A Protocol for the Estimation of Parameters in Process Models

by

Alojz Loui Polic

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemical Engineering

Waterloo, Ontario, Canada, 2001

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

As our understanding of chemical processes increases the models created to describe them also increase in complexity. These models usually consist of sets of differential equations, containing multiple response variables which are a function of multiple input or design variables and a potentially large number of parameters. In most cases the equations are nonlinear in the inputs and parameters, and must be solved by numerical integration. An example of such a model is the Watpoly polymerization model (Gao and Penlidis, 1996 and 1998) developed in the polymer research group at the University of Waterloo. To be able to use these models effectively the parameters that they contain have to be known.

The bulk of the literature dealing with parameter estimation has only considered small models. At present in estimating parameters for large process models, there are two shortcomings in the existing knowledge about parameter estimation. The first is, how effective is the present parameter estimation methodology when applied to large models, and the second is, can any advantage be gained from considering the parameter estimation problem as a whole. This work will try to address this limitation, by revisiting the parameter estimation process and developing a protocol for the estimation or updating of the parameters within process models.

The projected use of the parameter estimation protocol is as part of a model based experimentation program. Therefore it considers actual experimental conditions, where the number of experiments that can be carried out is limited due to the expense of performing experiments and analysis.

In the development of a parameter estimation protocol all of the steps of the parameter estimation will be revisited. The parameter estimation steps are: parameter sensitivity analysis, statistical design of experiments, estimation of parameters and confidence regions. Where these four steps correspond to answering the following questions;

1. Is it possible to estimate the parameters with the chosen responses and which response or responses will provide the most information?

2. At what conditions (i.e. temperature, conversion, initial feed composition) should the data in the experiment be collected?

3. What is the best method to estimate the parameters with the data that was collected?

4. How good are the parameters that were estimated?

Of the parameter sensitivity analysis methods available it was found that the best method to present sensitivity information is a plot of the gradient values of the responses with respect to the parameters. The gradients are normalized and plotted as a function

of independent variables such as initial feed composition, time or conversion. This is performed so that responses of different magnitudes and different measurement errors can be compared to each other.

In designing experiments the D-optimality criterion is generally used. One of the implementation challenges in designing experiments is local optima. One possible method found to deal with this difficulty, is to provide the optimization algorithm with a good initial guess that is based on the information provided by the gradient plots.

To estimate the parameters with multiple responses the Determinant criterion is used. When estimating multiple parameters in a large model a large number of local optima can exist. To overcome this difficulty, different approaches are available, such as a robust optimization algorithm (e.g. simulated annealing) or the use of multiple starting points.

Confidence regions of the parameter estimates will provide a measure of the quality of the parameter estimates. The true shape approximate level confidence regions were found to be an adequate compromise between information provided and computation required. It was found that the true shape joint confidence regions can be incorrect if multiple responses are used and the sample size is small.

The parameter estimation protocol is a series of actions or steps that can be followed in the course of obtaining parameter estimates. By following these actions the overall parameter estimation procedure can be more efficient and some pitfalls such as local optima and incorrect confidence regions, may be dealt with in an appropriate manner. To illustrate the application of the protocol, three case studies are presented. These case studies illustrate some of the problems that may be encountered in the parameter estimation process and how the proposed protocol can aid in overcoming them.

Acknowledgements

I would like to express my sincere gratitude and thanks to my supervisor Professor Tom Duever, and to Professor Alex Penlidis. As the completion of this thesis would not have been possible if it were not for the support and guidance that I received from them. In particular I would like to thank them for their patience with my struggles and their encouragement and support when things were not going well.

I would also like to thank my friends and colleagues at Waterloo, for their support and all of the good times.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

As our understanding of chemical processes increases the models created to describe them also increase in complexity. These models usually consist of sets of differential and algebraic equations, containing multiple response variables which are a function of multiple input or design variables and a potentially large number of parameters. In most cases the equations are nonlinear in the inputs and parameters, and must be solved by numerical integration. An example of such a model is the Watpoly polymerization model (Gao and Penlidis, 1996 and 1998) developed in the polymer research group at the University of Waterloo. To be able to use these models effectively the parameters that they contain have to be known.

The case studies presented in the bulk of the literature dealing with parameter estimation have only considered small models. A small model has less parameters, less equations and can frequently be solved analytically. An example of a 'small' model is the A to B to C model that represents the sequence of first order irreversible chemical reactions of A producing B and B producing C (Bates and Watts, 1998). The use of small models in the parameter estimation literature in the past is due to the limited computing power that was available. Therefore to make the problems tractable small models were used which often had analytical solutions available.

The problem of nonlinear parameter estimation has been addressed by a number of authors, including the early work by Box and Draper (1965) in the development of the

determinant criterion, to more recently where there are a number of texts that discuss the subject such as Seber and Wild (1989), Bates and Watts (1988) and Bard (1974). These texts provide a general discussion of the theory associated with nonlinear parameter estimation and show how it has evolved from the linear case. Various issues in the different steps of parameter estimation have also been addressed by Duever and Penlidis (1998), Bartus (1987), Klaus (1981), Box et al. (1973) and Hemker and Kok (1993) among others. While a number of authors discuss the individual steps of parameter estimation in detail, there is very little discussion of the how the steps are related.

In estimating parameters for large process models, at present there are two shortcomings in the existing knowledge about parameter estimation. The first is, how effective is the present parameter estimation methodology when applied to large models, and the second is, can any advantage be gained from considering the parameter estimation problem as a whole. Therefore to address this limitation, the objective of this work was to revisit the parameter estimation process and to develop a protocol for the estimation or updating of the parameters within process models. The protocol was then tested by using it in several case studies to estimate parameters in the Watpoly polymerization modelling package.

This research into the development of a protocol for estimation of parameters within large models is part of the larger research program being performed in the polymer research group at the University of Waterloo. The objective of this work is to develop a set of tools to support a model-based experimentation approach. The model in this case is not only a store of information but an integral part of the research process of proposing various hypothesis, which are then experimentally verified.

In the development of a parameter estimation protocol all of the steps of the parameter estimation will be revisited. The parameter estimation steps are: parameter sensitivity analysis, statistical design of experiments, estimation of parameters and confidence regions. Each of these steps correspond to answering one of the following questions:

1. Is it possible to estimate the parameters with the chosen responses and which response or responses will provide the most information?

2. At what conditions (i.e. temperature, conversion, initial feed composition) should the data in the experiment be collected?

3. What is the best method to estimate the parameters with the data that was collected?

4. How good are the parameters that were estimated?

The projected use of the parameter estimation protocol is as part of a model based experimental program. In this work simulated experiments will mimic actual experiments. In practice the number of experiments that can be carried out is limited due to the expense of performing experiments and analysis. In the case studies 'actual' experiments with a limited number of trials will be simulated by using the model to simulate the responses and then adding reasonable experimental errors. In this work it is assumed that the model is correct and the error that is added to the simulation responses represents only the variation associated with reproducibility (i.e. experimental error). If it is suspected that the model is not correct, then trials for testing lack of fit should be included in the experiment design. The focus of this work was on the parameter estimation process and hence model lack of fit was not considered.

The parameter estimation protocol is a series of steps that should be followed in the course of obtaining parameter estimates. By following these steps the overall parameter estimation procedure can be made more efficient and some pitfalls such as local optima, and inaccurate confidence regions can be dealt with in an appropriate manner. Each of the four parameter estimation steps will be discussed in a separate chapter, which will be followed by three case studies to illustrate the proposed parameter estimation protocol.

The first step in the parameter estimation process is a sensitivity analysis and will be discussed in Chapter 2. Sensitivity analysis will identify which of the responses will

provide the most information about the chosen parameters. This is achieved by perturbing the parameters and examining the change in the responses. This is equivalent to calculating the gradient of the response with respect to the parameters as shown in equation 1.1.

$$Gradient = \frac{\Delta f(x, \theta_i)}{\Delta \theta_i} \tag{1.1}$$

Sensitivity analysis can be broadly divided into two areas, global and local. Local sensitivity is equal to the gradient at a point, while global sensitivity will obtain an average measure of sensitivity over a space of points. Of the available methods for sensitivity analysis, plotting of the gradient values was found to be the best approach. This approach provides a significant amount of information while still being simple to interpret.

The second step of parameter estimation is the design of experiments and will be discussed in Chapter 3. The objective in the design of experiments is to create an experiment that will maximize the information about the parameters from the data collected. The criterion used in this work is the D-optimality criterion, developed by Box and Lucas (1959) and extended to the multiresponse case by Draper and Hunter (1966) and Atkinson and Hunter (1968). The D-optimality criterion is shown in equation 1.2.

$$\det[(V'V)^{-1}] \tag{1.2}$$

Where V is the matrix of gradient values of the responses with respect to the parameters evaluated at the experiment trial points. The application of this criterion to large models is a difficult task, as the resultant optimization problem will in general have a large number of local optima. One method to overcome this problem, is to use the information obtained from the sensitivity analysis. How this information is obtained and used will be discussed.

The third step in the parameter estimation process is the estimation of the parameter values from the data collected and will be discussed in Chapter 4. The principal criterion used in this thesis is the determinant criterion developed by Box and Draper (1965) and

shown in equation 1.3, where Z is the matrix of deviations.

$$min|Z'Z| \tag{1.3}$$

The optimization problem in parameter estimation was found to be very challenging due to the large number of local optima that were found and the computational expense of working with large models. Unfortunately no simple, generic solution was found to overcome these problems. However a set of recommendations is given that will allow the researcher to deal with the problem of local optima.

The last step in parameter estimation is the calculation of joint confidence regions of the parameter estimates and is discussed in Chapter 5. Confidence regions will provide an indication of the uncertainty that is present in the estimated parameters and thus are a reflection of the quality of the parameter estimates. They also allow for a comparison of parameter estimates from different experiments. In the generation and interpretation of these confidence regions a number of guidelines are presented as some limitations exist in their use that are not readily apparent. An example is the limitation of the current theory for the generation of confidence regions when the sample size is small and multiple responses are used. The confidence regions obtained under these conditions can be much smaller than the true confidence regions which can be obtained from a Monte Carlo analysis.

The proposed protocol for parameter estimation is illustrated in Chapter 6 with three case studies involving the estimation of parameters in the Watpoly model, which is a polymerization simulation model based on a mechanistic consideration of the polymerization process. These cases will illustrate some of the problems that may be encountered in the parameter estimation process and how the proposed protocol can aid in overcoming them. The conclusions and recommendations for future work will then be presented in Chapter 7.

Throughout this thesis the terms 'true parameter value', 'initial parameter value', 'parameter value', 'input variable' and 'response variable' will be used. To clarify their

use within this thesis, they are defined in the following manner. The term 'parameter' will refer to a value within a model that represents a physical or chemical constant, such as reaction rate or rate of diffusion. Since in this thesis the experiments will be simulated, the 'true parameter values' will refer to the parameter values used to generate the response values. The 'initial parameter value' will refer to the value used as the starting point within an optimization or the values used to design experiments. This value will be different from the true value as defined above. An 'input variable' is a variable manipulated by the researcher, such as the initial feed composition or the temperature. A 'response variable' is a variable that would normally be measured by a researcher such as the final composition, the molecular weight of a polymer, or the reaction rate.

Chapter 2

Sensitivity Analysis

2.1 Introduction

The goal of sensitivity analysis is to determine how a response will be affected by a parameter perturbation. In the parameter estimation process this information is used to answer the following related questions.

1. Can we estimate a parameter, θ_i, using the response y_j?

2. Which response will provide the most information about a chosen parameter?

One of the major difficulties in conducting a sensitivity analysis with nonlinear models, is that values of the unknown parameters are required. This is a common limitation of working with nonlinear models, and numerous methods have been developed to minimize this effect of unknown parameters.

A number of different methods have been developed to perform a numerical sensitivity analysis and are discussed in section 2.2. The approach of these methods can be generally divided into global or local sensitivity. A more detailed comparison and discussion of global and local sensitivity is given in section 2.3. Of the available methods for sensitivity analysis, in this work gradient plots were found to be the best approach and will be

discussed in section 2.4. This section will discuss the implementation and interpretation of the gradient plots and an example will be presented to illustrate the method.

2.2 Sensitivity Analysis Review

A number of different methods exist to perform a sensitivity analysis. A review of sensitivity analysis in general is given by Vajda et al. (1989), Markussen and DiStephano (1982), Rabitz et al. (1983), Sulieman (1998) and Seber and Wild (1989). The objective of all of these methods is to identify which of the responses within a model will provide the most information about a given set of parameters. This is achieved by perturbing one of the parameters at a given set of conditions (i.e. values of the inputs and other parameters) and examining the corresponding effect in the response. This effect can be quantified by calculating the numerical gradient of the model response with respect to the parameter of interest, θ_i, at the chosen parameter values and process conditions, as shown in equation 2.1. Where $f(x, \theta_i)$ is the model response and x is the set of input variables.

$$Gradient = \frac{\Delta f(x, \theta_i)}{\Delta \theta_i} \tag{2.1}$$

The simplest method of sensitivity analysis involves an examination of the gradient values. This is usually achieved by plotting the gradients at the chosen parameter values and process conditions versus one of the independent variables such as initial composition or time for a dynamic response. As the model being used becomes more complex and the number of parameters and responses to be considered increases, the amount of information that needs to be analysed increases considerably. To overcome this, various methods to summarize the sensitivity information have been proposed.

