection of discrete sampling points. The second is the ability to tackle problems regardless of any hypothesis on the nature of the variance covariance matrix of the parameter estimates. The algorithm is therefore, at least from a theoretical point of view, capable of addressing real world problems. Experiments such as those of section 3.5 could not have been designed using the optimal control approach.

The advantages of the algorithm presented in this chapter should not be hidden by the relative simplicity of the cases presented or by the intuitive results. The good physical sense of the results of sections 3.5 and following ones enables one to trust the method also for those systems where such an immediate physical insight is not possible.

In section 3.5.1 the use of the initial conditions as design variables is discussed. When the initial condition \( c_{A0} \) of system 3.26 is left to vary freely, it is found that it is always set to the upper bound. The argument that this generates the required maximum excitation of the system, proposed in section 3.5.1, can be misleading without further comment. It can be objected that the solution of system 3.26 depends on the ratio \( c_A/c_{A0} \) and is therefore independent from the value of \( c_{A0} \) and that since the error on the computer generated experimental data has constant variance, the increase in \( c_{A0} \) just reflects an increase in the relative accuracy of the measurements. This is accurate and certainly true. However, the design algorithm does not have any information on the error structure and it would give the same design also when the variance of the experimental error is a constant percentage of the quantity being measured.

Although in the case of constant variance, maximising the initial concentration of \( A \) is desirable, an important effect of the increase of \( c_{A0} \) is to delay the time when the system reaches steady state.

Similar conclusions can be drawn from the case presented in section 3.2.8. Unlike the previous cases it has not been possible to assess the statistical improvement in parameter estimates achieved by the use of optimally selected sampling points. This is not totally surprising as the Arrhenius expressions of equation 3.28 normally show very high level of cross-correlation and set a very severe test for most of the estimation algorithms.
Such a situation could possibly have been tackled through the use of parametric transformations, such as those already included in model 3.26. These have not been implemented because the scope of this chapter was not to test the use of parametric transformations. Moreover, parametric transformations in dynamic model should be used with great care with dynamic models. Biegler [1986] shows that parametric transformations on a dynamic model may change considerably the nature of the model itself leading even to changes in the number of state variables yielding a model rather different from the one originally thought by the experimenter. Although parametric transformations are useful and necessary tools, there is a definite need for improved dynamic parameter estimation algorithms which can tackle highly non linear systems.

The case study of section 3.6 shows that rather long computing times are required to find an optimal design. As reported in table 3.19 most of the time was spent in the gradient evaluations necessary for the many line searches carried out by the optimiser. Each gradient evaluation involves several integrations of the system in order to compute by finite differences the terms $\frac{\partial m_{ij}}{\partial \phi_k}$ of equation 2.37. Such a large number of line searches indicates a flat objective function and a certain level of numerical noise. It may be necessary to focus future research on the evaluation of such numerical derivatives, in particular on improving the finite difference scheme and optimal selection of perturbation step. Details on the method currently used are given in Appendix D.

During the analysis of these aspects, some problems were identified with the sensitivity trajectories calculation. However, The examples reported were carried out using a computer code for integration and dynamic sensitivity analysis of DAE systems produced in a related research project, but still in its early stage of development. This code has been selected based on its unique ability to handle nonlinear differential and algebraic systems, to provide a consistent initialisation for such systems, to handle discontinuities during the integration and to provide the required dynamic sensitivity analysis for such systems. Initial testing has confirmed the good performance of the code. As the testing of the dynamic experiment design progressed, however, it become apparent that many of the noise problems experienced during the optimisation were in fact related to the sensitivity calculations. A throughout review of the numerical methods employed for such calculation is reported in the following Chapter 4 and Appendix E.

Further progress with the algorithm for the design of dynamic experiments proposed in this chapter will require in order:

1) Fixing the present inaccuracies of the sensitivity calculation so that any kind of DAE system can be correctly tackled

2) Possibly, the determination of the optimal calculation procedure for the numerical derivatives needed in the gradient formula 2.37.

The previous two points should be tackled in that order, as any analysis of improved differentiation schemes would be pointless until the calculation of the sensitivities is robust.

A comparison between the design criteria 3.16 and 2.19 shows some important differences. In the steady state design, the design matrix $S_r$ increases its number of rows as the sequential design proceeds. Only the last rows of $S_r$ vary during the optimisation, as the previous ones are due to experiments already carried out and they need to be computed just once with the last estimates of $\hat{\theta}$. Such a situation is not repeated in the dynamic case where the matrix $Q_r$ has rows representing different sampling points in different experiments. During an optimisation run each row of $Q_r$ may change. In the stationary design discussed in Chapter 2, previous information is directly put in the objective function as the setting of the n previous experiments and the correspondent parameter estimate $\hat{\theta}_n$ while in the dynamic case only this last information is supplied. Such a situation is mostly evident by the values of the objective function which, unlike in the stationary case, do not increase steadily with the number of experiments, but stay in the
vicinity of a relatively constant value.

Other matters of future investigation would be to quantify whether the inclusion of previous experiments may lead to any increase in the content of information. Previous experiments could be taken into account by defining the design matrix \( Q_r \) as:

\[
Q_r = \begin{bmatrix}
Q_1^r(\hat{\theta}_n) \\
Q_2^r(\hat{\theta}_n) \\
\vdots \\
Q^{n+1}_r(\hat{\theta}_n)
\end{bmatrix}
\]

which is a \((n_{sp} \times (n+1)) \times p\) dimensional matrix. Only the last \( n_{sp} \) rows of \( Q_r \) would be variable during the optimisation. As the previous \( n \) rows need to be calculated only once and do not need recalculation for gradient evaluation, the extra computational load in using \( Q_r \) as design matrix should be limited.

Observing the sampling points determined in the cases of this chapter, it can be seen how the constraints which require a minimum interval between consecutive sampling points are often active. The design tends to locate sampling points about peaks in the sensitivities. Without any minimum interval constraint sampling points would be overlapping on these peaks. This is the situation observed in steady state design. Of course, in dynamic case this cannot be accepted as repeated samplings at the same time would require to repeat the experiment as well.

The previous discussion of the areas where further investigation is necessary, does not decrease neither the value nor the novelty of the algorithm proposed in this chapter. These can be summarised as follows:

1) - Extension to dynamic models of the probabilistic approach to the design of experiments described in Chapter 2.

2) - Solution of the algorithmic and mathematical limitations (such as the very restrictive requirement for the matrix of variance covariance of parameter estimates to be diagonal) which precluded wider application of previous existing algorithms.

3) - Use of an efficient optimisation method which can tackle without difficulties the large scale problems possibly arising from such kind of designs.

4) - Capability to address practical problems through the optimal selection of discrete and independent sampling points. It should be remarked that this is a very important requirement, as it is a common occurrence in kinetic modelling that the sampling points are the only design variable left to the experimenter. This capability could not be found in previous algorithms dealing with the optimal design of dynamic experiments for deterministic models in the time domain.

103
Sensitivity analysis of dynamic systems. Computational methods and use as an aid for model building.

The previous chapter has shown that the dynamic sensitivity coefficients are essential in order to carry out the design of experiments for dynamic systems. In this chapter, methods to compute such coefficients are discussed. Moreover, sensitivity analysis is, on its own, a valuable method to investigate the deterministic behaviour of dynamic systems. This is shown with an example. In this example, the concept of identifiability of a model is also considered. Identifiability analysis provides an investigation technique which relies only the structural property of the model and therefore it suits the first stage of conceptual development of a model. A detailed analysis and discussion of this technique is not among the aims of this work, however a short introduction is reported in order to point out its role as an useful tool in the general framework of the mathematical model building activity.

4.1 Sensitivity analysis of dynamic systems: computational methods.

Given a dynamic system such as 3.1, the dynamic sensitivity coefficients are defined by the partial derivative of the state variables with respect to the parameters:

$$q_{ij}(t) = \frac{\partial y_i(t, \hat{\theta})}{\partial \hat{\theta}_j} \quad i = 1 \ldots ns \quad j = 1 \ldots p$$

(4.1)

where $ns$ indicates the total number of variables in the model, both algebraic and differential. The actual computation of the $q_{ij}$'s is called sensitivity analysis. The coefficients defined by equation 4.1 are also referred to as first order sensitivity coefficients, higher order sensitivities being defined by successive derivatives.

The simplest approach to calculate the sensitivity coefficients involves one at a time variations of each parameter. The trajectories of the perturbed states are computed and stored. The sensitivity coefficient $q_{ij}$ is then approximated by the
ratio $\frac{\delta y_i(t, \delta \theta_j)}{\delta \theta_j}$. Computationally, such an approach is extremely simple as only a robust DAE's solver is required. However, it has the drawback that the calculation of each of the $q_{ij}$'s requires one integration of the dynamic system. The complete sensitivity analysis of a dynamic system requires a fairly large number of successive integrations of the model 3.1. Moreover the results can be heavily affected by the size of the perturbation $\delta \theta_j$. A similar approach was used by Seigneur and al. [1982] in computing the sensitivity of an objective function to the uncertainty in the value of the parameters.

The most common approach considers the sensitivity coefficients as state variables and it develops differential equations which describe the time evolution of the $q_{ij}$'s. Differentiating 4.1 with respect to the time one obtains:

$$q_{ij}' = \frac{\partial f_i}{\partial \theta_j} + \sum_{k=1}^{n_s} \frac{\partial f_i}{\partial y_k} q_{ik} \quad i = 1 \ldots n_s \quad j = 1 \ldots p$$

Equation 4.2 is rewritten in matrix notation as:

$$Q(t)' = J(\theta) + J(y)Q(t)$$

where $J(\theta)$ is the matrix of partial derivatives of the model with respect to the parameters $\theta$ and $J(y)$ is the Jacobian matrix. Equations 4.2 or 4.3 define a $ns \times p$ dimensional system ordinary differential equations with initial condition (1):

$$q_{ij}(t_0) = 0 \quad i = 1 \ldots n_s \quad j = 1 \ldots p$$

Equations 4.3 can be solved simultaneously with the state equations 3.1 (Direct differential method). The size of such combined DAE's system can be rather large. The sensitivity equations 4.2 are linear with respect to the sensitivity coefficients $q_{ij}$ and such a linearity is irrespective of the nonlinearity of the original differential systems. It has been suggested [Seinfeld and Lapidus, 1974] to take advantage of this and to solve the state equations first, then to use the trajectories of the state variables, appropriately stored, to compute the matrices $J(\theta)$ and $J(y)$ at

---

(1) This is not true if the initial conditions of the dynamic system 3.1 are a function of the parameters [Pantelides and Jarvis, 1990] However such a case is not of interest in this context where the $y_0$'s are always independently set.
the different times and calculate the $q_{ij}$'s by interpolation. This procedure has been sometimes preferred, because, although it is subject to interpolation errors, it requires smaller computer recourses. Presently, this is less of a problem. However the use of the direct differential method requires the sensitivity equations to be explicitly coded. This is a tedious and error prone activity if done manually.

The linearity of equations 4.2 enables one to compute the sensitivity coefficients simultaneously during the integration of the system 3.1 without explicitly defining the sensitivity equations. This approach is called decoupled direct method [Dunker, 1984]. For sake of simplicity, let us consider a system of ordinary differential equations $f(y', y, \theta, t) = 0$ with set initial conditions $y_0$. If a BDF method is used for integration, the predictor for the state variables at time $t_{n+1}$ is computed by the Gear's backward differentiation formula, (BDF), [Gear, 1971] as follows:

$$y_{n+1}^0 = \sum_{i=0}^{k-1} a_i y_{n-i} + h b_{-1} y_{n+1}'$$  \hspace{1cm} (4.5)

where $a_i$’s, $b_{-1}$ are polynomial interpolation coefficients and $h$ is the integration step. The predictor $y_{n+1}$ should satisfy the relation:

$$f\left(\frac{1}{h b_{-1}} \left[ y_{n+1}^0 - \sum_{i=0}^{k-1} a_i y_{n-i} \right], y_{n+1}, \theta, t \right) = 0$$  \hspace{1cm} (4.6)

Equation 4.6 is solved using Newton’s method. The $m^{th}$ step of the Newton iteration is:

$$[J(y') + h b_{-1} J(y)]_{n+1}^m [y_{n+1}^m - y_{n+1}] = h b_{-1} f(y_{n+1}', y_{n+1}; \theta, t)$$  \hspace{1cm} (4.7)

where $J(y')$ is the matrix of partial derivatives with respect to the differential variables. If the difference $\Delta y_{n+1}^m = [y_{n+1}^m - y_{n+1}^m]$ is equal to or less than a set relative error, the value $y_{n+1}^m$ is accepted, otherwise another iteration is carried out. Upon acceptation of the $y_{n+1}^m$, if the norm $\|y_{n+1}^m - y_{n+1}^0\|$ is less than a set absolute error then the step is accepted, otherwise $h$ is reduced.