To compare the sensitivity of two responses to a given parameter using gradient values, the response magnitude and measurement error have to be taken into account. If

this is not done the sensitivity analysis will be biased. An example of this bias would be the case where two responses have the same gradient values, but the measurement error of one of the responses is much larger than the other. If the measurement error is not considered the responses would be considered equal based on their gradient values. But the response with the smaller measurement error will provide more information about the parameter. To remove this bias, the gradient value is divided by the standard deviation of the response measurement error, σ_{resp}, as shown in equation (2.2). The normalized gradient is discussed further in section 2.4.2.

$$Normalized\ Gradient = \frac{\Delta f(x, \theta_i)}{\Delta \theta_i\, \sigma_{resp}} \tag{2.2}$$

In performing a sensitivity analysis a further point to consider is the uniqueness of the response. By this we mean if both parameter θ_1 and θ_2 affect response y_1 in the same manner, then it will not be possible to determine which parameter produced the change in y_1. Thus the two parameters θ_1 and θ_2 will not be observable (i.e. estimable) together and if considered individually (i.e. the other is fixed), their estimates will be highly correlated.

Global sensitivity methods address two difficulties in sensitivity analysis. The first is that the parameters are unknown but are required for the analysis if a nonlinear model is used. The second is the difficulty of interpreting the large volume of information that may exist if point gradient values are used. Global sensitivity methods obtain an 'average' sensitivity value over the chosen range of parameter values and process conditions. By averaging the sensitivity values over a range of parameter values and a range of input variables, the uncertainty due to the unknown parameter values is decreased. A further advantage is that the sensitivity analysis is summarized in a few values and therefore is easier to interpret.

The simplest form of global sensitivity is obtained by calculating a numerical average of the gradient values over a defined parameter and input variable space. The implementation of this global sensitivity measure involves the following steps.

1. Define a range of certainty for each of the parameters (i.e. the likely minimum and maximum parameter values).
2. Define a range for each of the input variables.
3. Discretize each of the parameter and input variable ranges into n_i steps, where $i = 1...k$.

The above steps will produce a discretized parameter and input variable space with N points, where $N = \prod_{i=1}^{k} n_i$. The global sensitivity value of this space is obtained by taking the average of the absolute gradient values at each point in the grid, as shown in equation 2.3.

$$Global\ \ Sensitivity = \frac{\sum_{j=1}^{N} \left| \frac{\Delta f_j(x,\theta)}{\Delta \theta} \right|}{N} \tag{2.3}$$

N is the total number of points in the grid, j is the grid index, $\Delta f_j(x,\theta)$ is the change in the response at the j-th point in the grid that covers the desired parameter and process condition ranges, and $\Delta\theta$ is the parameter perturbation. Although simple to calculate this approach will not reveal if any parameter interaction exist. To overcome this Pierce et al. (1981) propose a more sophisticated approach based on modulating the parameters at given frequencies and then performing a Fourier analysis of the responses. A comparison of four global sensitivity methods was performed by Markusen and Di Stefano (1982), and they determined that integration of the gradient curves is the best approach for large models.

A related technique to global sensitivity is the global structural identifiability. The objective of these methods is to determine if a parameter can be estimated (i.e. is observable) with a given set of responses assuming no measurement error is present. These methods have been used primarily with compartmental models in the pharmacokinetic field. The various methods are discussed by Chappell et al. (1990), Chappell and Godfrey (1992) and Seber and Wild (1989) among others. Vajda et al. (1989) compare the linearization, power series expansion and similarity transform approaches to sensitivity

analysis with an emphasis on chemical kinetic modelling. Ljung and Glad (1994) propose a method based on differential algebra that shows how global structural identifiability is determined if the given model structure can be rearranged as a linear regression. A limitation of all of these methods is that they only determine if a parameter is observable with a given response. They do not consider the quality of this observability as the response measurement error is not considered in the analysis. With this limitation and the requirement of algebraic manipulation, this approach was not considered a feasible option to use with the type of models considered in this thesis.

A further application of sensitivity analysis is as a screening method to determine which parameters are the most significant. This is usually achieved by performing a fractional factorial type of experiment on the model where the experiment factors are the model parameter values and the experiment responses are the changes in the model responses. This type of sensitivity analysis using different types of factorial experiments is discussed by Andres (1997) (fractional factorial designs), DeWit (1997) (Plakett-Burman designs) and Rahni et al. (1997) (factorial designs). This approach to sensitivity analysis is only applicable to the initial stages of a model investigation where it is not known which parameters are the most important, and is not applicable as part of a parameter estimation process as considered in this thesis.

The numerical aspect of calculating gradients in a sensitivity analysis for a dynamic model has been addressed by a number of authors such as Leis and Kramer (1985), Leis and Kramer (1988), Caracotsios and Stewart (1985) and Guay and McLean (1995) among others. Guay and McLean present a method of obtaining the gradient values while solving the system of differential equations and a method to obtain the second order sensitivity values. In calculating the gradient, the change in the response (i.e. Δy) is obtained by running the simulation twice and taking the difference between the two sets of results. If the gradients can be obtained in only one simulation pass, by extracting them from the ODE solver while the differential equations are being solved,

then the overall computation time required to obtain the gradient plots is decreased by half.

2.3 Global and Local Sensitivity

Sensitivity can be broadly subdivided into two areas, local and global sensitivity. Local sensitivity is the gradient value of a response with respect to a parameter at a given set of conditions (i.e. input conditions and other parameter values), while global sensitivity is an 'average' sensitivity value over a given range of conditions. Which of these two approaches is best as part of the parameter estimation protocol will be addressed in this section, where the advantages and disadvantages of each approach will be discussed.

Global sensitivity has two advantages. The first is its ease of interpretation, as it summarizes all of the sensitivity information for the chosen response into one value. The second advantage is that by covering a range of parameter values it will reduce the amount of uncertainty in the sensitivity analysis due to the true parameter values being unknown. The disadvantage of a global sensitivity measure is that it will not provide any information about the location of the sensitivity. This information about the location and distribution of the sensitivity was found to be very useful in the design of experiments as described in Chapter 3.

Local sensitivity has the advantages of providing both the location of the sensitivity information and a greater understanding of the parameter/response relationships in the model. Its disadvantages are twofold: The first is that the sensitivity values are dependent on the parameter, and input variable values. The second is that the amount of data that has to be analysed increases significantly as the number of parameters and/or responses considered increases. To deal with these limitations, an effective method is to plot the gradients versus the input variable at different parameter values. This allows the researcher to quickly analyse the gradient data with a visual inspection of each plot. More detail on the implementation of this approach and how it is part of the parameter

estimation protocol is given in section 2.4.

To illustrate the difference between local and global sensitivity and how information can be lost by using an average (i.e. global) sensitivity value an example will be presented. In this example the global and local sensitivities will be calculated for two responses with respect to the parameter r_1 in the batch copolymerization of Styrene / Methyl Methacrylate at a run time of thirty minutes. The sampling time of thirty minutes was chosen arbitrarily here, as only the effect of initial feed composition is being considered in this example and not sampling time. The responses used are copolymer composition (F1) and the cumulative triad fraction of monomer1–monomer1–monomer1 (A111). The input variable which is changed is the initial mole fraction of Styrene in the feed.

The local sensitivity analysis involves calculating the gradients of the two responses with respect to the parameter r_1 at different values of the independent variable. This information is then summarized by plotting the gradient values versus the independent variable. These plots are shown in Figure 2.1, where the horizontal axis is the initial mole fraction of Styrene and the vertical axis is the normalized gradient value as given by equation 2.2. A visual inspection of the gradient curve of the triad fraction and copolymer composition, indicates that the triad fraction will provide more information about the parameter due to its larger gradient values in general. A further piece of information that can be obtained, is at what initial feed composition will the information be maximized. This occurs where the absolute gradient values are the largest, which is when the initial feed is near 0.8 mole fraction Styrene for both curves.

The global sensitivity value is obtained by averaging the gradient values over the range of the independent variable (0.1 mole fraction Styrene to 0.9 mole fraction Styrene). This results in global sensitivity values of 8.9 for the triad fraction and of 2.8 for the copolymer composition. From these values we can conclude that the triad fraction A111 contains more information about the parameter r_1 than copolymer composition, due to its larger value, but not at what initial feed to sample.

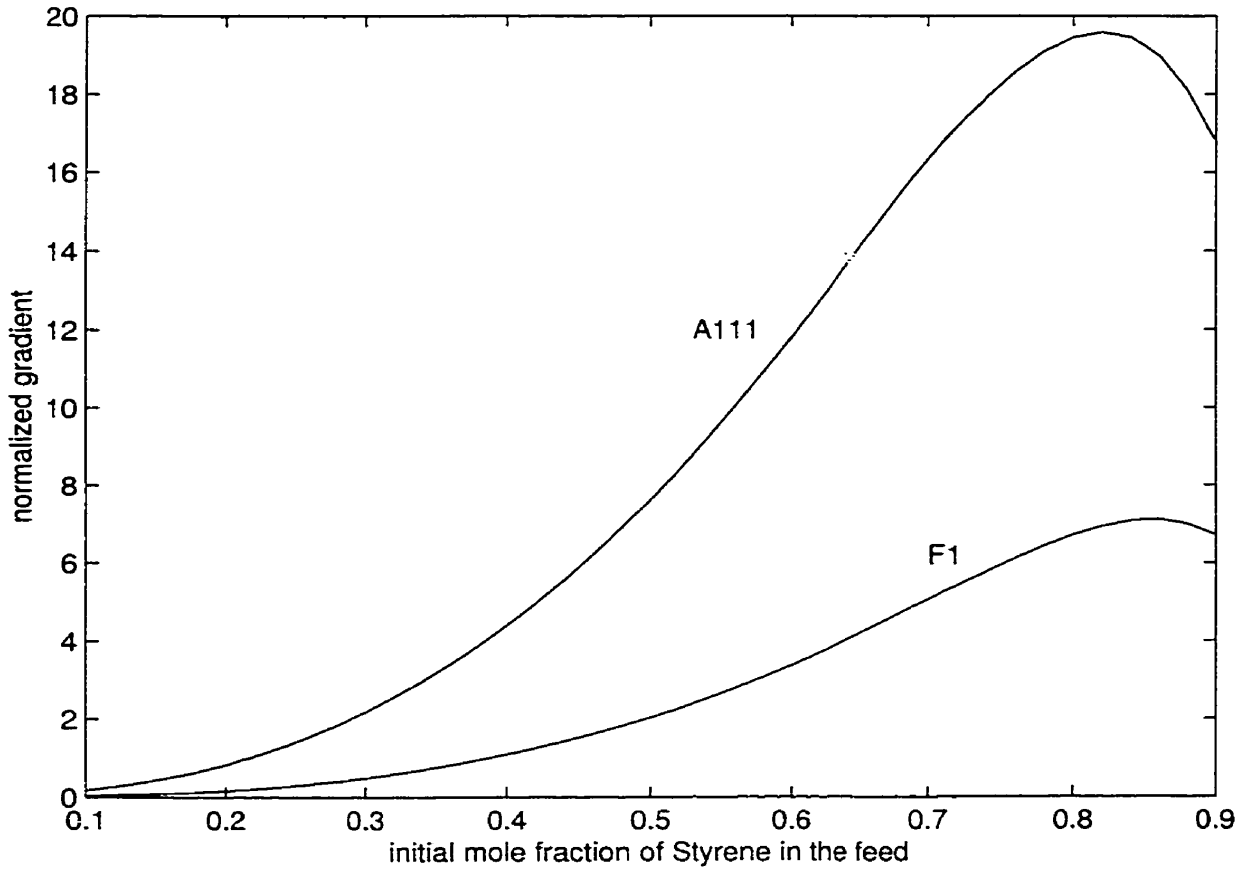


Figure 2.1: Gradient curves with respect to the parameter r_1 in the copolymer system Styrene/Methyl Methacrylate at 30 minutes.

In general, as shown in the above example both local and global sensitivity will indicate which of the responses will provide the most information. Only local sensitivity will indicate where this information is located. This is a significant asset, as this information was found to be very useful in the design of experiments.

2.4 Gradient Plots

2.4.1 Overview

This section will discuss how plots of the gradient data can be used to perform a parameter sensitivity analysis and in what other areas of the parameter estimation process can this information be used to make the overall process more efficient. This method was found to be an adequate compromise between the amount of information provided and ease of analysis. A description of the method and how to interpret the information provided will be given and highlighted with an example.

As discussed in the previous section, using local sensitivity information is a better approach as it provides insight into the location and distribution of the information about the parameter values. However, the use of local sensitivity is accompanied with the challenge of how to effectively present a large amount of information (i.e. the large number of gradient values) in a manner that is easy to interpret while still containing the desired information.

The most effective way to present the sensitivity information is to plot the gradient values with respect to input variables (i.e. process conditions) such as sampling time or initial feed composition. This approach produces a large number of graphs to interpret, but the interpretation can be accomplished quickly and easily. When plotting the gradient values versus one input variable, a 2D graph is produced and if two input variables are used a 3D graph is produced. Examples of these plots are shown and discussed in section 2.4.3. The difference between the 2D and 3D plots is in the type of information that is

provided. The 3D plot will show the interaction effect that two input variables have on the gradient, while the 2D plot is better at comparing different responses to determine which will provide more information about a parameter. If more than 2 input variables are to be considered then multiple 3D plots can be generated. In analyzing the gradient plots the goal is to identify regions where the gradients have large values.

Section 2.4.2 will discuss issues in the generation of the gradient plots. How to interpret the plots is discussed in section 2.4.3.

2.4.2 Implementation

In plotting the gradient values for analysis a number of issues arise, such as; how to compare different responses; how to compare the observability of two different parameters; and how to determine the quality of a response. The methods to address these issues will be described next.

To compare multiple responses to determine which response will provide the most information about a parameter, a direct comparison of the gradient values is not usually feasible. This is due to the different measurement errors that may exist and/or the different scales of the responses. To account for this, the gradient values are normalized by dividing them by the standard deviation of the measurement error, σ_{resp}, as originally shown in equation 2.2 and repeated in equation 2.4 below for the readers convenience. The value of σ_{resp} is obtained based on prior experience with a given type of analysis and is assumed to be constant (i.e. homoscedastic). If σ_{resp} is not constant then its variance should be accounted for by using different values of σ_{resp} as required.

$$Normalized\ Gradient = \frac{\Delta f(x, \theta_i)}{\Delta \theta_i\ \sigma_{resp}} \tag{2.4}$$

It is not possible to compare the observability of two or more parameters using the normalized gradients given by equation 2.4 if the parameters are of different orders of magnitude. For example, in the system Styrene / Methyl Methacrylate, it might be of interest

to determine which of the two parameters, the reactivity ratio ($r_1 \approx 0.5$) or the activation energy in the Arrhenius expression for the rate of propagation ($k_{p\,act-eng} \approx 7700$), is more observable using the response of copolymer composition. A direct comparison of the normalized gradient values is not adequate. This is because the normalized gradient value as given by equation 2.4 is the number of measurement error standard deviations a response will change for a unit change in the parameter. This unit change in the parameter will distort the normalized gradient values due to the different scales of the parameters. To deal with this, the normalized gradient values are multiplied by a percentage of the parameter value, as shown in equation 2.5. Ten percent is used in the equation below, this value was chosen arbitrarily and any small value (i.e. less than 30%) can be used.

$$Normalized\ Gradient = \frac{\Delta f(x, \theta_i)}{\Delta\theta_i\, \sigma_{resp}} (0.1\theta_i) \tag{2.5}$$

When the gradients are plotted versus two input variables it was found to be useful to also generate a contour plot of the gradient surface. This, as well as the ability to rotate the gradient surface plot to view it at different angles, makes the plot easier to interpret. An example of a normalized gradient plot and accompanying contour plot for two responses are shown in Figures 2.2 and 2.3 in the next section.

2.4.3 Interpretation

To best describe how to interpret information provided by the gradient plots an example will be used. The following example uses the Watpoly model to simulate the copolymerization of Styrene / Methyl Methacrylate at 65 °C , with the initiator AIBM-VAZO-64 at a 0.25 weight percent loading and the simulation options listed in Appendix A. The parameter of interest is the reactivity ratio r_1 and the responses that are considered are conversion and copolymer composition.

Figures 2.2 and 2.3 show the gradient plots and gradient contours for parameter r_1 with respect to time and initial feed composition for the responses of conversion and copolymer composition, respectively. It was found that it is useful to generate the contour

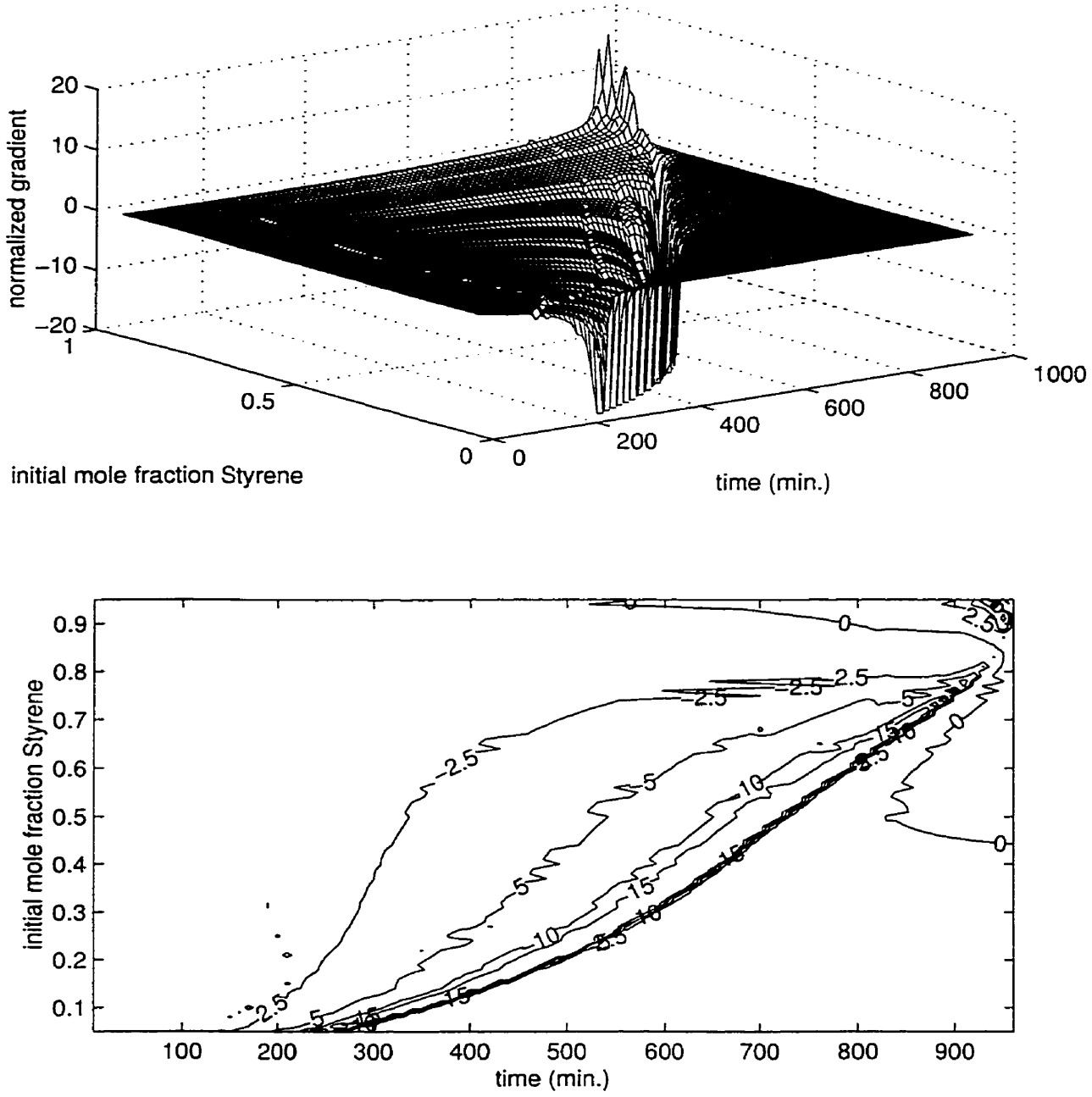


Figure 2.2: Gradient plot of the conversion response with respect to the parameter r_1 versus sampling time and initial feed composition for the system Styrene / Methyl Methacrylate.

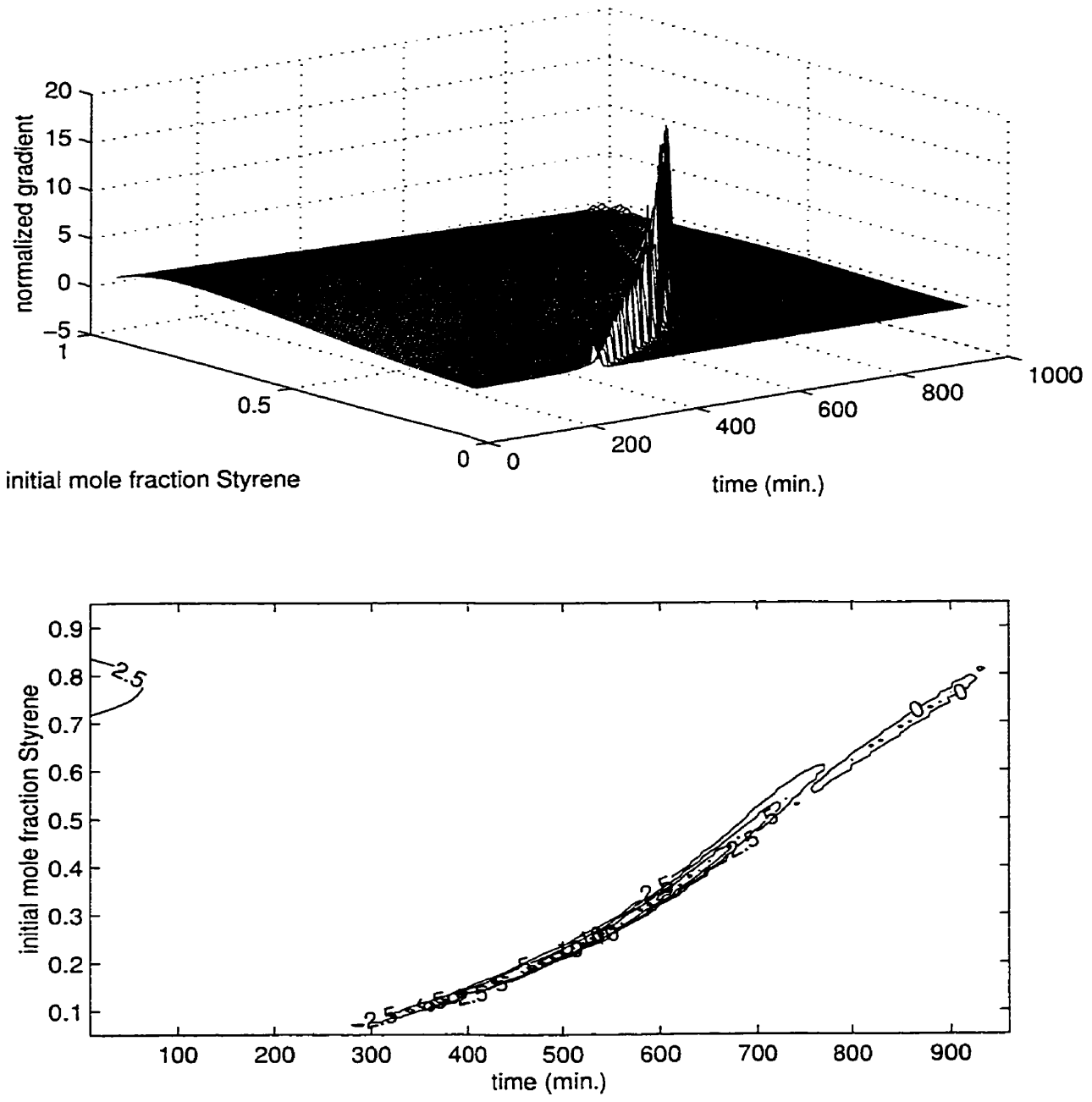


Figure 2.3: Gradient plot of the composition response with respect to the parameter r_1 versus sampling time and initial feed composition for the system Styrene / Methyl Methacrylate.

plot of the gradient surface, as this aids in the interpretation of the data. The raggedness in the surface plot and the contour plot is a numerical artifact of the resolution used in their generation. In both figures the vertical axis of the 3D plot is the normalized gradient value, while the horizontal axes are the reaction time and the initial feed composition. The gradient value was normalized by dividing it by the response measurement error standard deviation and multiplying it by a percentage of the parameter value, as discussed in the previous section and shown in equation 2.5. To determine where the experiment trials should be placed, large values of the gradients, either positive or negative, are desired. As this indicates a region where the response is very sensitive to the chosen parameter. An examination of the 3D gradient plot for the response of conversion indicates a valley (i.e. an area of large negative gradient values) where the information is maximized. This valley begins at 300 minutes and a low Styrene feed composition and continues to 900 minutes and a high styrene feed composition. An analysis of the gradient plot of the copolymer composition response (Figure 2.3) yields a similar observation, except that the area of high observability is a ridge instead of a valley (i.e. large positive values of the gradient) and the ridge is also much narrower.