The application of a Gear’s BDF formula to the sensitivity equations 4.3 yields the following predictor:

$$Q_{n+1}' = \frac{1}{h b_{-1}} \left[ Q_{n+1} - \sum_{i=0}^{k-1} a_i Q_{n-i} \right]$$  \hspace{1cm} (4.8)
The substitution of the predictor 4.8 into the sensitivity equations 4.3 yields a linear system which can be solved without need for Newton iterations:

\[
[J(y') + hb_{-1}J(y)]_{n+1}Q_{n+1} = hb_{-1} \left( \sum_{i=0}^{k-1} a_i Q_{n-i} \right) - J(\theta)
\]  \hspace{1cm} (4.9)

The iteration matrix \([J(y') + hb_{-1}J(y)]_{n+1}\) is the same in both equations 4.9 and 4.7.

Such an approach has been implemented by Leis and Kramer [1988a,b] in the code ODESSA. An analysis of the extension to DAE's system has been also carried out [Leis and Kramer, 1985]. The code ODESSA integrates normally the state equations, but any time a successful integration step has been achieved, the sensitivity coefficients are computed by equation 4.9. The iteration matrix is exactly the same as last used during the integration and the same integration step and polynomial coefficients are used. This procedure implies that the iteration matrix may be computed at a different time than the actual time at which the \(q_{ij}\)'s are computed, since the integrator does not need to update the iteration matrix at each step. The authors of ODESSA claim that the approximation of such a procedure is small compared to the gain in computational efficiency due to the fact that the LU decomposition of the iteration matrix can be preserved. The implementation of ODESSA has required relatively marginal modifications to the stiff ODE's solver LSODE [Hindmarsh, 1980]. The only substantial extra information the user is required to supply is the matrix \(J(\theta)\). The same basic approach has been adopted by Caracotsios and Steward [1985] in the code DASAC, which is a modification of the stiff DAE solver DASSL [Petzold, 1982a]. Dynamic sensitivity analysis is also being implemented in the code DASOLV [Pantelides and Jarvis, 1990]. Unlike the previous two codes, DASOLV updates the iteration matrix before computing the sensitivity coefficients. This is thought to be necessary for accurate computation of the \(q_{ij}\), even though it is computationally more expensive.

In all these codes the sensitivity coefficients at the user required output points are computed by interpolation among the terms stored in a Nordsieck history array, as is done for the state variables.

The problem of numerical sensitivity analysis with simultaneous integration of the dynamic system has been tackled also with different approaches. Among others, the Green's function method has been successfully applied to large chemical
networks such as those arising in combustion problems and the chemistry of the atmosphere [Hwang and al., 1978] [Dougherty and al., 1979]. This approach, although suitable for large problems, is not effective for stiff systems. Moreover it has been shown, [Dunker, 1984], that the direct decoupled method is less demanding in terms of the coding necessary for implementation, and computationally more accurate. Tilden and al. [1980] report a complete review of all the then available methods.

4.1.1 Comparison of sensitivity analysis codes. ODESSA, DASAC and DASOLV.

The three codes ODESSA, DASAC and DASOLV are compared in order to establish their relative robustness and accuracy. Despite the similarities of the numerical approach used by the three codes, there are differences in the way error control on the sensitivities computation is carried out. The error in the sensitivities is proportional to the residual between predicted values (equation 4.8) and corrected values (equation 4.9).

ODESSA allows the user to specify individual values of the relative and absolute error for each of the sensitivity coefficients. These are used to compute individual error weights as done for the model equations. The values of the sensitivities are accepted if the following inequalities are satisfied:

\[
\frac{1}{n} \sum_{i=1}^{n} \left( \frac{r_{ij}}{\omega_{ij}} \right)^2 \leq 1.0 \quad i = 1 \ldots p \tag{4.10}
\]

where \( r_{ij} \) is the residual of the coefficient \( q_{ij} \) and \( \omega_{ij} \) is the correspondent user defined weight. If the inequalities 4.10 are not satisfied, then the integration is repeated with a reduced step length.

DASAC uses a similar error control strategy, however the vector of residuals \( r_j \) of the sensitivities with respect to \( \hat{\theta}_j \), is augmented with the model error vector \( \epsilon \). The inequalities to be satisfied are:

\[
\frac{1}{2n} \left[ \sum_{i=1}^{n} \left( \frac{\epsilon_i}{\omega_{i0}} \right)^2 + \sum_{i=1}^{n} \left( \frac{r_{ij}}{\omega_{ij}} \right)^2 \right] \leq 1.0 \quad i = 1 \ldots p \tag{4.11}
\]

where the \( \omega_{i0} \) are the user selected weights for the state variables. The use of the norm 4.11 might lead to accept a step which would not be accepted by using the norm 4.10, as some of the larger error terms are compensated. This eventually causes undesirable accumulation of error.
DASOLV uses a different error control strategy and the user is not required to supply extra tolerance other than those required for the model integration. DASOLV requires the following inequality to be satisfied:

\[ h \left( \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{z_{ij}}{nm} \right)^2 \leq 1.0 \]  \( (4.12) \)

where \( z_{ij} = q_{ij} \) for algebraic variables and \( z_{ij} = hq_{ij} \) for differential variables.

DASOLV and DASAC are both able to handle DAE systems but DASAC does not include any implicit discontinuity checking. DASOLV also incorporates a robust technique for handling explicit discontinuities. The initialisation of the DAE’s system is automatic in DASOLV, while in DASAC the user needs to supply consistent initial condition as the built in initialisation routine can only handle ODE’s systems.

A comparison among the performance of the three codes has also been carried using the ODE’s model describing the simple reacting system \( A \rightarrow B \rightarrow C \):

\[
\begin{align*}
c_A' &= -\exp(\theta_1)c_A \\
c_B' &= \exp(\theta_1)c_A - \exp(\theta_2)c_B \\
c_C' &= \exp(\theta_2)c_B
\end{align*}
\]  \( (4.13) \)

The values of the parameters are \( \theta_1 = -1.572 \) and \( \theta_2 = -0.702 \) and the initial conditions are \( c_{A0} = 1.0, \quad c_{B0} = c_{C0} = 0.0 \). Figure 4.1 reports the state trajectories, which, as expected, are the same for the three codes.

The sensitivity coefficients calculated by the three codes are compared with those obtained by simultaneous solutions of the model equations 4.13 and of the explicit sensitivity equations:

\[
\begin{align*}
q'_{11} &= -\exp(\theta_1)(c_A + q_{11}) \\
q'_{12} &= -\exp(\theta_2)q_{21} \\
q'_{21} &= \exp(\theta_1)(c_A + q_{11}) - \exp(\theta_2)q_{21} \\
q'_{22} &= -\exp(\theta_2)q_{12} - \exp(\theta_2)(c_B + q_{22}) \\
q'_{31} &= \exp(\theta_2)q_{31} \\
q'_{32} &= \exp(\theta_2)(c_B + q_{32})
\end{align*}
\]  \( (4.14) \)

Equations 4.14 are an example of the considerable enlargement of the system which takes place when explicit sensitivity equations are used. Throughout the study,
Figure 4.1. — Reaction $A \rightarrow B \rightarrow C$, evolution of the state variables.

Each relative and absolute tolerance required for state variables and sensitivities has been kept fixed to $10^{-5}$. The sensitivity coefficients considered for comparison are $q_{11}$, $q_{21}$ and $q_{22}$. The results are reported in the following tables and figures.
Table 4.1. — Sensitivity coefficients $q_{11}$ comparison for three codes. DDF: direct differential method.

<table>
<thead>
<tr>
<th>Time</th>
<th>DDF</th>
<th>ODESSA</th>
<th>DASAC</th>
<th>DASOLV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.500E - 1</td>
<td>-4.928E - 2</td>
<td>-4.928E - 2</td>
<td>-1.519E - 4</td>
<td>-4.922E - 2</td>
</tr>
<tr>
<td>1.000E + 0</td>
<td>-1.687E - 1</td>
<td>-1.687E - 1</td>
<td>-2.072E - 4</td>
<td>-1.686E - 1</td>
</tr>
<tr>
<td>2.000E + 0</td>
<td>-2.741E - 1</td>
<td>-2.741E - 1</td>
<td>-3.724E - 4</td>
<td>-2.741E - 1</td>
</tr>
<tr>
<td>3.000E + 0</td>
<td>-3.341E - 1</td>
<td>-3.341E - 1</td>
<td>-3.026E - 4</td>
<td>-3.341E - 1</td>
</tr>
<tr>
<td>4.000E + 0</td>
<td>-3.619E - 1</td>
<td>-3.619E - 1</td>
<td>-2.459E - 4</td>
<td>-3.619E - 1</td>
</tr>
<tr>
<td>5.000E + 0</td>
<td>-3.676E - 1</td>
<td>-3.676E - 1</td>
<td>-1.998E - 4</td>
<td>-3.676E - 1</td>
</tr>
<tr>
<td>7.500E + 0</td>
<td>-3.281E - 1</td>
<td>-3.281E - 1</td>
<td>-2.926E - 4</td>
<td>-3.281E - 1</td>
</tr>
<tr>
<td>1.000E + 1</td>
<td>-2.603E - 1</td>
<td>-2.603E - 1</td>
<td>-1.989E - 4</td>
<td>-2.603E - 1</td>
</tr>
<tr>
<td>1.250E + 1</td>
<td>-1.936E - 1</td>
<td>-1.936E - 1</td>
<td>-2.108E - 4</td>
<td>-1.936E - 1</td>
</tr>
<tr>
<td>1.500E + 1</td>
<td>-1.382E - 1</td>
<td>-1.382E - 1</td>
<td>-4.288E - 4</td>
<td>-1.382E - 1</td>
</tr>
<tr>
<td>1.700E + 1</td>
<td>-1.034E - 1</td>
<td>-1.034E - 1</td>
<td>-7.695E - 5</td>
<td>-9.967E - 2</td>
</tr>
<tr>
<td>1.950E + 1</td>
<td>-7.063E - 2</td>
<td>-7.062E - 2</td>
<td>-1.967E - 4</td>
<td>-6.791E - 2</td>
</tr>
</tbody>
</table>

Figure 4.2. — Evolution of the coefficient $q_{11}$. 1 — DDF, ODESSA and DASOLV. 2 — DASAC.
Table 4.2. — Sensitivity coefficients $q_{21}$ comparison for three codes.

<table>
<thead>
<tr>
<th>Time</th>
<th>DDF</th>
<th>ODESSA</th>
<th>DASAC</th>
<th>DASOLV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.500E-1</td>
<td>4.630E-2</td>
<td>4.630E-2</td>
<td>1.572E-4</td>
<td>4.872E-2</td>
</tr>
<tr>
<td>5.000E-1</td>
<td>8.251E-2</td>
<td>8.252E-2</td>
<td>2.198E-4</td>
<td>9.167E-2</td>
</tr>
<tr>
<td>1.000E+0</td>
<td>1.306E-1</td>
<td>1.306E-1</td>
<td>1.905E-4</td>
<td>1.620E-1</td>
</tr>
<tr>
<td>2.000E+0</td>
<td>1.610E-1</td>
<td>1.610E-1</td>
<td>3.158E-4</td>
<td>2.532E-1</td>
</tr>
<tr>
<td>3.000E+0</td>
<td>1.441E-1</td>
<td>1.441E-1</td>
<td>2.336E-4</td>
<td>2.968E-1</td>
</tr>
<tr>
<td>4.000E+0</td>
<td>1.089E-1</td>
<td>1.089E-1</td>
<td>1.758E-4</td>
<td>3.095E-1</td>
</tr>
<tr>
<td>5.000E+0</td>
<td>7.021E-2</td>
<td>7.022E-2</td>
<td>1.343E-4</td>
<td>3.027E-1</td>
</tr>
<tr>
<td>7.500E+0</td>
<td>-5.280E-3</td>
<td>-5.283E-3</td>
<td>1.775E-4</td>
<td>2.464E-1</td>
</tr>
<tr>
<td>1.000E+1</td>
<td>-4.085E-2</td>
<td>-4.086E-2</td>
<td>2.512E-4</td>
<td>1.789E-1</td>
</tr>
<tr>
<td>1.250E+1</td>
<td>-4.956E-2</td>
<td>-4.957E-2</td>
<td>2.038E-4</td>
<td>1.221E-1</td>
</tr>
<tr>
<td>1.500E+1</td>
<td>-4.534E-2</td>
<td>-4.534E-2</td>
<td>3.358E-4</td>
<td>8.037E-2</td>
</tr>
<tr>
<td>1.700E+1</td>
<td>-3.850E-2</td>
<td>-3.849E-2</td>
<td>2.052E-4</td>
<td>5.641E-2</td>
</tr>
<tr>
<td>1.900E+1</td>
<td>-3.113E-2</td>
<td>-3.113E-2</td>
<td>1.647E-4</td>
<td>3.915E-2</td>
</tr>
<tr>
<td>1.950E+1</td>
<td>-2.936E-2</td>
<td>-2.935E-2</td>
<td>1.397E-4</td>
<td>3.567E-2</td>
</tr>
</tbody>
</table>

Figure 4.3. — Evolution of the coefficient $q_{21}$. 1 — DDF and ODESSA. 2 — DASOLV. 3 — DASAC.
Table 4.3. — Sensitivity coefficients $q_{22}$ comparison for three codes.