Determining the location of the large gradient values will indicate the best areas to sample. By examining the size of the normalized gradient values it is possible to determine the amount of information that is present in the response. Therefore it is possible to obtain a qualitative indication of the number of trials (either replicated or unique) that will be required to estimate the parameters. If the largest absolute normalized gradient values are small (i.e. < 1), this is an indication that a large number of replicates may be required due to the low level of information provided by the responses

The argument could be presented that while the gradient plots will provide a lot of information about the individual parameter/response relationships, very little will be learned about the correlation between parameters (i.e. how a change in the value of one parameter will affect the observability of another parameter). To counter this argument,

a sensitivity analysis was found to be best used as an exploratory tool that will guide the researcher in the correct direction and is not meant to answer all of the questions about a parameter's estimability. This is an acceptable approach, since parameter correlation will be taken into account by the experiment design criterion. The sensitivity analysis of the designed experiment will then tell the researcher how sensitive the design is to the parameter values. More details on the experiment design and a sensitivity analysis of it is described in Chapter 3.

Chapter 3

Design of Experiments

3.1 Introduction

This chapter will discuss experiment design for large nonlinear dynamic models. The models used are assumed to be correct and the objective of the experiment is defined as 'the estimation of a set of parameters'. This objective may exist due to the desire to improve a set of parameter estimates (i.e. decrease the size of the confidence region of the current parameter estimates) to decrease the confidence region of the model prediction. Another reason may be to expand the model to a similar system with different components, a new monomer within a polymerization model for example.

A number of different criteria have been used to design experiments for parameter estimation in nonlinear models. A listing of the most common design criteria is given in Table 3.1. A further discussion of these criteria can be found in a number of sources, such as Atkinson and Donev (1992), Shaw (1994), Chaudhuri and Mykland (1993), Atkinson (1996), Draper and Pukelsheim (1996) and Ford et al. (1989). Atkinson and Donev present a more theoretical discussion of the criteria while Shaw has a more applied focus and provides a comparison of the criteria as applied to a number of nonlinear test cases.

Of the design criteria available, the criterion chosen is D-optimality. This criterion

Criterion	Reference	Criterion to be minimized
D-optimal	Wald (1943), Box and Lucas (1959)	$\det[(V'V)^{-1}]^{\dagger}$
A-optimal	Atkinson and Donev (1992)	$\text{trace}[(V'V)^{-1}]^{\dagger}$
E-optimal	Atkinson and Donev (1992)	max. eigenvalue of $[(V'V)^{-1}]^{\dagger}$
G-optimal	Atkinson and Donev (1992)	max. variance of the predicted response over the range of interest
Quadratic D-optimal	Hamilton and Watts (1985), O'Brien (1992)	Quadratic approximation to the joint confidence region
X-optimal	Villa (1990)	true shape joint confidence region

† V is the matrix of gradient values of the response with respect to the parameters evaluated at the experiment trial points

Table 3.1: Criteria used in the design of parameter estimation experiments for nonlinear models

is used since it will minimize the hypervolume of the linearized joint confidence region using multiple responses, and has been used previously by other researchers in the design of experiments for the estimation of parameters in nonlinear models.

Background on the D-optimality criterion is discussed in Section 3.2 and is followed by a general discussion of the implementation of the criterion in Section 3.3. Section 3.4 describes a diagnostic procedure to determine how sensitive the designed experiment is to the initial parameter estimates.

3.2 D-optimality: Background

The D-optimality criterion for the design of experiments was initially proposed by Wald (1943). The motivation behind the development of this criterion was to minimize the volume of the elliptical joint confidence region of the parameter estimates. For a linear

model , given by:

$$Y = X\beta \tag{3.1}$$

β is a vector of parameters, X is the matrix of regressor variables and Y is a vector of predicted values. The volume of the joint confidence region of the parameter estimates is proportional to $|X'X|^{-1/2}$. Therefore Wald proposed maximizing the determinant of $(X'X)$, as this would result in parameters with the smallest confidence regions and the smallest amount of uncertainty (note: maximizing $|X'X|$ is equivalent to minimizing $|X'X|^{-1/2}$). Box and Lucas (1959) extended this to nonlinear models by replacing the X matrix by the derivative matrix V^o. The D-optimality criterion applied to nonlinear models is given by:

$$max(|V^{o\prime}V^o|) \tag{3.2}$$

V^o is the gradient or Jacobian evaluated at the initial parameter values (θ^o) given by:

$$V^o = \left.\frac{\partial f(x,\theta)}{\partial \theta}\right|_{\theta^o} \tag{3.3}$$

In the D-optimality criterion extension derived by Box and Lucas, they consider the case where the number of trials in the experiment is equal to the number of parameters (i.e. $n = p$). Draper and Hunter (1966) extended this to the case where N runs have already been performed and it is desired to design n additional runs. This extension was developed based on a Bayesian approach where the previous N runs provide prior knowledge for the future n runs.

The criterion as given by equation 3.2 is for the single response case. Box and Draper (1965), and Draper and Hunter (1966) extended the D-optimality criterion to the multiresponse case as shown in equation 3.4.

$$\max \quad \left|\sum_{i=1}^{r}\sum_{j=1}^{r}\sigma^{ij}V_i^{o\prime}V_j^o\right| \tag{3.4}$$

Where V_i^o is the gradient matrix of response i (equation 3.3) evaluated at θ^o and σ^{ij} is the ij-th element of the inverse of the covariance matrix of the r responses.

A property of D-optimality is that D-optimal designs are invariant to any non-degenerative transformation applied to the model parameters, such as scaling. A further property of D-optimality is that a D-optimal design is equivalent to a G-optimal design (i.e. a design that minimizes the maximum variance of the predicted response) as shown in Kiefer and Wolfowitz (1960), and Kiefer (1974). This equivalence theorem has only been proven for linear models where the design points span a continuous space.

In using the D-optimality criterion to design experiments the following assumptions are made:

1. The expectation surface is close to linear in the neighborhood of the initial parameter values, therefore the linear approximation to the joint confidence region is adequate

2. The initial parameter values are close to the true parameter values, therefore the designed experiment will also be a good experiment for estimating the true parameter values.

3. The measurement errors are homoscedastic.

How the above assumptions affect the implementation of the D-optimality criterion is discussed in section 3.3.1

The standard D-optimality criterion uses the gradient matrix V^o, which is used in the first order (i.e. linear) approximation of the parameter estimate joint confidence region. Conditions exist where this approximation may not be adequate for nonlinear models. To address this limitation, Hamilton and Watts (1985) extended the joint confidence region approximation by including a second order term in the gradient expression. This extension requires the residuals of the future measurements and the measurement error. Since these values are not known at the design stage, Hamilton and Watts assume a value for the measurement error and set the residuals to zero. In making these assumptions the D-optimal quadratic criterion becomes less sensitive to changes in the initial parameter

values, but is affected by the method of parameterization and is very sensitive to the magnitude of the measurement error (Hamilton and Watts, 1985). Whether it is a better implementation of the D-optimality criterion is dependent on the value of the measurement error. If the error is too large very poor designs will result and there appears to be no single way of determining if a given value of the measurement error is too large for the D-optimal quadratic criterion to be used safely (Seber and Wild, 1989; Hamilton and Watts, 1985). O'Brien (1992) further extended the quadratic D-optimality criterion as proposed by Hamilton and Watts to include $n+1$ points in the design. This modified design has the advantage of being able to test for model fit which is not possible with the original criterion of Hamilton and Watts, where the number of support points (i.e. unique trials in the experiment design) equals the number of parameters to be estimated.

3.3 Designing D-optimal Experiments

This section will discuss the implementation of the D-optimality criterion to design experiments. A number of difficulties can be encountered while designing experiments for nonlinear models. These may be due to one or more of the following reasons.

1. The assumption that the parameter values used to derive the experiment design are close to the true values and that the difference between the parameter values used and the true values will not adversely affect the quality of the experiment.

2. The assumption that the linear approximation of the joint confidence region used in the experiment design criterion is an adequate representation of the true joint confidence region

3. The presence of locally optimal experiments.

How the above difficulties affect the design of experiments will be discussed in the sections that follow. While no method will guarantee that the globally optimal experiment is found, the proposed methodology will provide a systematic framework that is an improvement over a less formal empirical approach.

The numerical aspect of the design of experiments will not be addressed directly, as the focus of this work is the general procedure and the statistical assumptions that are made in designing experiments. A discussion of the numerical aspect of the design of experiments is given by Atkinson and Donev (1992), though they focus on linear models. In the design of experiments optimization plays a significant role due to the difficulty of local optima. This difficulty is not a particular characteristic of the design of experiments, but exists in a number of other fields. Therefore the performance of the optimization algorithm was not addressed, other than to choose a robust algorithm. A general overview of optimization methods can be found in a number of texts such as Fletcher (1987) and Press et al. (1989).

The next three subsections will discuss the effect that the assumptions have on the D-optimality design calculation, local optima in designing experiments and other implementation issues.

3.3.1 Assumptions Made in Using D-optimality

In implementing the standard D-optimality criterion, three assumptions are made. If these assumptions are not valid the quality of the designed experiment can be detrimentally affected. The first assumption is that the volume of the joint confidence region, considered to be proportional to $|V^{o\prime}V^{o}|^{-1/2}$, is an adequate approximation of the volume of the true joint confidence region. This assumption is based on the model being linear in the neighborhood of the parameter values. The second assumption is that the parameter values used to design the experiment are close to the true parameter values. This assumption is made because the experiment design is a function of the parameter values

and may change with a change in the parameter values. The third assumption is that the measurement errors are homoscedastic. To deal with the situation that one or more of the assumptions made are not valid, various methods are available and will be discussed below.

The third assumption is the easiest to deal with. If the measurement error is not homoscedastic, the model can be transformed so the error becomes homoscedastic or a diagonal matrix of weights can be used in the D-optimality criterion as described by Seber and Wild (1989, pg. 251).

If the first assumption, that the linearized joint confidence region is adequate, is not valid two different approaches have been proposed. The first is to extend the gradient equations to include second order terms, (Hamilton and Watts, 1985). The benefits of this modification is offset by its other limitations as discussed in Section 3.2. The second method is to use the true shape joint confidence region (i.e. X-optimality, Villa, 1990). This would represent the ideal case as it would consider all of the nonlinearity present in the model. The current limitation of this criterion is that with the current computing power available, it is not feasible to implement if the number of parameters is large (i.e. greater than 3 or 4) or if the model is computationally expensive to evaluate. If this assumption is not satisfied, then the designed experiment will be sub-optimal.

Failure of the second assumption, that the parameter values used are close to the true parameter values, can be dealt with by a number of different approaches. All of these will either minimize the effect of the unknown parameters or reveal it, so that appropriate action may be taken. The first approach will minimize the effect of the unknown parameter values by the use of a robust design criterion. This family of criteria take a conservative approach and strive to generate an experiment that, although suboptimal at the initial set of parameter estimates, will be good over a range of parameter values. An example is the minimax criterion proposed by Pronzato and Walter (1988). This criterion will try to obtain the best 'worst case', within a specified parameter range. To

achieve this the following two step optimization is used.

1. Given an experiment, the parameter values are varied within a pre-specified range until the D-optimal criterion is minimized (i.e. the worst case).

2. Using the worst case parameter values from step 1, the sampling conditions are varied until the D-optimal criterion is maximized (i.e. the best case).

The experiment from step two is then used in step one and the above two steps are repeated until convergence is achieved (i.e. the parameter values and the D-optimal experiment do not change at the chosen tolerance between iterations). A limitation of this method is choosing appropriate parameter ranges. Since the design is highly dependent on the ranges of the parameter values, a very poor experiment may result if the parameter ranges are too large. Alternatively, if the parameter ranges are too small, an experiment will be obtained which is not sufficiently robust.

The second approach to dealing with unknown parameter values and the simplest to use, is a sequential experiment design. This is analogous to the iterative approach of model development. This method can be applied to the standard D-optimality criterion in two ways.

1. An independent D-optimal experiment is designed at each step of the investigation, where the current estimates of the parameter values are used in the design of the next set of trials.

2. A Bayesian approach is taken, where there is formal use of the prior knowledge about each of the parameter estimates and their distribution in the design of the next experiment.

A third approach is to perform a sensitivity analysis of the experiment design with respect to the parameter values. This will provide the experimenter with information about the change in quality of the design with changes in the parameter values. The

procedure to perform the designed experiment sensitivity analysis is described in Section 3.4.

While no simple solution was found to deal with the problem of unknown parameter values, the sensitivity analysis will indicate how sensitive the given design is to the initial parameter values. Therefore the experimenter will have an idea of the risk associated with the given D-optimal design and can choose to modify the design to reduce this risk. How to best modify the experiment so that it is less sensitive to the initial parameter values is case dependent. As a guide to generating a more robust experiment, a list of possible approaches is given below.

- Design multiple experiments using different parameter values and make the final design a compilation of all of the designs. This can be achieved from a visual inspection of the different experiment designs to identify general regions where the trials are placed. Then the trials of the amalgamated design are placed in these regions.
- Use one of the robust criteria, such as the minimax criterion.
- In the experiment design process a number of locally optimal experiments are usually found. If the best experiment found is too sensitive to the parameter values used to design it (i.e. the quality of the experiment becomes very poor when evaluated using parameter values within the prior uncertainty range of the parameter values), then one of the other locally optimal experiments found should be considered, as it may be more robust to the uncertainty in the parameter values.

If a more robust design cannot be found, it is feasible to accept the current D-optimal experiment and the risk of it being inefficient, as the first part of a sequential parameter estimation process.

When designing experiments, selecting the best approach is very case dependent. Unfortunately no method is available that will indicate a priori which approach is best.

The method chosen in this work is to use the D-optimality criterion combined with a sensitivity analysis of the designed experiment to the parameter values. This approach has the advantage of being simple to implement while providing information about the quality of the experiment.

3.3.2 Local Optima

When designing experiments for the estimation of parameters in large models (as defined in the introduction) multiple local optima are usually present. This problem is not limited to the large models considered in this thesis but has also been observed with smaller nonlinear models. It was found that in general the magnitude of this problem is correlated to the size of the model, the number of responses used, the number of trials in the experiment and the number of parameters to be estimated.

To deal with local optima in the search for a global optimum a number of methods are available. These methods can only guarantee very good local optima, since there is no practical method that will always find the global optimum when working with large problems.

If the problem is considered from an optimization viewpoint, then the use of an algorithm that has been shown to work well when multiple local optima are present should be used. Simulated annealing is one such algorithm and it has been shown to work very well with problems such as the traveling salesman problem where a large number of local optima exist (Press et al., 1989). The disadvantages of this method are that it requires a greater number of function evaluations than other optimization methods and two tuning parameters need to be determined. How well the algorithm works is a function of the tuning parameters refered to as, the 'initial temperature' and the 'rate of cooling', and they are case specific. Further details about this method can be found in Press et al. (1989).

Another approach to dealing with local optima and perhaps the simplest is to design

multiple experiments, each with a different initial parameter values, and to then choose the best experiment. A limitation of using this method is the difficulty of determining how may different starting points should be used. A guideline used by the author to choosing this number is to design from p to $2p$ experiments with different initial points, where p is the number of parameters. The decision if more experiments should be designed, is a function of the number and distribution of unique local optima that are found. As an aid to making this decision the following guidelines are proposed.

1. If all of the experiments found are different but their criterion values are similar, any of the experiments is feasible and which one is chosen should be based on the results of a sensitivity analysis to the parameters.

2. If a subset of experiments is found that are similar, have similar criterion values and include the best experiment, any of the experiments in the subset can be chosen.

3. If all of the experiments found are different and their criterion values span a broad range, further experiments should be designed. If the experiment distribution does not change with the design of further experiments, the best experiment should be chosen.

4. If the best experiment is found more than once, it should be chosen.

The above are only guidelines used by the author based on experience. Although this approach is simple to apply it performs poorly if a large number of local optima exist and the method can be computationally expensive if a lot of different initial points are needed. There is also the difficulty of choosing an experiment if the results obtained are like point 3 above, which was found to be the common case.

Another possible solution to deal with local optima is to provide the optimization algorithm with a very good initial starting point that is hopefully in the neighborhood of the global optimum or a very good local optimum. This good initial starting point is

based on the information obtained from the sensitivity analysis described in Chapter 2, and has been found to work very well with the large models that are considered in this work. The reason why this approach is successful in generating a good initial starting point, is because the D-optimality criterion is a function of the gradients of the responses with respect to the parameters (i.e. the V^o matrix). Therefore by choosing locations where the gradient values are the largest, the maximization of the $|V^{o\prime}V^o|$ term will usually start at large values. The proposed method has been found in general to produce very good experiments, but it is not guaranteed to always do so.

To best describe the proposed method of how to generate a good initial starting point based on the sensitivity information, an example will be presented. This example will involve the estimation of the penultimate reactivity ratios (i.e. r_{11}, r_{12}, r_{21}, r_{22}, s_1 and s_2) for the system Styrene/Methyl Methacrylate using the responses of copolymer composition, cumulative triad fractions and conversion. The polymerization was simulated at 60 degrees Celsius and a 0.01 mol/L loading of the initiator AIBME. The designed experiment had 12 trials and the sample time was fixed at 300 minutes. In designing the experiment for this case it was found that multiple optima existed, since each starting point that was tried resulted in a unique locally optimal experiment. Figure 3.1 shows a sample of the locally optimal experiments that were found, where each row of circles represents an experiment and each circle represents an individual trial. The horizontal axis is the initial feed composition of styrene as a mole fraction for a given trial. For instance, the experiment represented by the row of circles with a criterion value of 251 (second from the bottom), has trials with the following feed compositions in mole fractions of styrene [0.16, 0.18, 0.36, 0.43, 0.59, 0.64, 0.68, 0.74, 0.75, 0.75, 0.77, 0.77]. A value proportional to the experiment D-optimality criterion is also given to the right of each row and will be referred to as the criterion value in the remainder of this discussion (the experiments are sorted by quality, where a larger criterion value indicates a better experiment).

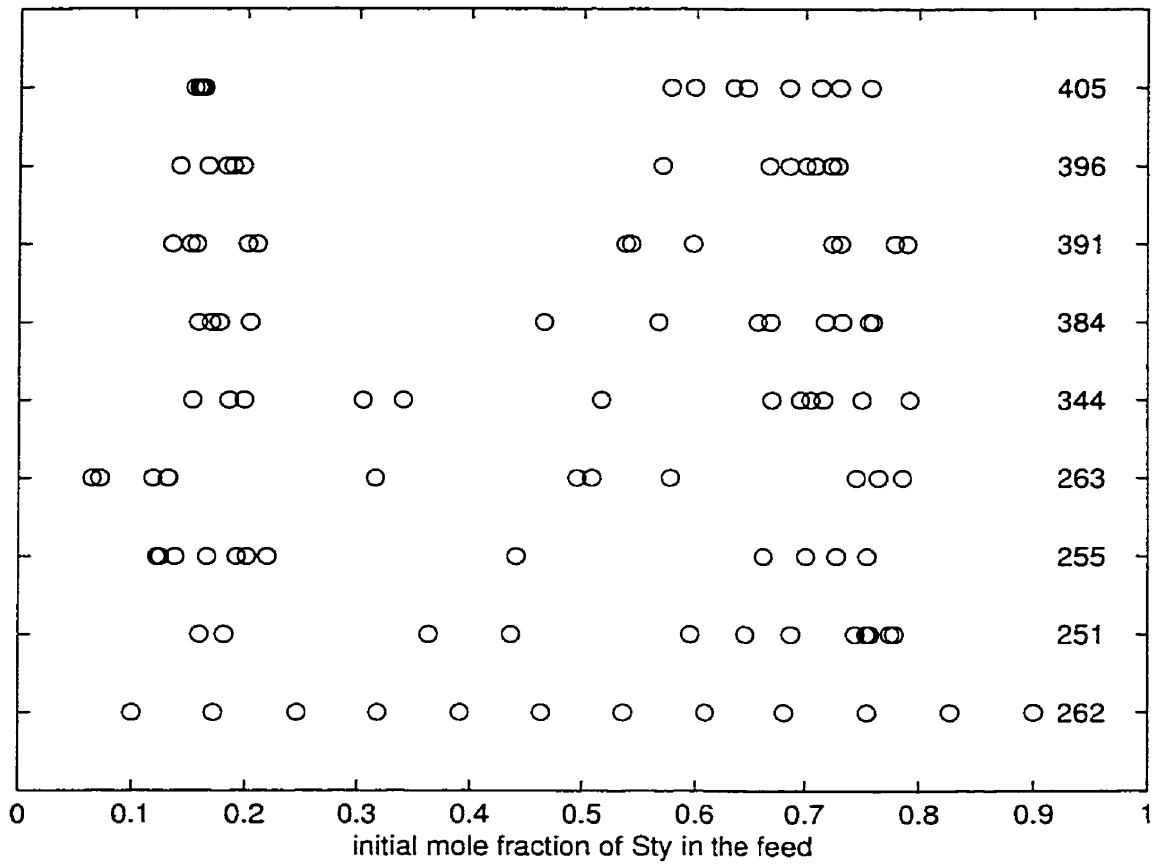


Figure 3.1: A sample of locally optimal experiments for the estimation of 6 parameters using a 12 trial experiment

Of the experiments shown in Figure 3.1, the best experiment found with a criterion value of 405 was obtained by using the information obtained from the sensitivity analysis to establish starting values for the optimization. The case where the initial values used in the optimization are an even distribution of points was also performed, and resulted in an experiment with a criterion value of 344. This experiment is the middle of the of the set of experiments found, with respect to experiment quality. The remainder of the experiments, except for the bottom one, were obtained using a random initial guess for the starting value.

The bottom row in Figure 3.1 is an experiment with evenly spaced initial feed compositions from 0.1 to 0.9 mole fraction. This experiment was included as a comparison of the empirical approach to the designed experiment approach. A comparison of the criterion values, 262 for the evenly spaced experiment versus 405 for the designed experiment, indicates that the designed experiment is better. Though how much better the designed experiment is versus the empirical one is cannot be determine based solely on the criterion values. Other factors that should be considered when performing the comparison are: The effect of the uncertainty in the parameter values, which will affect the robustness of the experiment, and that the criterion value is proportional to the volume of the linearized confidence region not the true confidence region. A further discussion on how to compare experiments is given in Section 3.4.

The best experiment was obtained using the gradient information to generate the optimization starting values. These values were obtained from an inspection of the gradient plots shown in Figure 3.2. This figure shows the sensitivity plots of the responses with respect to the parameters $r_{11}, r_{12}, r_{21}, r_{22}, s_1$ and s_2 at 300 minutes. The initial points are placed in areas where the response information is maximized. To determine where the amount of information from a response is maximized the areas where the gradient curves have the largest absolute magnitude are identified, as this is where the response is most sensitive to the parameter values. For example, this occurs near an initial feed of 0.8

mole fraction styrene for the parameter r_{11} with the responses of copolymer composition (F1) and cumulative triad fractions A111 and A211. For the parameter r_{22} it occurs near an initial feed of 0.2 mole fraction styrene with the responses of copolymer composition (F1) and the triad fractions A222 and A122.

The determination of the point of most information is not always as straightforward as for the parameters r_{11} and r_{22}. If the gradient plots for the parameters r_{12} and r_{21} are examined we can see that there is a region of high sensitivity for initial feeds from 0.4 to 0.6 mole fraction styrene. If we look at the gradient plots for the parameter s_1 we can see that for any initial feed value less that 0.6 mole fraction styrene, the information will be maximized, while for the parameter s_2 the smaller the mole fraction of styrene the better. In general, if the determination of the set of initial values is ambiguous, then several different sets of values should be used.

Based on the gradient plots the initial values chosen for the optimization were (0.2, 0.2, 0.5, 0.5, 0.8, 0.8, 0.2, 0.2, 0.5, 0.5, 0.8, 0.8). A combination of only three distinct values was used in the initial starting point for this case, as it is sufficient that the initial trials are in the neighborhood of the gradient value maxima. This is because the information from the gradient plots does not take into account the correlation between responses or between parameters, while the D-optimality criterion does.

3.3.3 Implementation Notes

This section will briefly list some of the observations made by the author in the process of designing experiments for the various cases studied.

- The gradient information was used to choose an initial guess for the experiment design optimization algorithm. Multiple initial guesses were tried if the information distribution in the gradients indicated that a range of sampling conditions were feasible and as a double check to ensure that a good experiment was obtained.