<table>
<thead>
<tr>
<th>Time</th>
<th>DDF</th>
<th>ODESSA</th>
<th>DASAC</th>
<th>DASOLV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.50E - 1</td>
<td>-2.909E - 3</td>
<td>-2.906E - 3</td>
<td>-2.355E - 5</td>
<td>-3.016E - 3</td>
</tr>
<tr>
<td>5.00E - 1</td>
<td>-1.054E - 2</td>
<td>-1.053E - 2</td>
<td>-4.325E - 5</td>
<td>-1.093E - 2</td>
</tr>
<tr>
<td>2.00E + 0</td>
<td>-9.341E - 2</td>
<td>-9.341E - 2</td>
<td>-2.240E - 4</td>
<td>-1.070E - 1</td>
</tr>
<tr>
<td>3.00E + 0</td>
<td>-1.426E - 1</td>
<td>-1.426E - 1</td>
<td>-2.322E - 4</td>
<td>-1.746E - 1</td>
</tr>
<tr>
<td>4.00E + 0</td>
<td>-1.730E - 1</td>
<td>-1.730E - 1</td>
<td>-2.193E - 4</td>
<td>-2.259E - 1</td>
</tr>
<tr>
<td>5.00E + 0</td>
<td>-1.853E - 1</td>
<td>-1.853E - 1</td>
<td>-1.968E - 4</td>
<td>-2.576E - 1</td>
</tr>
<tr>
<td>7.50E + 0</td>
<td>-1.661E - 1</td>
<td>-1.661E - 1</td>
<td>-3.303E - 4</td>
<td>-2.675E - 1</td>
</tr>
<tr>
<td>1.00E + 1</td>
<td>-1.217E - 1</td>
<td>-1.217E - 1</td>
<td>-3.814E - 4</td>
<td>-2.234E - 1</td>
</tr>
<tr>
<td>1.25E + 1</td>
<td>-8.094E - 2</td>
<td>-8.094E - 2</td>
<td>-3.219E - 4</td>
<td>-1.668E - 1</td>
</tr>
<tr>
<td>1.50E + 1</td>
<td>-5.119E - 2</td>
<td>-5.119E - 2</td>
<td>-6.800E - 4</td>
<td>-1.164E - 1</td>
</tr>
<tr>
<td>1.95E + 1</td>
<td>-2.112E - 2</td>
<td>-2.112E - 2</td>
<td>-2.894E - 4</td>
<td>-5.524E - 2</td>
</tr>
</tbody>
</table>

Figure 4.4. — Evolution of the coefficient $q_{22}$. 1 — DDF and ODESSA. 2 — DASOLV. 3 — DASAC.
It is striking to notice the inability of DASAC to obtain precise results. It seems that the error control strategy of DASAC is ineffective because the error control in the sensitivities is overridden by the error control in the states. No sensibly better results were obtained with more stringent tolerances on the sensitivity coefficients. Moreover the use of tighter tolerances leads to an excessive number of steps necessary to complete the integration. ODESSA manages to reproduce correctly the sensitivity trajectories with standard values of the tolerances. DASOLV, although it gives results far better than DASAC, seems to be affected by accumulation of error which makes its sensitivity trajectories diverge from the real ones. However, the general overall shape is still relatively well preserved. These results seem to justify the claim that the iteration matrix does not need to be updated when evaluating the sensitivity coefficients.

Despite the not totally satisfying results, DASOLV has been chosen to be used in a code for the design of dynamic experiments. At the time in which these tests were carried out DASOLV was still at an early stage of development, but, unlike ODESSA, it handles DAE's and it offers advanced features for the handling of discontinuities. These are very important features for a robust usage of time-dependent piecewise constant controls. The inaccuracies in the sensitivity calculation by DASOLV were increasing with the stiffness of the system. It is advisable that an improved error control strategy should be developed.

4.1.2 Comments on sensitivity analysis.

In the previous chapter, it has been noted how the nonlinear design depends on the values of the parameter estimates. Similar conclusions can be drawn for the dynamic case. It is obvious that, if system 3.1 is linear with respect to the parameters, the sensitivity equations 3.27 do not depend on $\hat{\theta}$. On the contrary, such a dependency exists when the dynamic system is nonlinear with respect to the parameters. In this case, the results of the sensitivity analysis are greatly affected by the value of the parameters, and the use of greatly approximated values when designing experiments might be grossly misleading. This kind of design is therefore inappropriate for the very first stages of the experimentation when no reasonably approximate first guesses of the parameters values are available.
It has been shown in the previous chapter that the gradient evaluation is the most demanding step of the proposed algorithm for the design of dynamic experiments. It is obvious that such a computation can lead to severe numerical problems due to the noise which accumulates during the many steps taken before the actual derivative is calculated. However, as shown, dynamic sensitivity analysis of stiff system is still relatively unstable making the computation of the such gradients relatively uncertain, and explains the difficulties experienced with the optimal design.

4.2 The use of sensitivity analysis as an aid for model building.

So far only the problems related to the mathematical modeling, as defined in chapter 1, have been considered. However, the importance of the conceptual development of a model is not to be underestimated in order to successfully approach the subsequent problem of experiments’ planning and design. In a nonlinear model the availability of sensible initial guesses is crucial to any further development of the investigation. At an early stage, sensitivity analysis can be used to point areas where some parameters are redundant (zero or very small sensitivities). This, in turn, may allow one to reduce the model to a smaller and more easily manageable one. This technique enables the experiment to achieve two goals: easier identification of initial guesses for unknown parameters and simplification of overparameterised models. Related to the topics above, one should know, when a model is chosen, whether the available set of measured variables is sufficient to uniquely identify parameters value. A possible solution to this problem is given by a technique which solely concentrates on the structural mathematical properties of the model, without the need for any numerical information. Although this technique is a stand alone tool, it is reported in this section because, in our opinion, it is most effective as a method to point out problems which can be efficiently tackled by sensitivity analysis.

4.2.1 Structural identifiability analysis.

Structural identifiability analysis, has been formally posed by Bellman and Anstrom [1970], mainly with reference to linear problems. Subsequently extensions to nonlinear dynamic systems have been proposed. However, as they only rely on symbolic manipulation of the model of interest, they have so far failed to raise a more
widespread interest. This should no longer be justified as powerful tools for symbolic computation, such as Mathematica [Wolfram, 1988], are easily available on a wide range of computers.

The concept of identifiability can be better expressed by using an example. Let us consider the linear dynamic model:

\[
x'_1 = -\theta_1 x_1 + \theta_2 u(t) \\
x_2 = \theta_3 x_1 \\
x_1(t_0) = 0
\]  

(4.15)

where \(x_2\) is the only measured variable and \(u(t)\) is a time dependent input into the system. Equation 4.15 admits the following analytical expression for the trajectory of \(x_2\):

\[
x_2 = \theta_1 \theta_3 \int_0^t \exp(-\theta_1 (t - \tau)) u(\tau) d\tau
\]  

(4.16)

When the input \(u(t)\) is an impulse at time \(t_0\), such as \(u(t_0) = \delta\), and zero at any other time, equation 4.16 becomes:

\[
x_2 = \theta_2 \theta_3 \exp(-\theta_1 t)
\]  

(4.17)

which clearly shows that only \(\theta_1\) and the group \(\theta_2 \theta_3\) are uniquely identifiable, but \(\theta_2\) and \(\theta_3\) cannot be uniquely and independently estimated.

A complete review of the identifiability analysis techniques is reported by Godfrey and DiStefano [1987]. In the following a particular method of this type, suitable for nonlinear dynamic model is briefly discussed. The following discussion, far from being complete, is an overview of this technique which might constitute an interesting area of future research.

4.2.2 The power expansion method for identifiability analysis.

The power series expansion method [Pohjanpalo, 1978] is, thus far, the only reported identifiability analysys method suitable for nonlinear dynamic models. Let us consider a dynamic model such as:

\[
x'(t) = f(x(t), u(t), \theta) \\
y(t) = g(x(t), \theta) \\
x(t_0) = x_0
\]  

(4.18)
where \( y(t) \) is a subset of measured variables, \( u(t) \) a vector of time dependent inputs and \( \theta \) are the unknown parameters.

It has been shown that under certain hypotheses of relatively general validity [Pohjanpalo, 1978], a condition necessary and sufficient for the identifiability of the system 4.18 is that the equations:

\[
\frac{d^k g}{dt^k} = a_k(t_0) \quad k = 1 \ldots \infty
\]

(4.19)

have an unique solution for \( \theta \). The \( a_k \)'s are the \( k^{th} \) derivatives of the measurement vector \( y \) with respect to time, computed at the origin.

From a practical point of view, a finite number of steps is necessary to determine if a system is structurally identifiable or not. Considering model 4.15, one may obtain:

\[
\begin{align*}
   x_2(0) &= \theta_3 x_1(0) = a_0 \\
   x_2(0)' &= \theta_3 \theta_1 x_1(0)' = -3 \theta_3 x_1(0) + \theta_3 \theta_2 u(0) = \theta_3 \theta_2 \delta = a_1 \\
   x_2(0)'' &= \theta_3 x_1(0)'' = \theta_1 \theta_3 \theta_2 \delta = a_2 \\
   \vdots
\end{align*}
\]

(4.20)

The second equation of 4.20 shows that the group \( \theta_3 \theta_2 \) is identifiable and the third equations shows that \( \theta_1 \) is also identifiable. This is consistent with the results obtained by the analytical solution of the model. No further improvement can be achieved by subsequent derivatives. It should be noted that for this particular example the identifiability depends on the particular nature of the input \( u(t) \). This situation is quite common and it as been defined by Bellman and Anstrom [1970] as deterministic identifiability. This concept has an important consequence for the design of experiments, as the determination of the particular trajectory \( u(t) \) which makes the system identifiable can be used to generate a rather new design criterion.

This method has the obvious drawback that when the system’s equations are nonlinear, it is not trivial to see whether there are multiple solutions for the system 4.23 or if any redundancy exists. It should be borne in mind, moreover, that the concept of structural identifiability refers to an idealized, noise free situation. The satisfaction of the conditions for the structural identifiability is only a necessary condition for the subsequent successful parameter estimation. The actual parameter estimation could still fail due either to convergence problems or excessive experimental error. Such a situation is referred to as (lack of) numerical identifiability.
An example of the application of the method to complex nonlinear dynamic systems is given by Anstrom and Kallstrom [1976]. Pohjanpalo [1978] uses the method to study the identifiability of systems described by Michaelis-Menten rate equations and Langmuir isotherms. A similar example, of interest in biochemical systems, is reported by Travis and Haddock [1981] and Walter and Lecourtier [1981].

4.2.3 Example: preliminary investigation of a complex kinetic model.

In this example, the situation is supposed facing an engineer who, having found in literature a model such as the following 4.21, wants to verify whether the model can be a reasonable description of a process of interest. Such a scenario usually represents a very preliminary stage in the kinetic investigation. During such an early stage limited experimental resources are allocated or archive experimental data are used.