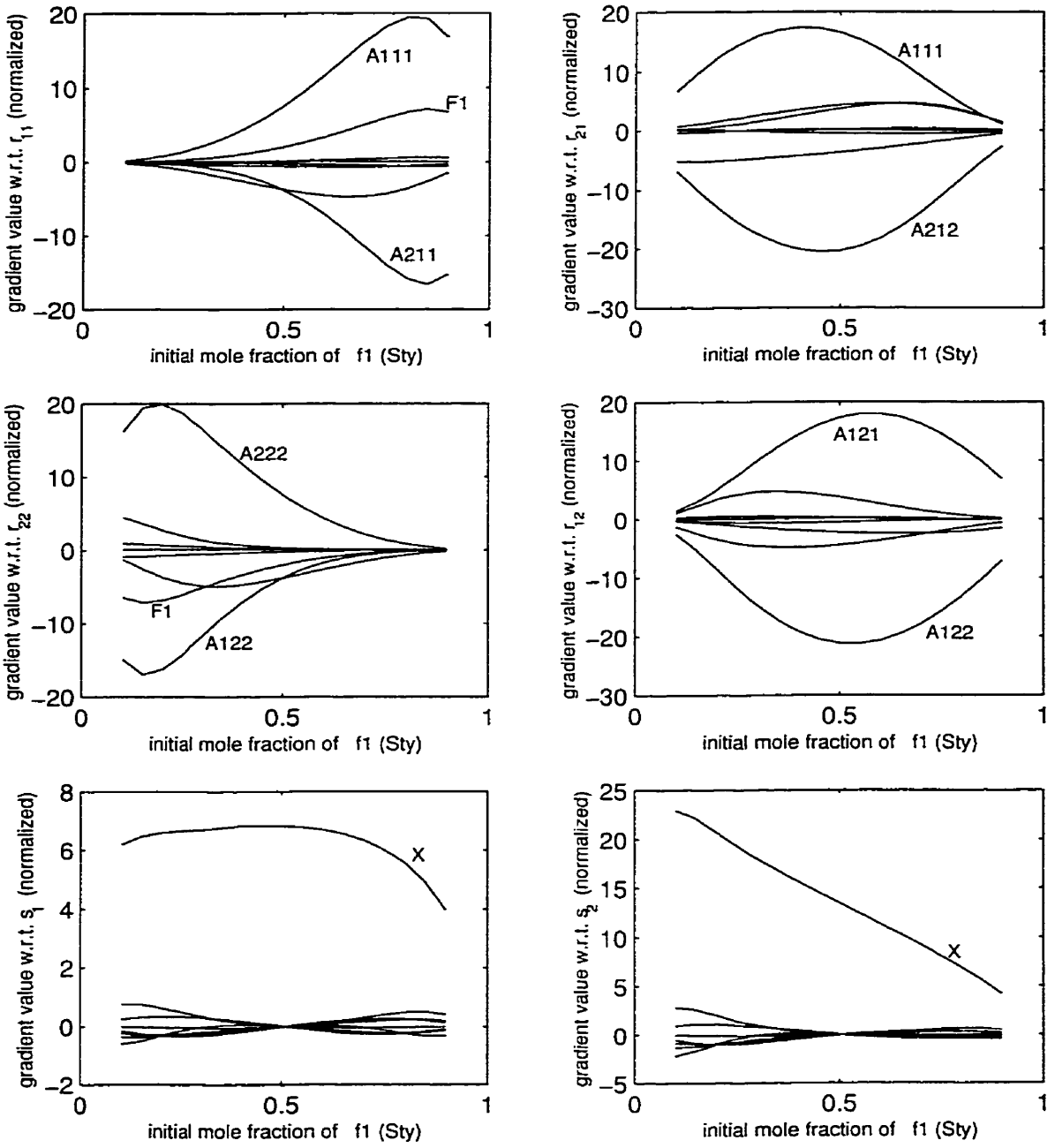


Figure 3.2: Gradient plots to generate an initial point for the experiment design optimization algorithm

- In implementing the multiresponse D-optimality criterion as given by equation 3.4, the off diagonal measurement covariance values (i.e. σ_{ij}) were assumed to be zero, while for the diagonal values of the covariance matrix (i.e. σ_{ii}), the square root of $1/\sigma_{ii}$ is included in the normalized gradient values that make up the V^o matrix. Therefore the term σ^{ii} (which in this case is equal to $1/\sigma_{ii}$) is included in the term $V^{o\prime}V^o$. This simplification is only possible because the off-diagonal terms in the covariance matrix are assumed to be zero.

- The optimization algorithm used was a modified simplex algorithm included in the Matlab optimization toolbox, the function `fmins.m`. This routine is very robust and could be used on a range of problems without intervention by the user. Though it is very slow to converge, it was felt that the advantages of robustness outweighed the slow convergence.

3.4 D-optimal Experiment Sensitivity Measure

The objective in performing a sensitivity analysis with respect to the parameter values on the designed experiment is to determine the quality of the chosen design over the range of possible parameter values. This analysis is recommended since the initial parameter values will rarely be equal to the true parameter values. If the quality of the experiment design is highly sensitive to the parameter values, then the designed experiment can result in parameter estimates with a low level of uncertainty being obtained.

To determine how sensitive the chosen experiment is to the parameter values it is desired to see how the design criterion will change as the parameter values change. By calculating the D-optimal criterion value of the chosen experiment with different parameter values, it is possible to obtain a qualitative measure of how well the experiment will perform with different parameter values. One issue in applying this method is deciding how much of a change in the criterion value is large enough to warrant a reevaluation

of the experiment design. Based on the author's experience, a rough guide is that if the criterion (normalized to one dimension from n-dimensions, see section 3.4.1) over a significant portion of the parameter space are less than 50

An effective method to summarize the experiment efficiency over a range of parameter values is to plot the D-optimality criterion of the designed experiment using a range of parameter values. If the number of parameters is greater than three this approach is no longer feasible, and a two-step approach is proposed. The first step is to calculate the D-optimality criterion over the parameter values arranged in a factorial layout, where the minimum and maximum feasible parameter values are the high and low values in the factorial experiment. The objective of the first step is to act as a screening design to determine where in the parameter space the chosen experiment will be inefficient. The second step is to focus on this area of the parameter space, by adding more points to obtain more information about the experiment efficiency. An example of this approach is given in case study 1, Section 6.3.3.

To illustrate why the designed experiment sensitivity analysis is an important and useful diagnostic in determining the quality of the designed experiment an example will be presented. The following simulation example involved the estimation of the reactivity ratios (r_1 and r_2) for the system Styrene/Methyl Methacrylate. The polymerization simulation was carried out using the conditions given in Appendix B. The response used was copolymer composition and there are four trials in the experiment design. The initial values used by the experiment design optimization routine were obtained using the information provided by the gradient plots as described in section 3.3.2. The four trial experiment that was designed based on the initial values from the gradient plot is shown in Table 3.2, and will be referred to as the 'high conversion' experiment.

Another experiment was also designed using initial values based on the conventional approach, the Tidwell-Mortimer method, (Tidwell and Mortimer, 1965). Using these values a low conversion experiment, that is a local optimum, was found. This experiment

	Initial feed (mol fr. Styrene)	Sampling time (minutes)	Criterion
high conversion	0.20	475	8917
	0.20	475	
	0.30	580	
	0.30	580	
low conversion	0.20	5	169
	0.20	5	
	0.80	5	
	0.80	5	

Table 3.2: Experiment designs and their D-optimal criterion values for the estimation of the reactivity ratios

is shown in Table 3.2 and will be referred to as the 'low conversion' experiment. A comparison of the two experiments based on the criterion values indicated that the high conversion experiment is a much better experiment as its criterion value is approximately sixty time larger than that of the low conversion experiment.

A sensitivity analysis was then performed on the high conversion experiment, where the D-optimality criterion was calculated using the chosen experiment over the parameter space. Since there are only two parameters to be estimated the D-optimality criterion was plotted as a function of the parameter values and is shown in Figure 3.3. In the upper plot of this figure the horizontal axes are the two parameters to be estimated and the vertical axis is the log of the criterion value. To facilitate the analysis, the log of the criterion values is used to compress the data and a contour plot of the surface is included (bottom plot). By examining Figure 3.3 we can see that the chosen experiment is good for a band of parameter values that correspond to a ridge across all values of r_2 and near a value of 0.55 for r_1. For parameter values to the left of the ridge (i.e. $r_1 < 0.5$), the chosen experiment is very poor as the criterion values decreases by five orders of magnitude from the ridge values. Thus if the true parameter values were in this range the chosen experiment would be very poor. While to the right of the ridge (i.e. $r_1 > 0.6$)

the experiment is still very poor but the decrease is not as large as to the left. A further observation is that the high conversion experiment is only sensitive to the r_1 parameter, since the ridge spans all of the values of r_2 considered.

As a comparison, the experiment design sensitivity analysis was performed on the low conversion experiment. The resultant D-optimality criterion surface is shown in Figure 3.4. The vertical axis of the upper plot is again the log of the D-optimal criterion value and the horizontal axes are the parameter values. The axes scales for both Figures 3.3 and 3.4 are the same so that a direct visual comparison can be made between the figures. The most important feature of Figure 3.4 is that the surface is flat. This implies that the low conversion experiment is not very sensitive to the parameters within the chosen ranges. While the 'low conversion' experiment is not as good as the 'high conversion' experiment, it is much more robust (due to the flatness of the sensitivity surface). Therefore which experiment is used depends on how well the parameter values are known a priori. In this case, if the chosen parameter ranges used in the figure represented the uncertainty in the parameters, then it is recommended that the low conversion experiment be used.

3.4.1 Comparing Experiments

The experiment sensitivity analysis requires that a number of experiments are compared. This is a very difficult task to accomplish as the common approach would be to linearly associate a change in the criterion value with a change in quality of the parameter estimates. That is, if the criterion value doubles, the experiment would be twice as good and the parameter estimates obtained would be twice as good. This is not correct as the D-optimality criterion is inversely proportional to the p-dimensional hypervolume of the linearized joint confidence region and a doubling of the criterion does not result in parameter estimates that are twice as good. Therefore the following method is proposed as an alternative to the direct comparison of criterion values. While this method makes a number of assumptions, it is intended as a qualitative measure that is intuitive and

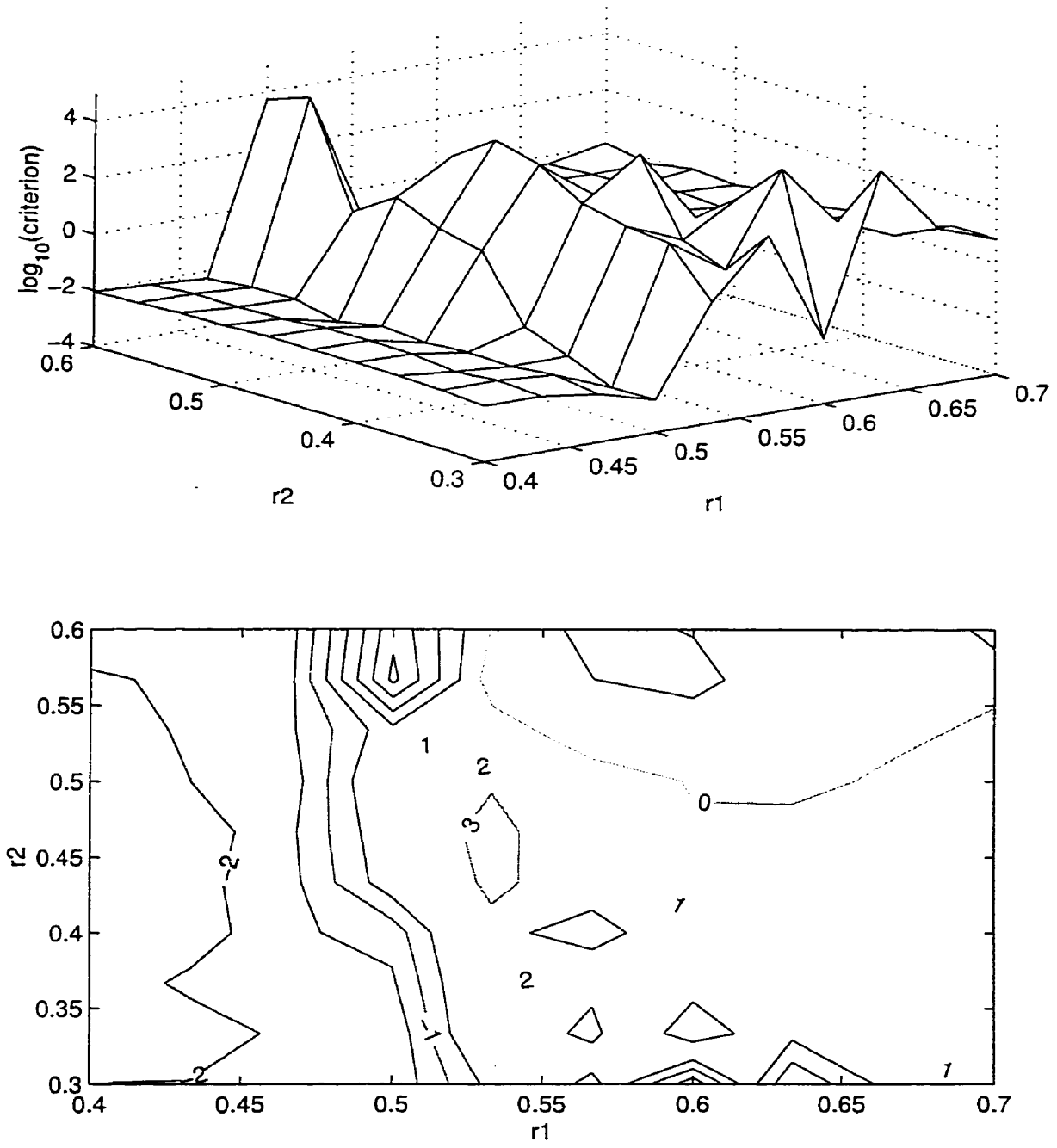


Figure 3.3: The D-optimal design criterion of the 'high conversion' experiment over the feasible parameter space