A complex model describing the free radical polymerisation of methylethacrylate has been considered. The model was proposed by Chiu and al. [1983] to describe the time evolution of monomer conversion and molecular weight of a growing polymer. The model takes into account the onset the gel effect, caused by the dramatic increase of viscosity in the reacting mixture. Such an effect is accounted in mechanistic terms using a theory (free volume) which correlates the length of polymeric chains with their mobility. The main kinetic implication of the gel effect is that the reactivity of the species involved is not constant, therefore, unlike other chemical systems, kinetic coefficients are time dependent.

The model consists of the following system of equations:

\[
\begin{align*}
\frac{dI}{dt} &= -(k_d + \varepsilon \phi M_0 k_p)I \\
\frac{dx}{dt} &= k_p(1 - x)I \\
\frac{dM_0}{dt} &= -(\varepsilon \phi k_p + k_t)M_0^2 + 2f k_d I \\
\frac{1}{k_p} &= \frac{1}{k_p^0} + \theta_p \frac{M_0}{\exp \left( \frac{2.3 \phi}{A+B\phi} \right)} \\
\frac{1}{k_t} &= \frac{1}{k_t^0} + \theta_t \frac{M_0}{\exp \left( \frac{2.3 \phi}{A+B\phi} \right)} \\
\phi &= \frac{1 - x}{1 + \varepsilon x}
\end{align*}
\]  

(4.21)
where $x$ is the fractional conversion, $I$ the initiator concentration and $M_0$ the zeroth moment of the growing chain. Six parameters are unknown: $k_p^0$ (propagation coefficient at initial time), $k_t^0$ (termination coefficient at initial time), $\theta_t$ and $\theta_p$ (characteristic migration times of growing chain and dead polymer respectively) and $A$ and $B$ which are diffusivity coefficients.

Other parameters appearing in the model are assumed to be either known or independently estimable. These are $\epsilon$, which is the volume expansion factor, $k_d$, dissociation coefficient for the initiator, and $f$, which is an efficiency factor expressing which fraction of the radicals generated by the dissociation of the initiator actually give birth to a polymeric chain.

Only the conversion $x$ and the zeroth moment of the growing chain, $M_0$ can be measured. However the analytical technique used to measure $M_0$ (gel permeation chromatography) is subject to large experimental error at low conversions. These two variables do not allow model 4.21 to be identifiable. In particular, the application of the power expansion method shows that four parameters, $\theta_p$, $\theta_t$, $A$ and $B$ cannot be uniquely identified. For sake of brevity, the derivatives are omitted.

Sensitivity analysis is used to establish whether any model reduction is possible. By reducing the model to a subset of the model’s equations to describe the process relative to a limited time interval, it can be shown that the estimation of some parameters is allowed. These parameters are subsequently kept as constants in estimations carried out with a more complete equation set. In mechanistic models, parameters are representative of well determined physical phenomena. Dropping a parameter from a model amounts to acknowledging the non relevance of the underlying physical phenomenon. Such a non relevance is formally expressed by low or zero values of the correspondent dynamic sensitivity coefficient.

Such an approach is quite common in experimental practice, however it mostly relies on the experimenter’s ingenuity and physical insight into the process. Dynamic sensitivity analysis may provide a formalized measure for this procedure. This would help in highlighting hidden or not obvious relations. In the case of nonlinear models, however, sensitivity coefficients depends on the values of the parameters. It is necessary to assess whether this procedure is viable also with grossly approximated parameter estimates.
The parameter values chosen to carry out the sensitivity analysis are shown in table 4.4. These values are chosen accordingly to the suitable ranges reported in literature [Bamford, 1988], these are also reported in the table 4.4.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Guessed value</th>
<th>Suitable range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_0^p$</td>
<td>$5.0 \times 10^4$</td>
<td>$3.0 \div 6.0 \times 10^4$</td>
</tr>
<tr>
<td>$k_0^t$</td>
<td>$2.0 \times 10^9$</td>
<td>$1.0 \div 30.0 \times 10^8$</td>
</tr>
<tr>
<td>$\theta_p$</td>
<td>100.0</td>
<td>$50.0 \div 150.0$</td>
</tr>
<tr>
<td>$\theta_t$</td>
<td>50.0</td>
<td>$30 \div 100.0$</td>
</tr>
<tr>
<td>$A$</td>
<td>0.10</td>
<td>$\leq 0.50$</td>
</tr>
<tr>
<td>$B$</td>
<td>0.01</td>
<td>$\leq 0.05$</td>
</tr>
</tbody>
</table>

Table 4.4. — Polymerisation model. Initial guesses for parameter values and suitable ranges as reported by literature.

Figures 4.5 and 4.6 show the sensitivities of the fractional conversion with respect to $k_0^p$ and $k_0^t$. Examination of these curves show that both sensitivities are initially very low and the first decreases while the second increases following an almost symmetrical pattern.

![Figure 4.5](image-url)  

*Figure 4.5. — Sensitivity of the fractional conversion $x$ with respect to the parameter $k^0_p$*
Figure 4.6. — Sensitivity of the fractional conversion $x$ with respect to the parameter $k_t^0$.

Each curve, after reaching an extreme value, goes to zero (not shown in the figures). Such asymptotic behaviour is consequence of the conversion asymptotically converging towards unity. Similar behaviour is shown by the sensitivities of the fractional conversion with respect to the parameters $A$ and $B$. In this case however, the sensitivities are smaller than before and almost negligible for any conversion less than 0.4. This is shown in figures 4.7 and 4.8. Analogous observation can be drawn by the observation of the sensitivity of $x$ with respect to $\theta_t$ and $\theta_p$.

An important conclusion, which confirms physical intuition, is that the four parameters describing the gel effect ($\theta_t$, $\theta_p$, $A$, $B$) are redundant at low conversion and can be discarded. At low conversions ($x \leq 0.4$), $k_p$ and $k_t$ can be considered constant and equal to $k_p^0$ and $k_t^0$ respectively. This assumption is used in conjunction the widely assessed quasi steady state assumption (QSSA), which states that at low conversion concentration of radicals and average molecular weight are constant. This enables one to set $I = I_0$ and $P_0 = \text{const}$. A suitable value for $P_0$ is found in literature: $P_0 = 1.67 \times 10^{-7}$ [Bamford, 1988].

The estimation of $k_p^0$ is thus reduced to an easy problem of linear regression since only the second equation of the model 4.25 is required. After this estimation
Figure 4.7. — Sensitivity of the fractional conversion $x$ with respect to the parameter $A$.

Figure 4.8. — Sensitivity of the fractional conversion $x$ with respect to the parameter $B$.

has been carried out, the value for $k_i^0$ is immediately evaluated from the third equation of 4.25 with $dP_0/dt = 0$. The apparent kinetic coefficients $k_p$ and $k_t$
cannot be measured, therefore it is necessary to have an independent estimate. In a generic radical free polymerisation, the propagation rate may be described by:

\[ r_p = -\frac{dM}{dt} = k_p P_0 M \]  (4.22)

\( M \) being the monomer concentration, which, in a normalized fashion, is expressed by the variable \( \phi \):

\[ \phi = \frac{M}{M_0} = \frac{1 - x}{1 + \varepsilon x} \]  (4.23)

where \( M_0 \) is the initial monomer concentration. Substituting 4.23 into 4.22, one obtains:

\[ r_p = \frac{d\phi}{dt} = -k_p P_0 \phi \]  (4.24)

Equation 4.24 allows the estimation of \( k_p(x) \), provided that a value of \( r_p(x) \) is computed. This is given by numerical differentiation of \( \phi \). Measurement of \( P_0 \) is needed, however, for \( x \geq 0.4 \) \( P_0 \) is in excess of \( 10^{-3} \) and the experimental measurements are reliable because of the acceptable level of error in the analytical method.

Figure 4.9. — Evolution of the estimated and expected apparent propagation coefficient with respect to time (upper graph) and fractional conversion (lower graph).
The soundness of this procedure is proven by figure 4.9, where the evolution of the estimated $k_p$ is compared with the evolution of the expected one computed with the exact parameters used in the computer simulation (Table 4.5).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value in simulation</th>
<th>Estimated value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_p^0$</td>
<td>$4.965 \times 10^4$</td>
<td>$5.019 \times 10^4 \pm 9.38 \times 10^2$</td>
</tr>
<tr>
<td>$k_t^0$</td>
<td>$2.102 \times 10^9$</td>
<td>$2.018 \times 10^9$</td>
</tr>
<tr>
<td>log($\theta_p$)</td>
<td>5.521</td>
<td>5.520 $\pm$ 0.128</td>
</tr>
<tr>
<td>$\theta_t$</td>
<td>49</td>
<td>20.59 $\pm$ 8.47</td>
</tr>
<tr>
<td>$A$</td>
<td>0.152</td>
<td>0.145 $\pm$ 0.25</td>
</tr>
<tr>
<td>$B$</td>
<td>0.030</td>
<td>0.074 $\pm$ 0.042</td>
</tr>
</tbody>
</table>

Table 4.5. — Polymerisation model. Parameter values as reported in the original paper and used in the computer simulation and as estimated.

Parameters $A$, $B$ and $\theta_p$ are estimated by linearising the fourth equation of model 4.21 as follows:

$$f(\phi) = \log(\theta_p) - \frac{2.3A}{A + B\phi}$$

$$f(\phi) = \frac{1}{k_p} - \frac{1}{k_p^0}$$ (4.25)

The remaining parameter, $\theta_t$, is easily estimated using the whole model 4.21. Table 4.5 reports the parameter estimates compared with the expected values as used in the computer simulation. The parameter estimation was carried out using the nonlinear weighted least square code NPLEST [Espie, 1986].

Figure 4.10 shows the evolution of the fractional conversion with time as computed with the intial guesses (table 4.4), with the expected parameters and with the parameter estimates of table 4.5.

4.2.4 Comments.

The previous example, although far from being exhaustive, shows how a limited number of experimental measurements can be used to assess whether a given model deserves more attention. Such assessment is not in terms of statistical adequacy to fit the data, but only in terms of qualitative ability of the model to describe the process. This kind of analysis, therefore, does not guarantee the model to be
the right one but it enables, on a preliminary stage, a more sensible qualitative screening.

This example also confirms that sensitivity analysis is an effective way to study the dynamics of a reacting system as it helps the experimenter to augment his/her physical insight into the system. This is particularly appreciable for fast reactions such as combustions. This kind of application has been studied by Lutz and al. [1988].

The results achieved in the previous section were relatively predictable, however the good agreement between physical intuition and mathematical results enables one to trust the procedure. Similar considerations should be made for the identifiability analysis, even though, at the present stage, its application is still rather cumbersome.

The use of the procedure described in the previous section should be greatly improved because, at the moment, it does not have the rigour necessary to be implemented as a formalized and fully automated computer algorithm.
In this chapter, the results achieved so far are discussed and future areas of research pointed out. In particular, the previous chapter has shown that the lack of an adequate code for dynamic parameter estimation is a serious obstacle to improving the techniques of computer aided experimental design. This aspect is analysed with some detail in the subsequent sections. A novel algorithm for parameter estimation in dynamic models described by stiff systems of mixed algebraic and differential equations is proposed. This algorithm, although just a proposal and yet to be implemented, appears very attractive as it may solve some problems of general relevance. Moreover, such an algorithm, as it shares many basic routines with the algorithm for the dynamic design of experiments discussed in chapter 3, could be implemented in the same code generating an integrated package for the analysis and design of dynamic experiments.

5.1 Practical problems in dynamic multivariate parameter estimation.

It has been well demonstrated by Biegler and al. [1986] that parameter estimation in multivariate dynamic models is still a very complex problem requiring the intervention of specialists. In that work, the results of a difficult industrial problem, submitted by six independent teams of researchers are presented and discussed. The findings of the paper suggest the following observations:

i) All but one of the solutions involved extensive rearrangement and reduction of the proposed model.

ii) The values of parameter estimates reported by the different authors are greatly different, however they all manage to predict quite satisfactorily the experimental data.
iii) Few of the parameter estimates are supplied with the interval of confidence. Moreover, some of those supplied with interval of confidence exhibit an unacceptable high level of variability (12.392 ± 50 or even 9.457 ± 233!).

This last point is particularly worrying from the point of view of the statistical validation of the estimation results. When the term parameter estimation is used, it refers to a procedure which enables one to evaluate the parameters’ values of a certain model along with relevant statistical properties, namely the interval of confidence and crosscorrelation. When this information is missing, the estimation can not be assessed from a statistical point of view, regardless of the ability of the parameter estimates to reproduce the experimental data. Moreover, there is no formal definition of the acceptable level of variability in parameter estimates. However, it is quite intuitive that an interval of confidence grossly exceeding the value of the parameter itself can hardly indicate an estimate of some physical significance.

These findings correspond to the problems experienced in testing the algorithm for the design of dynamic experiments, where it has been impossible to assess statistically the improvement in the parameter estimates.

Parameter estimates are evaluated using some criterion which tends to minimize the discrepancy between predicted values and measured values. The most general criterion is the maximisation of the likelihood function:

\[
L(\hat{\theta}) = 2\pi^{-mn/2}|W(\epsilon)|^{-1/2}\left\{ -\frac{1}{2} \sum_{i=1}^{n} e^T W(\epsilon)^{-1} e \right\}
\]

(5.1)

where \(|W(\epsilon)|\) is the variance covariance matrix of the experimental error. Mostly, the problems in applying a maximum likelihood criterion are related to the matrix \(|W(\epsilon)|\). Such a matrix, which gives the prior information, is usually unknown and an independent estimate is not easily available. The estimation is simplified by assuming some knowledge on the error structure. In particular it is assumed that:

1) The errors at different sampling points are uncorrelated

2) The errors are normally distributed with zero mean and constant variance covariance matrix.

3) The errors are independent so that the variance covariance matrix is diagonal.
Under such conditions equation 5.1 reduces to the well known criterion of weighted least squares, which requires the minimisation of the expression below:

\[ F(\hat{\theta}) = \sum_{i=1}^{n} e^T V^{-1} e \]  

(5.2)

where the matrix \( V \) is a matrix of weights usually proportional to the variances of experimental errors. Some of the solutions submitted to Biegler [1986] used objective functions intermediate between 5.1 and 5.2, however the results achieved do not justify the extra expense due to criteria more complex than 5.2.

The assumptions under which equation 5.1 is simplified need to be commented. The first condition is easily satisfied by the adoption of sensible experimental practice, while the following ones require some extra consideration before being accepted.

It is easy to show that if a measured variable spans over some orders of magnitude and the error is a fixed percentage of the quantity being measured, then the error structure is no longer normal. Several analytical techniques used in kinetic investigation exhibit such a behaviour. Moreover, multivariate weighted least squares consider, during the optimisation, all the residuals simultaneously. In this case, due to the differences in the order of magnitude in the measured variables, the assumption of normal error structure is a gross approximation, even though the residuals relative to each response, individually taken, have normal error structure.

This can be more easily explained by the use of an example. Let us consider the following model:

\[
y_1 = \beta_1 / x_1 \\
y_2 = \beta_2 x_1^2
\]  

(5.3)

where \( \beta_1 = 1.862 \) and \( \beta_2 = 2.040 \). Computer simulated measurements are taken in the interval \( 1.0 \leq x \leq 10.0 \) with samplings each 0.5 units. A normally distributed error term is added to \( y_1 \) and \( y_2 \). The standard errors are respectively \( \sigma_1 = 0.12 \) and \( \sigma_2 = 12.0 \). Table 5.1 reports the exact and the measured values of the dependent variables. The residuals, on each dependent variable, are normally distributed. However, when they are considered altogether there is a large drift from the normal distribution.
Table 5.1. — Exact and measured states in model 5.3

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y_1$ Exact</th>
<th>$y_1$ Measured</th>
<th>$y_2$ Exact</th>
<th>$y_2$ Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.862E+00</td>
<td>1.861E+00</td>
<td>2.040E+00</td>
<td>5.852E+00</td>
</tr>
<tr>
<td>1.5</td>
<td>1.241E+00</td>
<td>1.434E+00</td>
<td>4.590E+00</td>
<td>2.278E+01</td>
</tr>
<tr>
<td>2.0</td>
<td>9.310E-01</td>
<td>1.032E+00</td>
<td>8.160E+00</td>
<td>1.715E+01</td>
</tr>
<tr>
<td>2.5</td>
<td>7.448E-01</td>
<td>7.769E-01</td>
<td>1.275E+01</td>
<td>6.657E+00</td>
</tr>
<tr>
<td>3.0</td>
<td>6.206E-01</td>
<td>5.098E-01</td>
<td>1.836E+01</td>
<td>2.898E+01</td>
</tr>
<tr>
<td>3.5</td>
<td>5.320E-01</td>
<td>5.235E-01</td>
<td>2.499E+01</td>
<td>2.201E+01</td>
</tr>
<tr>
<td>4.0</td>
<td>4.655E-01</td>
<td>4.832E-01</td>
<td>3.264E+01</td>
<td>2.392E+01</td>
</tr>
<tr>
<td>4.5</td>
<td>4.137E-01</td>
<td>3.469E-01</td>
<td>4.131E+01</td>
<td>3.596E+01</td>
</tr>
<tr>
<td>5.0</td>
<td>3.724E-01</td>
<td>3.319E-01</td>
<td>5.100E+01</td>
<td>4.364E+01</td>
</tr>
<tr>
<td>5.5</td>
<td>3.385E-01</td>
<td>3.883E-01</td>
<td>6.171E+01</td>
<td>5.920E+01</td>
</tr>
<tr>
<td>6.0</td>
<td>3.103E-01</td>
<td>4.972E-01</td>
<td>7.344E+01</td>
<td>8.018E+01</td>
</tr>
<tr>
<td>6.5</td>
<td>2.864E-01</td>
<td>6.854E-03</td>
<td>8.619E+01</td>
<td>7.342E+01</td>
</tr>
<tr>
<td>7.0</td>
<td>2.660E-01</td>
<td>1.342E-01</td>
<td>9.996E+01</td>
<td>1.041E+02</td>
</tr>
<tr>
<td>7.5</td>
<td>2.482E-01</td>
<td>3.829E-01</td>
<td>1.147E+02</td>
<td>1.283E+02</td>
</tr>
<tr>
<td>8.0</td>
<td>2.327E-01</td>
<td>3.025E-01</td>
<td>1.305E+02</td>
<td>1.323E+02</td>
</tr>
<tr>
<td>8.5</td>
<td>2.190E-01</td>
<td>1.864E-01</td>
<td>1.473E+02</td>
<td>1.558E+02</td>
</tr>
<tr>
<td>9.0</td>
<td>2.068E-01</td>
<td>2.565E-01</td>
<td>1.652E+02</td>
<td>1.646E+02</td>
</tr>
<tr>
<td>9.5</td>
<td>1.960E-01</td>
<td>7.866E-02</td>
<td>1.841E+02</td>
<td>2.083E+02</td>
</tr>
<tr>
<td>10.0</td>
<td>1.862E-01</td>
<td>6.362E-02</td>
<td>2.040E+02</td>
<td>2.150E+02</td>
</tr>
</tbody>
</table>

This is illustrated by figure 5.1, where the normal probability plot of the residuals is shown. In such a plot, deviation from the diagonal indicates deviation from the normal distribution. A similar situation is encountered when outliers are present, but unlike with outliers, in this case it is not possible to ignore the offending points. Weighted least squares enable one to partially compensate for the discrepancy from the normal distribution by correctly choosing the values of the weights $V$. However, such a precise tailoring of the objective function often requires sophisticated knowledge of the system under investigation.

A criterion which allows a maximum likelihood estimation without the need for prior information was introduced by Box and Draper [1965]. This is discussed in the following section. So far this criterion, commonly used for algebraic models, has not been widely used for dynamic estimation problems. This is due to the form of the objective function which prevents a precise calculation of the gradients by
finite differences. This criterion is implemented in the code GREG used for the parameter estimations carried out in the previous chapters.

Numerically, the evaluation of the estimation criterion requires the integration of the system with the current values of parameter estimates $\hat{\theta}_n$. The values of the states at the sampling points are subsequently used to compute the residuals. Similarly, the evaluation of the gradients of the estimation criterion requires to perturbate of the parameters and the calculation of the perturbed state trajectories. Optimal values of the perturbation value $\delta \hat{\theta}$ are suggested by Bard [1974]. This approach works satisfactorily when least squares are used as the estimation criterion and makes it relatively easy to adapt a least squares algorithm for dynamic estimation with the use of a differential equations solver. Such a simple calculation of the gradients is however ineffective when the estimation criterion is the Box-Draper one. In this case the more complex objective function requires a different approach.
5.2 Multiresponse parameter estimation in dynamic models. An application of the Box-Draper criterion.

Let us consider a \( m \) dimensional dynamic model described by a system of mixed algebraic and differential equations:

\[
f(y', y, t, \theta) = 0 \]
\[
y(t_0) = y_0
\]

\( n \) measurements are taken at discrete sampling points \( t_{sp} \). \( Y \) is defined as the \( m \times n \) matrix of measured responses. \( H(\hat{\theta}) \) is a \( m \times n \) matrix of expected responses defined as follows:

\[
H(\hat{\theta}) = \begin{pmatrix}
\hat{y}_{11} & \hat{y}_{12} & \cdots & \hat{y}_{1m} \\
\hat{y}_{21} & \hat{y}_{22} & \cdots & \hat{y}_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{y}_{n1} & \hat{y}_{n2} & \cdots & \hat{y}_{nm}
\end{pmatrix} = \begin{pmatrix}
\hat{y}_1^T \\
\hat{y}_2^T \\
\vdots \\
\hat{y}_m^T
\end{pmatrix}
\]

(5.5)

where the \( \hat{y}_{ij} \)'s are the states as predicted by 5.4 with the present parameter values.

The matrix of residuals \( Z(\hat{\theta}) \) is given by:

\[
Z(\hat{\theta}) = Y - H(\hat{\theta}) = \begin{pmatrix}
\epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1m} \\
\epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\epsilon_{n1} & \epsilon_{n2} & \cdots & \epsilon_{nm}
\end{pmatrix} = \begin{pmatrix}
\epsilon_1^T \\
\epsilon_2^T \\
\vdots \\
\epsilon_m^T
\end{pmatrix}
\]

(5.6)

Box and Draper [1965] used the matrix \( Z(\hat{\theta}) \) to define the criterion for multivariate parameter estimation below:

\[
\min \Phi(\hat{\theta}) = |Z^T Z|
\]

(5.7)

Criterion 5.7 was, in the original paper, derived from a bayesian approach, However, Bard [1974] shows that a similar result is achieved with a likelihood approach. Criterion 5.7 corresponds to a maximum likelyhood estimator, but, unlike criterion 5.1, it does not require any information other than the experimental measurements and the model. In statistical terms, it allows likelihood estimation with a non informative prior. Before the parameter estimation is carried out, one should check if the available data set is suitable for unbiased estimation. Box and al. [1973] have shown that, if any linear dependency is present in the data set, the matrix \( Z^T Z \) is
singular. A preliminary check with the initial guess of $\hat{\theta}$ should be carried out. An effective detection of dependency is performed by inspecting the singular values of $Z^T Z$. If any singular value is very close or below the machine round off error, linear dependency is likely. When such kind of dependency is detected, new unbiased data should be collected. If this is impossible, the offending column(s) should be dropped from $Y$ and $H$. Modification of the model may also be necessary. McLean and al. [1979] have shown that there are a few cases where $Z^T Z$ is singular even though no dependency exists in the data set. However, such cases are not very general and can be neglected in this context. An extensive review of the use of the criterion with algebraic kinetic models is reported by Ziegel and Gorman [1980].

Given any estimator $\Phi$, Bard [1974] shows that the gradient with respect to the parameter estimate $\hat{\theta}$ is:

$$\frac{\partial \Phi}{\partial \hat{\theta}} = \sum_{k=1}^{n} \frac{\partial \Phi}{\partial e_k} \frac{\partial e_k}{\partial \hat{\theta}}$$

(5.8)

where $e_k = y_k - \hat{y}_k$ is the vector of residuals at time $t_k$. Equation 4.8 can be rewritten as:

$$\frac{\partial \Phi}{\partial \hat{\theta}} = -\sum_{k=1}^{n} \frac{\partial \Phi}{\partial e_k} \frac{\partial y(t_k)}{\partial \theta} = -\sum_{k=1}^{n} \frac{\partial \Phi}{\partial e_k} Q(t_k)$$

(5.9)

where $Q(t_k)$ is the sensitivity matrix computed at time $t_k$. It can be observed that the derivative $\partial y(t_k)/\partial \theta$ in the second term of 5.9, for an algebraic model are equal to $\partial f/\partial \theta$. This can be easily computed by finite differences. The use of the Box-Draper criterion with algebraic models does not require the user to supply any extra information than those necessary for least squares estimations.