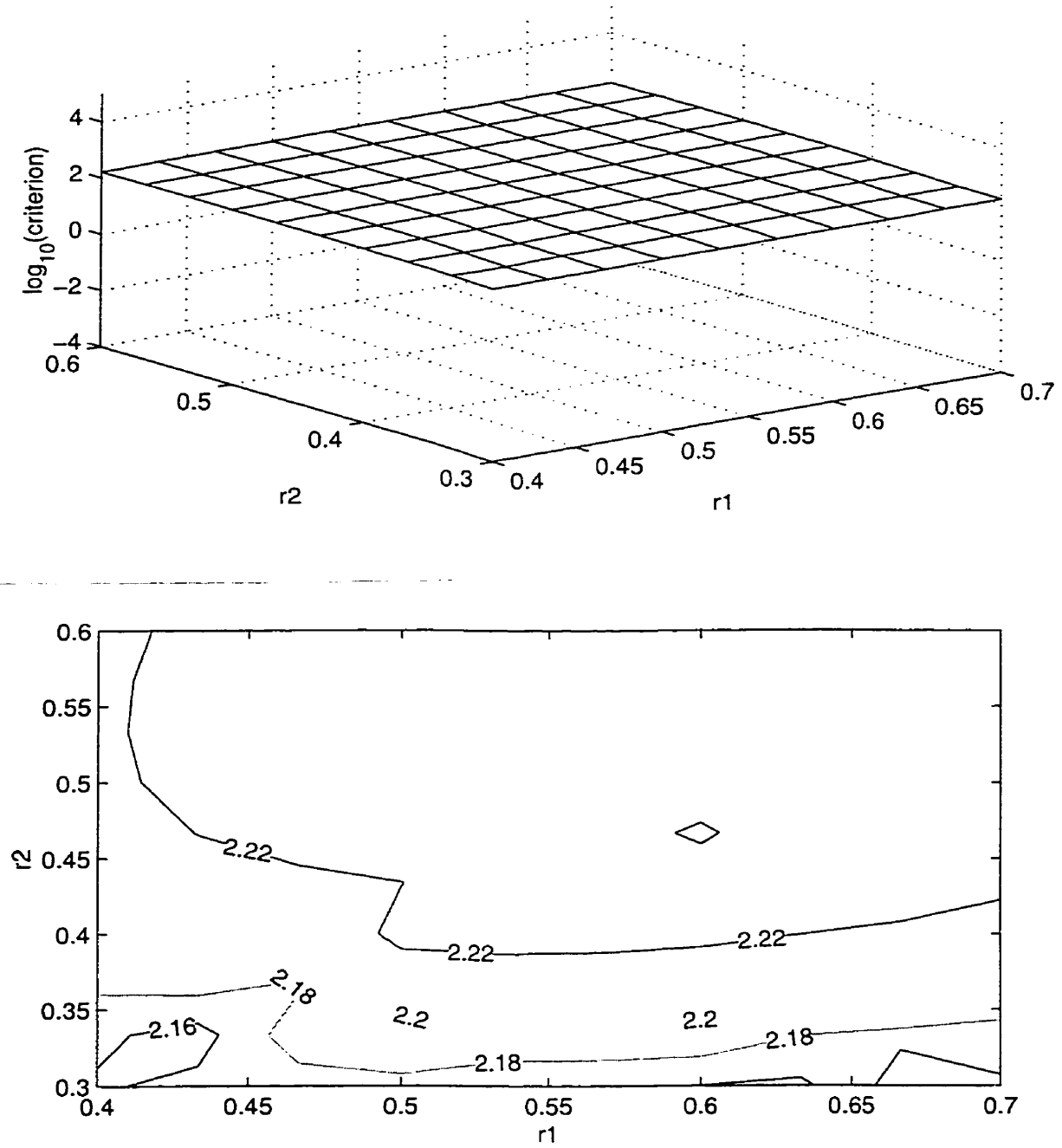


Figure 3.4: The D-optimal design criterion of the 'low conversion' experiment over the feasible parameter space

easy to interpret.

In performing a comparison of the different experiments obtained by spanning the parameter space, it is more intuitive to compare a value that is proportional to the standard deviation of the parameter estimates than the experiment criterion value which is inversely proportional to the volume of the linearized joint confidence region. Since the parameter standard deviation within a set of parameters can vary widely, it is desired to obtain a value that can be used for the comparison of experiments. To achieve this the following approximation is used. If it is assumed that the parameter estimate joint confidence region is a p-dimensional spheroid whose volume is proportional to the criterion value, then the radius of this spheroid will be proportional to the criterion value to the power of $(\frac{-1}{2p})$. This radius is then proportional to a value that can be referred to as a generalized standard deviation of the parameter estimates. This value (i.e. the radius) is a crude approximation but has the advantage of making the sensitivity analysis much easier to interpret. This is achieved because we are comparing a value that is linearly proportional to the standard deviation of the parameter estimates. Therefore if this value doubles, the uncertainty range of the parameter estimates will double. While this sensitivity analysis method makes a number of assumptions it is a feasible approach which is intended to be used as a qualitative measure of how the quality of the experiment will change as the parameter values change. It's application is shown in all three of the case studies discussed in Chapter 6.

Chapter 4

Parameter Estimation

The objective in nonlinear parameter estimation is to obtain a set of parameter values that are robust and show the least possible uncertainty and correlation. This can be a very difficult task to accomplish and is a function of a number of factors, which can affect the quality of the parameter estimates to different degrees. The principal factors are, the model structure, the quality of the measurements and the experiment design. Parameter estimation problem has been addressed by a number of authors such as Biegler et al. (1986), Bilardello (1993), Stewart et al. (1992), Watts (1994) and Ziegel and Gorman (1980) for example.

To estimate the parameters different criteria can be used. In Section 4.1 a listing of available criteria is given, as well as a discussion of the determinant criterion and the multiweighted least squares criterion (MWLS). The methodology of parameter estimation is discussed in Section 4.2. This includes a discussion of the optimization methods that can be used and the specific problems of local optima and correlation in the parameter estimates.

4.1 Estimation Criteria

Various criteria can be used for the estimation of parameters using multiple responses. A listing of available criteria is given in Table 4.1. All of these criteria are related in two ways. First, they are all a function of the residuals and second, they are all based on the same assumptions about the model and the error structure, as outlined below.

1. The model structure is correct.
2. The errors from trial to trial are independent of one another, within a trial they are assumed to be normally distributed with a mean of zero and a constant covariance matrix.
3. The measurement error is additive.
4. The measurement error is homoscedastic

The validity of the above assumptions is usually tested by an analysis of a plot of the residuals and a plot of the predicted and measured values. A failure of any of the above assumptions can result in erroneous parameter estimates. To deal with this failure the following methods are available;

- if 1 is violated, residual plots can be used to help adjust the model structure
- if 2 is violated, transformation of the data, (Box and Cox , 1964) or a Time Series Analysis can be used, (Box and Jenkins, 1976)
- if 3 is violated, transformation of the data (Box and Cox, 1964) or the model is required
- if 4 is violated, transformation of the data or use of an error model with heteroscedastic regression are required (Beal and Sheiner, 1988; Welsh et al., 1994)

Although the criteria in Table 4.1 are all related, there are distinct differences in their application and the assumptions that they further make. The least squares criterion is the oldest and most widely used criterion for parameter estimation in nonlinear models. When used in multi response parameter estimation it makes the further assumption that the magnitude of the measurement error of each response is similar. If this assumption

Criterion	Reference
least squares	Seber and Wild (1989)
weighted least squares	Seber and Wild (1989)
L_1-Norm	Oberhofer (1982)
determinant	Box and Draper (1965)
MWLS	Oxby (1997)

Table 4.1: Criteria for multiresponse parameter estimation

is not valid then a bias will exist in the parameter estimates obtained. To overcome this, the weighted least squares criterion was developed, where the weight for each response is usually the inverse of the standard deviation of the response measurement error. A limitation of weighted least squares is that the measurement error variance is required and it is usually assumed that there is no measurement error correlation. A detailed discussion of the least squares and weighted least squares criteria can be found in a number of statistical texts, such as Seber and Wild (1989). An extension of the least squares criterion is the L_1-Norm criterion. It takes the sum of the absolute deviations instead of the squared deviations. It was developed as a robust criterion to outliers, as it will be affected to a lesser degree by outliers than least squares. Its properties, assumptions and use are discussed by Oberhofer (1982) and Gonin and Money (1985). The determinant and MWLS criteria are discussed in sections 4.1.1 and 4.1.2 respectively.

The criterion chosen for the bulk of the simulation studies is the determinant criterion. It was chosen since it handles multiple responses when the response error covariance matrix is not know, which is the usual case. A limitation of the determinant criterion, is that it may not be appropriate when the sample size is small, as shown by Oxby (1997). To overcome this small sample size limitation Oxby (1997) proposes the use of the MWLS criterion.

4.1.1 Determinant Criterion

The most popular criterion used to estimate parameters from multiple responses is the determinant criterion as discussed by Box and Draper (1965). This section will present the derivation of the determinant criterion from a Bayesian approach. Its properties and application will also be discussed.

The determinant criterion was developed by Box and Draper (1965) using a Bayesian argument, though it can also be derived using a likelihood approach as described by Bates and Watts (1988), for example. The following is a description of the derivation of the determinant criterion using a Bayesian approach. Given the general multiresponse model.

$$y_{ui} = f_i(x_u, \theta) + \epsilon_{ui} \quad , i = 1 \ldots r, u = 1 \ldots n \tag{4.1}$$

where, i is the response number from 1 to r, u is the observation number from 1 to n, y_{ui} is the measured data point of trial u and response i, $f_i(x_u, \theta)$ is the expected value of response i at conditions x_u, x_u is the set of input variables for observation u, θ is the vector of parameters, and ϵ_{ui} is the random normally distributed error associated with the data point ui. To illustrate how the above general model relates to the model used in the case studies, y would be responses used such as conversion, composition and radical concentration; x would be input variables such as the initial feed composition or sampling time; and θ represents the parameters in the model such as $k_{fm\,act-eng}$ and $k_{fm\,pre-exp}$, these parameters are part of the Arrhenius expression for the rate constant of radical transfer to monomer (k_{fm}). Further information about these parameters can be found in the model description, Section 6.2.

Let the covariance matrix of the responses y_i, where $i = 1$ to r, be

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1r} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{r1} & \sigma_{r2} & \cdots & \sigma_{rr} \end{bmatrix} = \{\sigma_{ij}\} \tag{4.2}$$

and

$$\Sigma^{-1} = \{\sigma_{ij}\}^{-1} = \sigma^{ij}$$

Where Σ^{-1} is the inverse of the covariance matrix and σ^{ij} is the ij-th element of Σ^{-1}. If the covariance matrix of the responses were known, then the likelihood function would be a monotonic function of the quadratic form

$$\sum_{i=1}^{r}\sum_{j=1}^{r} v_{ij}\ \sigma^{ij} \tag{4.3}$$

Where v_{ij} is the sum of the product of the deviations of responses i and j given by:

$$v_{ij} = \sum_{u=1}^{n}[y_{ui} - f_i(x_u, \theta)]\ [y_{uj} - f_j(x_u, \theta)] \tag{4.4}$$

Using a Bayesian approach and assuming that the observations y_{ui} are independent from trial to trial, the likelihood is given by

$$p(y|\theta, \Sigma^{-1}) = (2\pi)^{-\frac{1}{2}nr}|\Sigma^{-1}|^{\frac{1}{2}n} exp\left(-\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} v_{ij}\ \sigma^{ij}\right) \tag{4.5}$$

Minimization of equation (4.5) is equivalent to weighted least squares. Assuming that θ and Σ^{-1} are independent, the prior distribution for these parameters can be expressed as:

$$p(\theta, \Sigma^{-1}) = p(\theta)p(\Sigma^{-1}) \tag{4.6}$$

In their development Box and Draper (1965) assume that a locally uniform prior may be used for θ,

$$p(\theta) \propto d\theta \tag{4.7}$$

They use the invariance theory of Jeffreys (1961) to produce a noninformative prior for Σ^{-1},

$$p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{\frac{r+1}{2}} \tag{4.8}$$

The posterior *pdf* of the parameters, equation (4.9), is obtained by combining the likelihood (equation (4.5)) with the noninformative priors for θ and Σ^{-1} (equations (4.7) and

(4.8)).

$$p(\theta, \Sigma^{-1}|y)d\theta \prod d\sigma^{ij} \propto (2\pi)^{-\frac{1}{2}nr}|\Sigma^{-1}|^{\frac{1}{2}(n-r-1)}exp\left(-\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} v_{ij}\ \sigma^{ij}\right) d\theta \prod d\sigma^{ij} \quad (4.9)$$

We are interested in θ not Σ^{-1}, and therefore to obtain the marginal distribution for θ, we must integrate out Σ^{-1}. This is achieved by comparing the right hand side of equation (4.9) to the Wishart distribution, (Box and Tiao 1973). This leads to the following:

$$p(\theta|y) = C\ |\mathcal{V}|^{-\frac{1}{2}n} \quad (4.10)$$

where C is the normalizing constant and $\mathcal{V}$ is the matrix of the elements v_{ij}, obtained from equation 4.4.

$$\mathcal{V} = \begin{bmatrix} v_{11} & \cdots & v_{1j} \\ \vdots & \ddots & \vdots \\ v_{i1} & \cdots & v_{ij} \end{bmatrix} \quad (4.11)$$

Thus to obtain the parameter estimates the right side of Equation 4.10 is maximized, which corresponds to minimizing $|\mathcal{V}|$, the determinant of the dispersion matrix. Bates and Watts (1988) use the notation of Z to represent the matrix of deviations, given by:

$$Z = Y - f(x, \theta) \quad (4.12)$$

Therefore the $\mathcal{V}$ used by Box and Draper corresponds to $Z'Z$ used by Bates and Watts. The Z notation will be the one used in this thesis. The determinant criterion has a number of favourable properties as outlined by Box and Tiao (1973). These properties are:

- the expectation function can be linear or nonlinear
- the parameters can be common to more than one response
- the design variables can be common to more than one response
- the responses used can be rescaled or a linear combination of responses can be used

The application of the determinant criterion is discussed at length by Bates and Watts (1988). They discuss numerical stability issues such as taking the QR decomposition of the Z matrix so that the determinant of $Z'Z$ is numerically easier to calculate (i.e. more numerically stable). They also discuss different optimization methods to obtain the parameter estimates, such as the Newton-Raphson and Levenberg-Marquardt methods. It was found that these optimization methods, while performing well on smaller nonlinear parameter estimation problems, were not suitable to the type of problems considered in this thesis. This is due to the large number of local optima that are present.