Analogously to the design criterion 2.18, the optimum of the estimator 5.7 is not affected if a transformation is carried out:

$$\Psi = \log |Z^T Z|$$

(5.10)

Equation 5.10 makes it possible to use most of the results of section 2.6.2 in evaluating the objective function and gradient. The derivatives of $\Psi$ with respect to each residual are computed as:

$$\frac{\partial \Psi}{\partial \epsilon_{kl}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial z_{ij}}{\partial \epsilon_{kl}} \left[ (Z^+)^T \right]_{ij} \quad l = 1 \ldots m$$

$$\frac{\partial \Psi}{\partial \epsilon_{kl}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial z_{ij}}{\partial \epsilon_{kl}} \left[ (Z^+)^T \right]_{ij} \quad k = 1 \ldots n$$

(5.11)
\[ [(Z^+)^T]_{ij} \] indicates the \( ij^{th} \) element of the pseudoinverse of \( Z \). The \( \partial z_{ij}/\partial \epsilon_{kl} \) can be easily computed by finite differences perturbing the residuals. No extra integrations of the dynamic system are required to compute the gradient 5.8.

Bates and Watts [1984, 1985] computed the gradients \( \partial \Psi/\partial \theta \) according to equation 2.18, skipping the chain rule derivation of equations 5.8 and 5.9. In this way they avoided the calculation of the sensitivity coefficients. This is convenient if the sensitivity equations need to be explicitly specified, but each optimization step would require \( p + 1 \) integrations of the dynamic system.

The SQP algorithm used for the design of the experiments can be effectively adopted to solve the optimization problem posed by the minimisation of \( \Psi \). Unlike the previous chapter, the full matrix routine SQPD could be modified in order to supply statistical information on the parameter estimation. At the end of the \( r^{th} \) iteration of the optimiser, the current values of the parameters are updated by adding an increment \( \delta \theta^r \):

\[
\hat{\theta}^{r+1} = \hat{\theta}^r + \delta \hat{\theta}^r
\]

the newly evaluated vector of parameter estimates is then passed to the calling routine for new function and gradient evaluations. Before that, it is necessary to check that the increment \( \delta \hat{\theta}^r \) does exceed the statistical variability of the estimate \( \hat{\theta}^r \). If such a condition is not satisfied, the optimisation can be stopped, regardless of the fact that other iterations are numerically required. Such a statistical variability is defined in terms of posterior probability, which is proportional to the estimation criterion. The 95\% posterior probability is:

\[
P(\theta|Y) \propto |Z^T Z|^{-n/2}
\]

Using a linearisation of 5.13 about the parameter estimate \( \hat{\theta}^r \), one obtains:

\[
P(\theta|Y) \propto |Z(\hat{\theta}^r)^T Z(\hat{\theta}^r)|^{-n/2} \left\{ 1 + \frac{(\theta - \hat{\theta}^r)^T \Gamma(\theta - \hat{\theta}^r)}{2 |Z(\hat{\theta}^r)^T Z(\hat{\theta}^r)|} \right\}^{-n/2}
\]

Where \( \Gamma \) is the Hessian of \( \Psi \) computed at \( \hat{\theta}^r \). Such a Hessian is not the one computed by the optimization routine, as the quadratic programming routine augments the objective function to take in account the constraints. However, \( \Gamma \) can be extracted.
from the Hessian computed in the SQP routine and used to compute the statistical termination criterion:

\[
\frac{\|C\delta\hat{\theta}\|^2}{2ps^2} \leq \epsilon
\]  

(5.15)

where \(\epsilon\) is a tolerance and \(s^2\) is a scale factor computed as:

\[s^2 = \frac{|Z^TZ|}{\nu}\]  

(5.16)

and \(\nu = n - p\) is the number of degrees of freedom of the estimation problem. The matrix \(C\) is the Choleski factor of the Hessian \(\Gamma\). The termination criterion 5.15 has been introduced by Bates and Watts [1981], who also suggested that the value \(10^{-3}\) is, in many cases, appropriate as tolerance.

If the numerical criteria are satisfied first, the optimization can be terminated as usual. The knowledge of \(\Gamma\), however, is still necessary to compute the variance covariance matrix of parameter estimates:

\[W(\hat{\theta}) = 2s^2\Gamma^{-1}\]  

(5.17)

In order to compare designs or the level of correlation of different parameters, the correlation matrix should be computed. The elements of the correlations matrix are:

\[r_{ij} = \frac{w_{ij}}{\sqrt{w_{ii}w_{jj}}}\]  

(5.18)

where the \(w_{ij}\)'s are the elements of the variance covariance matrix \(W(\hat{\theta})\) computed by equation 5.17.

5.2.1 A new algorithm for parameter estimation in nonlinear dynamic models.

The analysis of the previous section leads to the definition of an algorithm for parameter estimation in nonlinear dynamic models:

1. Integrate the dynamic system storing the expected values of the states and the sensitivity coefficients at the sampling points.

3. Check the linear dependency in the data set by inspection of the singular values of \(Z^T Z\). If any singular value exceeds a given tolerance, the problem needs to be reformulated. If no linear dependency is found, the result of the singular
value decomposition is stored and used for the first evaluation of the gradient of the objective function.

4 - Compute the objective function using the QR decomposition of Z and the transformed objective function (equation 2.35).

5 - Compute the pseudoinverse $Z^+$ by singular value decomposition of Z.

6 - Calculate the gradient of the objective function (equations 5.8 and 5.9).

7 - Call the optimization routine.

8 - Compute the Hessian $\Gamma$ by extraction of the relevant information from the BFGS estimation computed by the SQP routine. (1)

9 - If the numerical conditions for an optimum are satisfied go to step 13.

10 - Compute the Choleski decomposition of $\Gamma$ and check the statistical significance of the new estimates (equation 5.15).

11 - If the statistical termination criteria are not satisfied integrate the system with the new value of $\hat{\theta}$.

12 - Update Z and go back to step 4.

13 - Calculate the variance covariance matrix of the parameter estimates (equation 5.17) and the correspondent correlation matrix (equation 5.18).

The appealing property of this algorithm is that only one integration of the dynamic system is required for each iteration of the optimizer. The gradient calculation appears to be quite simple as well. Moreover, it is evident that such an algorithm could be relatively easily implemented along with the dynamic design algorithm introduced in chapter 3, since they share many basic numerical routines.

A similar algorithm was implemented by Bates and Watt [1984], however, they do not take advantage of the sensitivity coefficients and, as an optimization strategy, they use an unconstrained Gauss-Newton method. Not using the sensitivity coefficients, they limit their approach to linear systems.

(1) The feasibility and the accuracy of this operation should be more carefully assessed during the implementation of the algorithm.
To complete the algorithm described in this section, it should be necessary to take into account possible missing observations (missing elements in the matrix $Y$). An efficient method to do so has been proposed by Steward and Sorensen [1981].

5.3 Critical remarks on the design of experiments to improve parameter estimates.

Although the results of the previous two chapters convey an overall positive feeling about the use of nonlinear experimental design, some extra comments and summarising remarks are necessary.

In the case of algebraic models, optimal experimental points were mostly found lumped together in the upper corner of the experimental range. This feature was expected, as it has been repeatedly observed before, and it is typical of sequential design. Many experimenters have criticized such a behaviour as scarcely intuitive from a physical point of view. In particular, it could happen that if the model is not perfectly adequate to describe the process, such an uneven distribution of experimental points does not allow an accurate determination of possible lack of fit. From a rigorous theoretical point of view, the D-optimal sequential design procedure should be applied only to models whose adequacy has been tested. Inadequate models should have been dropped in a preceding experimental phase. However, budgetary reasons often limit the overall amount of available experiments and only a less than accurate screening of possible models can often be carried out. Moreover in most cases there are no competitive models to choose from and the model building procedure takes places along with the experiments. In both the previous cases, D-optimal sequential design, by lumping the experimental points deprives the experimenter of the feedback arising from the qualitative observation of how the dependent variables change. In this case the experimenter might find more informative some static and a priori design such as factorial design. Such designs yield experimental points sparsely disseminated in the experimental range and the loss of performance, compared to a sequential design, is often relatively low as shown in chapter 2. Such designs are also quite economical as they do not require any extra computation.

From a practitioner point of view, this is the main objection to D-optimal sequential design. However, if the lumping of experimental points is acceptable,
the D-optimal remains a design criterion which conjugates a very high effectiveness with, by far, the largest ease of use. The main statistical objection to the use of the D-optimal design resides in the increase in the correlation terms which takes places simultaneously to the decrease in the intervals of confidence. This has been observed in chapter 2, however not to such an extent to justify the extra effort associated with the use of the E-optimal criterion.

The observations above are mostly valid for the dynamic case as well. In this case the lumping of the experimental points is shown by the grouping of the sampling points about peaks in the value of the observation matrix. This leads to results which appear somehow not intuitive as in the case of the fed batch bioreactor of section 3.6. In this example, (figure 3.9) a vast number of sampling points are concentrated where the behaviour of the system is fairly stationary. However in such a case, the overall value of the information matrix is relatively higher. The fact that the constraints on the minimum distance between adjacent sampling points is always active in such groups, shows that the tendency is to repeat a number of samplings at the same time. This would cause the loss of dynamic information and might appear to contradict the idea of dynamic experiments. It should be borne in mind, however that the design algorithm of section 3.3 deals with the process as if it was a static one. The difference consists in the way the elements of the information matrix are computed. Object of future investigation would be to assess in more detail the implications of such behaviour. This will first require the availability of a sound estimation algorithm as the one of section 5.2.1.

In both dynamic and steady state design, there is an evident asymptotic behaviour. In the cases of chapters 2 and 3, it can be observed that the reduction in the intervals of confidence is mostly concentrated in the first two or three steps of the design. Subsequent reductions are comparable with those achieved by unplanned experimentation and therefore due to the extra experiments rather than their optimality.

The obvious question one would now ask is whether the results obtained by such techniques are worth the extra cost associated with them. When the problem is tackled, as in this work, from a practical point of view, the answer is not clear cut.
In the steady state case there are obvious advantages and the use of the code is relatively simple, therefore, a more widespread usage of such techniques should be advocated. In the dynamic case, such advantages are not as obvious and the results obtained thus far, although promising, are preliminary and not complete. The technique is not mature for application to real world problems until more detailed theoretical investigations are carried out.

5.3.1 Future work on sequential design of dynamic experiments.

Any future work in optimal design of deterministic models, should require a robust implementation of the estimation algorithm of section 5.2.1. Such an algorithm would use an improved version of the present DAE solver DASOLV and could be implemented together with the design algorithm of section 3.3, as many basic numerical routines can be shared. Since the estimation algorithm uses a Box–Draper objective function with unknown prior, it will be possible to have cycles of parameter estimation and subsequent design of the next experiment without intermediate user’s intervention.

Before such a stage can be reached, an extensive investigation on the effect on the sensitivities trajectories of variations of the controls and switching times should be carried out. This is necessary in order to enable efficient and reliable computation of the partial derivatives of the elements of the information matrix.

Investigation on an efficient method to compute the gradients of the objective function expressing an E-optimal criterion should be tackled. This would expand the capability of the algorithm with respect to both stationary and dynamic design. This is desirable as it would allow the change of the design criterion when the D-optimal one does not achieve any considerable improvement. Moreover, the information matrix for the dynamic case should be improved possibly as discussed in section 3.7, in order to accommodate true sequential design.

5.4 Future areas of work in experimental design and mechanistic model building.

So far model building has been mainly an activity separate from experimental design. Also in the design for model discrimination, different, concurrent models are proposed and used to design experiments. The statistical techniques used allow
the experimenter to reject a model if inaccurate, but supply little or no information on how the model could be improved or about its shortcomings.

The way in which experimenters design experiments has an heuristic aspect which, thus far, did not find a translation in the statistical techniques which are commonly used. Many experimenters still rely on empirical considerations when planning experiments and find the statistical criteria hard to accept because of their lack of physical feeling. Moreover the paramount constraint encountered in planning experimental investigations, is often the allocated budget. The necessity of studying both qualitative and quantitative aspects of a process in a limited number of experiments could be more important than the necessity to satisfy optimal criteria. These problems should be tackled by future research in the area.

In the foreseeable future, the process of model building and experimental design will still need considerable operator intervention and it cannot be predicted if and when it could be totally automated. However the number of separate computational tools available to the experimenter should increase. The present statistical algorithms can still be improved in both robustness and ease of use. Much of the future effort should be targeted to create new tools and techniques for model building. Some areas which might achieve successfull application are listed below:

1 – Tendency models. They aim to build a kinetic model from a framework of possible reaction paths and some experimental information. A tendency model is an approximate kinetic network, which lumps unknown or unmeasured states together with a set of mass and energy balances [Filippi-Bossi and al., 1986, 1987]. Sets of feasible reaction paths are modeled as systems of algebraic equations which permit the determination of approximate rate constants. Tendency models have been used when little or no knowledge is available on the system and a preliminary mechanistic model is being developed out of suboptimal experimental data or industrial plant runs. This approach so far has been applied to homogeneous reactions in liquid phase.

2 – Thermodynamically feasible paths. This approach was initially developed to study the possible states of an heterogeneous catalytic reaction [Shinnar and Feng, 1985]. Subsequently it has been applied to more general reacting systems [Shinnar, 1988]. The energetic levels of different reaction paths are different,
and only the one allowing the maximum reduction of free energy is feasible. This permits a deterministic discrimination criterion among models.

3 – Lumping of states. In many complex kinetic networks only few components are of interest. These usually have concentrations much higher than the byproducts. Moreover, in industrial runs, the concentrations of such byproducts are not measured. However, the byproducts cannot be ignored when modeling the systems. Lumping techniques should allow to lump all these unknown states in one or more pseudocomponents which can be more easily modeled. In doing so, however, thermodynamic and physical considerations should be taken into account, in order to avoid the lumping of incompatible species. So far lumping techniques have been mainly used in modeling cracking processes where hundreds of species, many only slightly different, are present.

The previous list is not intended to be exhaustive, but only to point out some of the directions in which future research may lead.
References


Atkinson A.C., [1966], “Statistical design for pilot plant and laboratory experiments – Part I”, Chemical Engineering, 73, 149-154.

Atkinson A.C., [1981], “A comparison of two criteria for the design of experiments for discriminating between models”, Technometrics, 23, 301-305.


Box G.E.P., Hill W.J., [1967], “Discriminating among mechanistic models”, Technometrics, 9, 57-71


Galil Z., Kiefer J., [1980], "Time and space saving computer methods related to Mitchell's DETMAX, for finding D-optimal design.", Technometrics, 22, 301–313.


Hindmarsh A.C., [1980], “LSODE and LSODI, two initial value ordinary differential equation solvers”, ACM SIGNUM Newsletter, 15, 10-11.


Steward W.E., Sorensen J., [1981], “Bayesian estimation of common parameters from multiresponse data with missing observations”, Technometrics, 23, 131


Welch W.J., [1982], “Branch and bound search for experimental design based on D-optimality and other criteria.”, Technometrics, 24, 41–48.


Appendix A

On the use of the optimal control approach to design experiments for model discrimination.

Unlike design to improve parameter estimates, the design for model discrimination can be effectively tackled by the optimal control approach. The design criterion chosen is the Hunter-Reiner criterion. Let us suppose two models are being examined:

\begin{align*}
    y'_1 &= f_1(y_1, u(t), t, \theta_1) \\
    y'_2 &= f_2(y_2, u(t), t, \theta_2)
\end{align*}

(A.1)

where \( u(t) \) is a common set of controls. The two models constitute a supersystem which is used by VOPCON as a single set of mixed differential and algebraic equations. In this case the objective function can be defined as:

\[
    \sum_{i=1}^{n} \| y_{i1} - y_{i2} \|
\]

(A.2)

\( n \) being the number of variables considered. The distance between the points is defined by an appropriate, usually Euclidean, norm. Maximisation of equation A.2 yields the maximisation of the area bounded by couple of equivalent trajectories. It should be however remarked that the set of controls \( u(t) \), must be the same in both models. Moreover if the number of state variables in the two models differs, it may arise the problem of arbitrarily decide which states of the larger models are not to be included in the objective function.
The need for an objective function analytical in the state variables in the optimal control problems

The necessity to define the objective function as analytical in the state variables is shown. For sake of brevity a reduced objective function is considered, the terminal time $t_f$ is kept fixed and no terminal constraints are imposed. Let us consider the dynamic model:

$$y' = f(y(t), u(t), t)$$
$$y(t_0) = y_0$$  \hspace{1cm} (B.1)

where $u(t)$ is a vector of controls. The aim is to find $u(t)^*$ so that the following objective function, or performance index, is maximized:

$$J = \int_{t_0}^{t_f} L(y(t), u(t), t) \, dt$$  \hspace{1cm} (B.2)

$u(t)$ being a vector of controls for the system. The objective function is adjoined [Bryson and Ho, 1975] as follows:

$$\overline{J} = \int_{t_0}^{t_f} \{ L(y(t), u(t), t) + \lambda(t)[f(y(t), u(t), t) - y'] \} \, dt$$  \hspace{1cm} (B.3)

where $\lambda(t)$ is a real vector of arbitrary variables. The elements of $\lambda$ are called the Lagrange multipliers. The argument of the integral in equation B.3 defines the Hamiltonian function $H(y(t), u(t), \lambda(t), t)$. Solving the integral in equation B.3 by parts, one obtains:

$$\overline{J} = \lambda(t_0)y(t_0) - \lambda(t_f)y(t_f) + \int_{t_0}^{t_f} [H(y, u, \lambda, t) + \lambda'y] \, dy$$  \hspace{1cm} (B.4)

If an arbitrary variation $\delta u$ on the control is considered, the following variation is induced in the adjoint objective function B.4

$$\delta \overline{J} = -\lambda(t_f)\delta y(t_f) + \int_{t_0}^{t_f} \left\{ \left[ \frac{\partial H}{\partial y} + \lambda'y \right] \delta y + \frac{\partial H}{\partial u} \delta u \right\} \, dt$$  \hspace{1cm} (B.5)
Being $\lambda(t)$ arbitrary, it can be chosen so that the coefficients of $\delta y$ would vanish. This is achieved by imposing:

$$
\lambda'(t) = -\frac{\partial H}{\partial y} - \frac{\partial L}{\partial y} - \lambda \frac{\partial f}{\partial y}
$$

\hspace{1cm} (B.6)

$$
\lambda(t_f) = 0
$$

If (B.6) is satisfied the variation (B.5) reduces to:

$$
\delta \overline{J} = \int_{t_0}^{t_f} \frac{\partial H}{\partial u} \delta u dt
$$

However, condition for an extremum is $\delta \overline{J} = 0$ for any $\delta u$, such a condition requires the satisfaction of the following:

$$
\frac{\partial H}{\partial u} = \lambda \frac{\partial f}{\partial u} + \frac{\partial L}{\partial u} = 0
$$

\hspace{1cm} (B.8)

The vector $u(t)$ simultaneously satisfying (B.6) and (B.8) is the solution. Equations (B.6) and (B.8) are called Euler-Lagrange equations.

The code VOPCON [Morison, 1984] integrates forward the system equations (B.1), stores the trajectory and, subsequently, integrates backward the adjoint system (B.6). States and multipliers trajectories are then used to verify the satisfaction of (B.8). It appears, that, in order to solve the adjoint system (B.6), the objective function $L(x, u, t)$ needs to be analytical in the independent variables $y$.

In the optimal control approach to the design of dynamic experiments, the system (B.1) is defined as the supersystem made up by the state equations and the sensitivity equations. If the hypothesis of diagonal $W(\hat{\theta})$ holds, then the design criterion can be reduced to an analytical function of the sensitivity coefficients, as seen in chapter 3. If the full matrix is considered, the original objective function $|\int_{t_0}^{t_f} Q(t)^T Q(t) dt|$ cannot be simplified as done by Espie [1986] and no suitable expression is found.
Calling statement and user instructions of the routine PRODES.

SUBROUTINE PRODES (DMPAR, DM2PAR, NLCR, XEX, PAR, XC, XI, WE, NEQ, 
+ NP, NLC, MEQ, PNAME, WSQ, LWSQ, WKS, LWKS, 
+ IWK, LIWK, JDERR)

C PURPOSE ----------------------------------------------- -------
C Subroutine to design D-Optimal experiments by selection of an
C experimental grid maximizing the determinant of the information
C matrix. The design variables can be subject to simple bounds
C inequality constraints and/or equality constraints. However
C the use of equality constraints is not advised as it may lead
C to highly correlated independent variables. The program handles
C monovariate or multivariate models described by algebraic
C equations.

C FORMAL ARGUMENT -------------------------------------------------
C ARGUMENT TYPE USAGE
C
C DMPAR external subroutine containing the sensitivities
C of the model. To be declared external in
C the calling program
C
C DM2PAR external subroutine containing the second derivatives
C of the model wrt the parameter. To be
C declared external in the calling program
C
C NLCR external subroutine containing inequality and/or
C equality constraints. To be declared external
C in the calling program
C
C XEX real*8 XEX(NE, NP) array containing the experimental
C grid. On exit it contains either the solution
C or the point at the failure.
C
C PAR real*8 PAR(NP) vector containing the current value
C of parameters. Unchanged by the program.
C
C XC real*8 XC(NI, 2) array containing the simple bounds
C for the independent variables. These should
C satisfy the conditions:
C XC(i,1) <= XEX(i,j) <= XC(i,2)
C j = 1, NE
C However no equality constraints in the form
C XC(i,1) = XEX(i,j) = XC(i,2)
C are accepted.
C
C NI integer Number of independent variables in the model
C
C NE integer Total number of experiments
NEO integer
Number of experiments to be optimized, it should satisfy: NEO <= NE. If NEO < NE, only the last NEO rows of XEX are affected by the design.

NP integer
Number of parameters in the model

NEQ integer
Number of equations in the model. It can be left non initialized if the model is mono-variate.

NLCS integer
Number of inequality constraints
- NOT SIMPLE BOUNDS -

MEQ integer
Number of equality constraints. It should be MEQ <= NLCS

PWAME char*20
String containing a short description of the problem

WSQ real
WSQ(LWSQ) real workarea required by the optimiser

LWSQ integer
length of WSQ.
LWSQ >= (10*NI*NEO+30)*((NI*NEO)+1)

WKS real
WKS(LWKS) real workarea required by PRODES

LWKS integer
length of WKS
LWKS >= 7(NP*NP)+NP(4+7*NEO)+2(ME*NP)+4*(NI*NEO)+NP*NI+NI+NEO+3*MAX(NLCS,1)*
(4+NI+NI*NEO)

IWK integer
IWK(LIWK) integer workarea required by PRODES

LIWK integer
length of IWK
LIWK >= 10*NI*NEO+10

JDER integer
flag to indicate the calculation method for the second derivatives of the sensitivities (monovariate case only):
1 --> use analytical supplied by the user in the subroutine DM2PAR
0 --> numerical derivatives using a forward difference formula
Do not initialize when multivariate case is considered.

USER SUPPLIED SUBROUTINE ---------------------------------------------
MONOVARIATE CASE

The design matrix is generated in the subroutine DMPAR which returns the derivatives of the model wrt parameter in the form
DPAR(J) = df(X(1)...X(NI))/dPAR(J)
where X is the vector of independent variables at the n-th experiment.

SUBROUTINE DMPAR (X,NI,NP,PAR,DPAR)
IMPLICIT DOUBLE PRECISION (A-H,O-Z), INTEGER (I-N)
DIMENSION X(NI), PAR(NP), DPAR(NP)

DPAR(1) = ....
DPAR(2) = ....
............
DPAR(NP) = ....
RETURN
END

The routine DM2PAR contains the second derivatives of the design matrix as: D2PAR(J, I) = \frac{\partial^2 f}{\partial X(I) \partial PAR(J)}.

SUBROUTINE DM2PAR(D2PAR, NI, NP, X, PAR)
IMPLICIT DOUBLE PRECISION (A-H, O-Z), INTEGER (I-N)
DIMENSION D2PAR(NP, NI), X(NI), PAR(NP)
D2PAR(1, 1) = ....
..............
D2PAR(NP, NI) = ....
RETURN
END

This subroutine is needed only if JDER = 1

MULTIVARIATE CASE

The design matrix is generated in the subroutine DMPAR which returns the derivatives of the model wrt parameter in the form:

DPAR(I, J) = \frac{\partial f}{\partial X(I) \partial PAR(J)}

where X is the vector of independent variables at the n-th experiment.

SUBROUTINE DMPAR(X, NI, NP, NEQ, PAR, DPAR)
IMPLICIT DOUBLE PRECISION (A-H, O-Z), INTEGER (I-N)
DIMENSION X(NI), PAR(NP), DPAR(NEQ, NP)
DPAR(1, 1) = ....
DPAR(1, 2) = ....
..............
DPAR(NEQ, NP) = ....
RETURN
END

BOTH CASES

The constraints are supplied along with the relevant gradients in the subroutine NLCR. The array DCON contains the derivatives of the constraints wrt the design variables in the form:

DCON(j, i) = \frac{\partial C_i}{\partial X(j)}

SUBROUTINE NLCR(X, CON, DCON, NCON, DCON, CON, NI, PAR, NP, IDER)
IMPLICIT DOUBLE PRECISION (A-H, O-Z), INTEGER (I-N)
DIMENSION X(NI), CON(NCON), DCON(NDCON, CON), PAR(NP)
IF (IDER .LE. 0) THEN
CON(1) = ....
CON(2) = ....
........
CON(NCON) = ....
ENDIF
IF (IDER .NE. 0) THEN
DCON(1, 1) = ....
......
DCON(NCON, NDCON) = ....
If equality constraints are present, they should occupy the first MEQ locations in the vector CON.

**COMMON BLOCKS**

Further control and/or information on the execution can be obtained using a common blocks:

<table>
<thead>
<tr>
<th>BLOCK NAME</th>
<th>VARIABLE</th>
<th>PURPOSE</th>
<th>DEFAULT VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLOUT</td>
<td>LOUT</td>
<td>default output unit</td>
<td>6</td>
</tr>
<tr>
<td>CDEBUG</td>
<td>IDEBUG</td>
<td>amount of diagnostic output (0,1,2,3)</td>
<td>0</td>
</tr>
<tr>
<td>UTIL</td>
<td>ICALL</td>
<td>counter</td>
<td>n/a</td>
</tr>
<tr>
<td>CSTEP</td>
<td>STEP</td>
<td>perturbation value in numerical differentiation</td>
<td>1e-7</td>
</tr>
<tr>
<td>CVAR</td>
<td>VARI</td>
<td>variance covariance matrix of measurement error</td>
<td>Identity matrix</td>
</tr>
<tr>
<td>CSVD</td>
<td>TSVD</td>
<td>tolerance for singular value rejection</td>
<td>TINY</td>
</tr>
<tr>
<td>CSQP1</td>
<td>IOP</td>
<td>User defined optimizer options</td>
<td>0 (no)</td>
</tr>
<tr>
<td>CSQP2</td>
<td>ACC</td>
<td>required tolerance for the optimizer</td>
<td>1e-5</td>
</tr>
</tbody>
</table>

When applicable more detailed information is available in comments in the code. Other common blocks are for internal use only.

**OTHER**

Machine dependent quantities EPS, TINY, HUGE are computed initially by call to function MACH. The values are stored in the common block CMACH.

**KNOWN BUGS**

1) In the monovariate case the program does not handle the case NEO=NE
2) In the multivariate case the variance-covariance matrix can be only supplied by means of an external unformatted file which should be written by an external program. No means to compute such a matrix from a set of previous experimental data are provided.

3) A few subroutines use local arrays of fixed size which do not belong to the global workspace supplied to the user. This might require to resize those array and recompile the code if the dimensions of the problem exceed those supplied. If this happens error messages are issued.

4) The subroutines DATE and TIME are VAX-Fortran specific.

LANGUAGE ----------------------------------------------
FORTRAN 77

Tested with the following compilers/operating systems:

VMS Fortran 5.1 / VMS 5.3
Sun Fortran / SunOs 4.0
Lahey F77L3 / MsDos 3.30 + OS386 Dos extender
MS-Fortran 4.01 / MsDos 3.30

REFERENCES ---------------------------------------------

i) BOX G. E. P, Lucas H. L. "The design of experiments in nonlinear situations"
   BIOMETRIKA, 49, 77-80, (1959)

ii) Bates D. "The derivative of |X'X| and its use"
    TECHNOMETRICS, 25, 373-376, (1983)

iii) Chen C. L. "A class of successive quadratic programming methods
     for flowsheet optimization"

---------------------------------------------------------------------
Coded by : Luca Zullo
At : IRC for Process Systems Engineering
     Dept. of Chemical Engineering
     Imperial College
     London, 1990

---------------------------------------------------------------------
This version dates: 25 June 1990.
Note on the numerical derivatives scheme used for PRODES.

The numerical derivatives necessary to compute the gradients reported in equations 2.36 and 2.37 are calculated using a variable step forward difference scheme:

\[
\frac{\partial f(x)}{\partial x_i} \approx \frac{f(x + e \delta_i) - f(x)}{\delta_i}
\]  

(D.1)

where \( e \) is a vector with all zero elements but the \( i \)th element which has a value of one. The variable step size \( \delta_i \) is chosen as follows:

\[
\delta_i = \max(\epsilon^2, |\delta_{\text{abs}} x_i|, \delta_{\text{abs}})
\]

(D.2)

where \( \epsilon \) is the machine precision and \( \delta_{\text{abs}} \) is an absolute step size. PRODES supplies a default value for \( \delta_{\text{abs}} \) (10\(^{-7}\)), however such a value can be overridden by the user. The default value, corresponding to the value recommended by other authors [Edgar and Himmelblau, 1988, page 232], has proved to be adequate for the case studies reported in chapter 2. Experiences of the users of PRODES [Bagatin, 1991] show that the optimal value of \( \delta_{\text{abs}} \) is problem dependent. In particular, when the range of variability of the design variables is wide, higher values may be necessary.

In solving the dynamic problems of chapter 3, \( \delta_{\text{abs}} \) was set at 10\(^{-5}\). Such a value has been satisfactory for the selection of the optimal sampling points.

Additional reference:
Further notes on the sensitivity analysis.

After the work in this thesis was finished, some corrections have been implemented in the DASOLV code which now allows correct calculation of the sensitivities for the case study presented in Chapter 4. The corrected trajectories are reported in the following tables E1.

<table>
<thead>
<tr>
<th>Time</th>
<th>$q_{11}$</th>
<th>$q_{21}$</th>
<th>$q_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5E-01</td>
<td>-0.492E-01</td>
<td>0.462E-01</td>
<td>-0.265E-02</td>
</tr>
<tr>
<td>5.0E-01</td>
<td>-0.935E-01</td>
<td>0.824E-01</td>
<td>-0.105E-01</td>
</tr>
<tr>
<td>0.1E+01</td>
<td>-0.168E+00</td>
<td>0.130E+00</td>
<td>-0.346E-01</td>
</tr>
<tr>
<td>0.2E+01</td>
<td>-0.274E+00</td>
<td>0.160E+00</td>
<td>-0.934E-01</td>
</tr>
<tr>
<td>0.3E+01</td>
<td>-0.334E+00</td>
<td>0.144E+00</td>
<td>-0.142E+00</td>
</tr>
<tr>
<td>0.4E+01</td>
<td>-0.361E+00</td>
<td>0.108E+00</td>
<td>-0.173E+00</td>
</tr>
<tr>
<td>0.5E+01</td>
<td>-0.367E+00</td>
<td>0.702E-01</td>
<td>-0.185E+00</td>
</tr>
<tr>
<td>7.5E+00</td>
<td>-0.328E+00</td>
<td>-0.528E-02</td>
<td>-0.166E+00</td>
</tr>
<tr>
<td>1.0E+01</td>
<td>-0.260E+00</td>
<td>-0.408E-01</td>
<td>-0.121E+00</td>
</tr>
<tr>
<td>1.25E+01</td>
<td>-0.193E+00</td>
<td>-0.495E-01</td>
<td>-0.809E-01</td>
</tr>
<tr>
<td>1.50E+02</td>
<td>-0.138E+00</td>
<td>-0.453E-01</td>
<td>-0.512E-01</td>
</tr>
<tr>
<td>1.70E+01</td>
<td>-0.103E+00</td>
<td>-0.385E-01</td>
<td>-0.347E-01</td>
</tr>
<tr>
<td>1.90E+01</td>
<td>-0.763E-01</td>
<td>-0.311E-01</td>
<td>-0.233E-01</td>
</tr>
<tr>
<td>2.00E+02</td>
<td>-0.652E-01</td>
<td>-0.276E-01</td>
<td>-0.190E-01</td>
</tr>
</tbody>
</table>

Table E1 - Corrected values for the sensitivities of model 4.13

Comparison of the values of table E1 with those of tables 4.1, 4.2 and 4.3 show that these values are now basically the same as those obtained by the DDF method. Further throughout testing of the revised code should however be undertaken to ensure that the sensitivities trajectories are always calculated correctly.
Appendix F

Nomenclature

Latin symbols

\(H(\hat{\theta})\)  
matrix of expected responses.

\(M\)  
information matrix.

\(m_{ij}\)  
\(ij^{th}\) element of \(M\).

\(n_a\)  
number of algebraic equations in a dynamic model.

\(n_c\)  
number of controls.

\(n_{ci}\)  
number of control intervals.

\(n_d\)  
number of differential equations in a dynamic model.

\(n_e\)  
number of experiments to be designed.

\(n_{eq}\)  
number of equations in a dynamic model \((n_{eq}=n_a+n_d)\).

\(n_{sp}\)  
number of sampling points.

\(n_{sw}\)  
number of switching points.

\(Q_{r}\)  
matrix of sensitivity coefficient of the \(r^{th}\) equation in the model \((eq.\ 3.17)\).

\(q_{ij}(t_i)\)  
sensitivity coefficient of the \(i^{th}\) state variable with respect to the \(j^{th}\) parameter computed at time \(t_i\) \((eq.\ 3.6)\).

\(Q(t_i)\)  
matrix of sensitivity coefficients of a dynamic model computed at time \(t_i\).

\(S\)  
design matrix.

\(S_n\)  
design matrix of row dimension \(n\).

\(S_r\)  
design matrix relative to the \(r^{th}\) equation in a multivariate model.

\(s_{ij}\)  
\(ij^{th}\) element of \(S\).

\(s_{r,ij}\)  
\(ij^{th}\) element of \(S_r\).

\(t_{\text{max}}\)  
maximum allowable final time for a dynamic experiment.

\(t_{sp_i}\)  
sampling point.

\(t_{sw_i}\)  
switching point.

\(u(t)\)  
vector of time dependent inputs into a dynamic system.

\(W(\hat{\theta})\)  
matrix of variance covariance of \(\hat{\theta}\).

\(w\)  
vector of time independent inputs into a dynamic system.

\(x\)  
vector of independent variables.
\( x(t) \) vector of time dependent algebraic state variables.
\( y \) vector of state variables.
\( y(t) \) vector of time dependent state variables.
\( \dot{y}, y(\dot{\theta}) \) vector of state variables as computed by the model using the parameter estimates \( \dot{\theta} \).
\( y_0, y(t_0) \) vector of initial conditions for a dynamic system.
\( Y \) matrix of measured responses.
\( z_{ij} \) \( ij^{th} \) element of \( Z \).
\( Z, Z(\dot{\theta}) \) matrix of residuals.

**Greek symbols**
\( \hat{e} \) residual \( y - \dot{y} \).
\( \Sigma \) matrix of variance covariance for the experimental error.
\( \hat{\Sigma} \) estimate of the matrix of variance covariance for the experimental error (eq. 2.13).
\( \sigma_{rs} \) \( rs^{th} \) element of \( \Sigma \).
\( \hat{\sigma}_{rs} \) \( rs^{th} \) element of \( \hat{\Sigma} \).
\( \sigma^{rs} \) \( rs^{th} \) element of \( \Sigma^{-1} \).
\( \hat{\sigma}^{rs} \) \( rs^{th} \) element of \( \hat{\Sigma}^{-1} \).
\( \theta \) vector of parameters.
\( \dot{\theta} \) vector of parameter estimates.
\( \dot{\theta}_n \) vector of parameter estimates as evaluated after \( n \) experiments.

**Superscripts**
\( L \) lower bound.
\( U \) upper bound.
\( + \) pseudoinverse.
\( T \) transpose.