4.1.2 MWLS

Oxby (1997) found that the use of the determinant criterion can result in poor parameter estimates when the sample size is small. To deal with this limitation he proposed a new criterion, the Multivariate Weighted Least Squares (MWLS) criterion. He shows, using Monte Carlo simulations, that the proposed criterion is a more robust criterion for multiresponse parameter estimation if a small sample size is used. As the sample size increases, the difference in quality (i.e. how well the criterion can estimate the parameter values) between MWLS and the determinant criterion decreases and eventually becomes zero. Then as the sample size increases further the determinant criterion becomes a better estimator. Oxby observed that the magnitude of the difference in quality between the MWLS criterion and the determinant criterion is case dependent. He suggests that the MWLS criterion be used for all cases when the sample size is small. A detailed discussion of the development of the criterion and its properties are given by Oxby (1997). A brief description of the implementation of the criterion follows.

The MWLS criterion is a two step process where the algorithm iterates between the two steps until convergence is obtained (i.e. the parameter values do not change at the desired tolerance between iterations). In the first step, the model parameters ($\hat{\theta}$) are obtained by changing the parameter values so that the weighted sum of squared

deviations is minimized, shown in equation 4.13).

$$min.\ \ tr[Z(\theta)WZ(\theta)'] \tag{4.13}$$

Where $Z(\theta)$ is the matrix of residuals as defined in equation 4.12 and W is a diagonal matrix of weights where each of the elements corresponds to a response. The W matrix is obtained in the second step of the algorithm. Therefore for the first iteration, where calculated values of the W matrix are not available, the W matrix is set to the identity matrix. The elements of the W matrix are obtained using equation 4.14.

$$W = \left[diag.\left(\frac{Z(\hat{\theta})'Z(\hat{\theta})}{n}\right)\right]^{-1} \tag{4.14}$$

Where $Z(\hat{\theta})$ is the matrix of residuals using the parameter estimates obtained in step one, equation 4.13, and n is the number of measurements per response. Equations 4.13 and 4.14 are then iterated to convergence.

The major difference between the determinant criterion and MLWS is that the MWLS criterion is more robust, with respect to parameter estimates, when the sample size is small. Oxby (1997) provides a detailed explanation of why this occurs. A simple explanation of the determinant criterion's poor performance is due to its assumptions that the residual covariance matrix is equal to the error covariance matrix. When the sample size is small the residual covariance matrix will likely be a poor approximation of the error covariance matrix, thus producing poor parameter estimates. A further difference between the criteria, is that the MWLS is more expensive to compute. In general, the computation time required for one iteration of the MWLS criterion is approximately equal to that required for the whole determinant criterion, though the MWLS criterion will usually converge in less than ten iterations.

4.2 Estimating Parameters

This section will discuss some of the practical aspects of implementing the determinant criterion related to local optima, parameter correlation and ill-conditioning of the determinant matrix. In general it was found that the parameter estimation problem is a difficult optimization problem that requires a robust optimization algorithm. The different optimization methods considered are discussed in Section 4.2.1. The particular problems of local optima and parameter correlation are addressed in Sections 4.2.2 and 4.2.3, respectively.

4.2.1 Optimization Methods

For the estimation problem considered in this thesis, involving relatively large numbers of response variables and parameters, it was found that a robust optimization method is required because the objective function surface is very complex exhibiting ridges and multiple local optima. In general, the large number of local optima present can trap in a local minimum most of the classical algorithms that follow a downhill path, such as the Newton-Raphson or BFGS algorithms. In general the only time that these algorithms will succeed is if the initial guess is such that there is no local optima in the algorithm's path. This can be the case when the initial guess is very close to the estimated parameter values.

The problem of parameter correlation while not directly causing the optimization algorithm to produce erroneous results, can result in slow convergence or instability in the algorithm. This is because the parameter correlation will cause a steep curved valley to occur on the objective function surface, which is a classic optimization problem (e.g. the Rosenbrock function; Fletcher, 1987). While methods have been developed that will adequately deal with the curved valley, they are not adequate to deal with the optimization problem in question due to the ridges and multiple local optima also

present in the objective function surface. Further considerations to take into account in the choice of an optimization algorithm are the high dimensionality of the problem if multiple parameters are to be estimated, and the expense of function evaluations of a large model.

In choosing an optimization method all of the above characteristics of the problem need to be taken into account. These characteristics make this a difficult optimization problem, as a method is required that is both robust with respect to local optima and that will minimize the number of function evaluations required. These two requirement are contradictory. Methods that are robust to local optima usually require a large number of function evaluations, while methods that are very efficient in minimizing the number of function evaluations do not perform well when local optima are present.

In choosing an optimization algorithm, the objective was to obtain good parameter estimates in a reasonable amount of time. To achieve this the simplex method and simulated annealing were used. Most of the time the simplex algorithm was used as it is the faster of the two, though it is less robust to local optima.

In the simplex method a triangle of $p+1$ dimensions in the objective function space, where p is the number of parameters to be estimated, is used. This triangle is then moved along the objective function surface by flipping, and resizing as required so that the new points chosen always result in a decrease in the objective function. A more detailed description of the algorithm and the details of its implementation are given by Press et al. (1989).

The simulated annealing algorithm is very robust to the folding in the objective function surface and can deal reasonably well with the problem of local optima due to the nature of the algorithm. In the simulated annealing algorithm, a random step within the defined parameter boundaries is chosen from the current point and the objective function is calculated. If the point is better, it is kept and becomes the current point. If it is worse, it may still be kept with a given probability. The size of the step and the

probability of accepting a worse step are decreased as the optimization proceeds. The initial values of the step and probability of accepting a worse step, and their rate of decrease, are determined by the optimization tuning parameters.

The limitations of the simulated annealing algorithm are that it is computationally expensive due to the large number of function evaluations that it requires, and that two tuning parameters have to be set within the optimization algorithm. While the algorithm may perform well with the default tuning parameters, this is case dependent and its performance will vary. A more detailed description of the algorithm and the details of its implementation are given by Press et al. (1989).

In using any algorithm the best that can be achieved is to obtain a very good local optimum, that hopefully is the global optimum. While with smaller models it is possible to determine if the global optimum was obtained, this is not possible with large models.

4.2.2 Local Optima

Multiple local optima can occur when estimating parameters in any nonlinear model, though it was found that they were the norm when estimating parameters in large models. To illustrate the magnitude of the problem of local optima with large models, an example will be presented.

This example is taken from the first case study, a simulation of Styrene homopolymerization where five parameters are estimated. The objective function surface is a five dimensional surface in this example. By fixing all of the parameters at the point estimate values and varying only one parameter at a time, a cross section of the objective function surface along the axis of the chosen parameter is obtained. A cross section with respect to the parameters $k_{fm\,act-eng}$ and $k_{fm\,pre-exp}$ was generated. These parameters are part of the Arrhenius expression for the rate constant of chain transfer to monomer. Further information about these parameters can be found in the model description given in Section 6.2. Figures 4.1 and 4.2 show a slice through the objective function surface while

varying the parameters $k_{fm\,act-eng}$ and $k_{fm\,pre-exp}$. In these figures, the horizontal axis is the value of the parameter being varied, the vertical axis is the criterion value of the objective function and the circle indicates the parameter point estimate. Both plots have been truncated on the vertical axis at a criterion value of 1400 and 1500, respectively. The convoluted surface and local optima create numerous difficulties for most optimization algorithms. In this case the simplex algorithm did not perform that well, as seen by the location of the point estimates.

An inspection of Figures 4.1 and 4.2 reveals that a large number of local optima are present and that the point estimate has converged to a local optimum. This observation provides valuable information about the magnitude of the optimization problem and therefore will suggest a course of action in obtaining the parameter estimates. In this case the large number of local optima suggests the use of an optimization algorithm that is more robust to local optima, such as simulated annealing. This optimization method might be more effective in the long run even though it is not as efficient as the simplex algorithm. The large number of local optima also indicates that some type of diagnostic should be run on any local optimum found. This diagnostic, such as the generation of cross sections of the criterion surface as shown in figures or an evaluation of the criterion in the neighborhood of found optimum, will indicate to the researcher if further optimization is required to obtain the parameter estimates.

4.2.3 Parameter Correlation

Parameter correlation can make the parameter estimation problem more difficult. This is because it usually produces a curved valley in the objective function surface when working with nonlinear models. A number of factors can cause the parameter estimates to be correlated. This may be due to the quality of the data collected (i.e. the experiment design), or the model structure. In certain cases it is possible to identify the model structure that is causing the parameter correlation, such as an Arrhenius relationship,

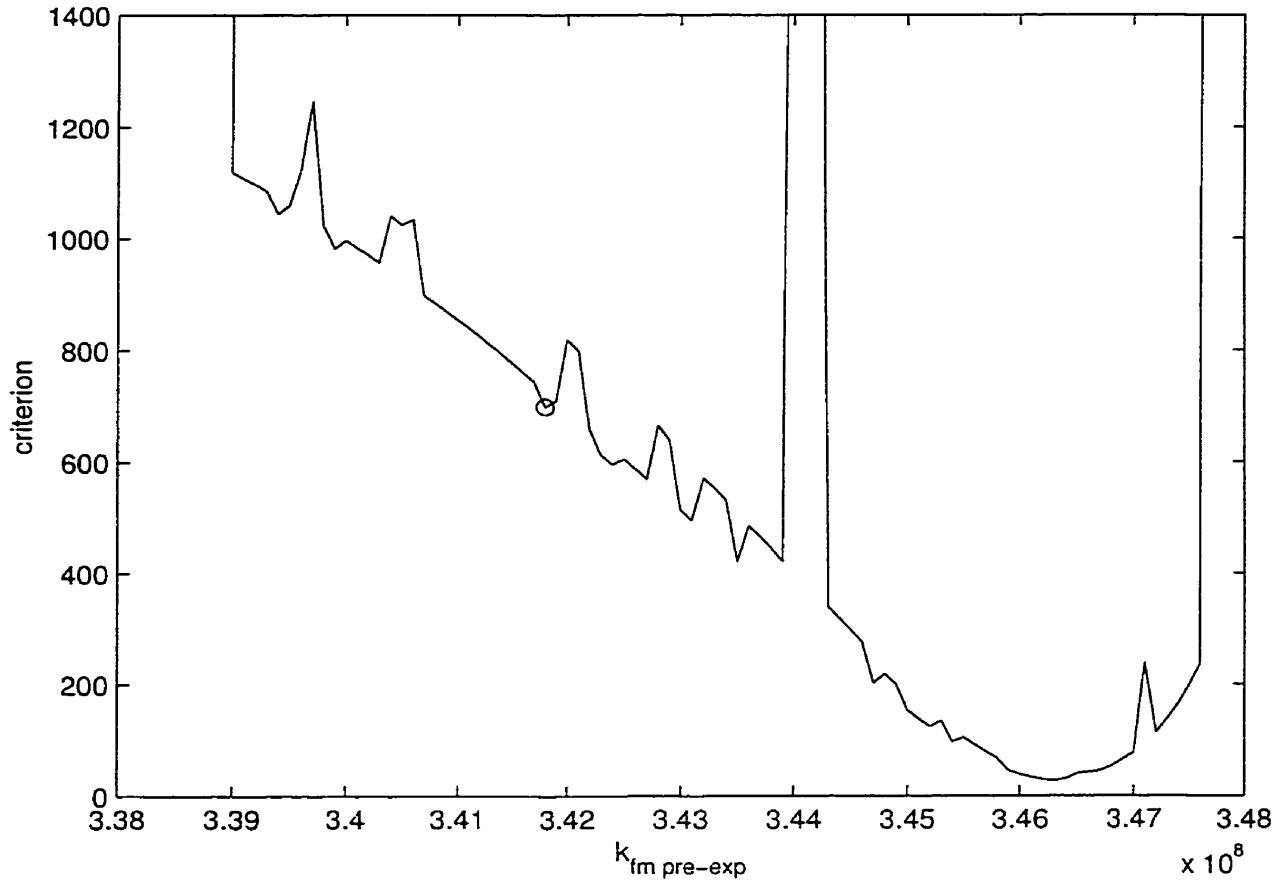


Figure 4.1: A cross section of the objective function surface, while varying $k_{fm\,pre-exp}$

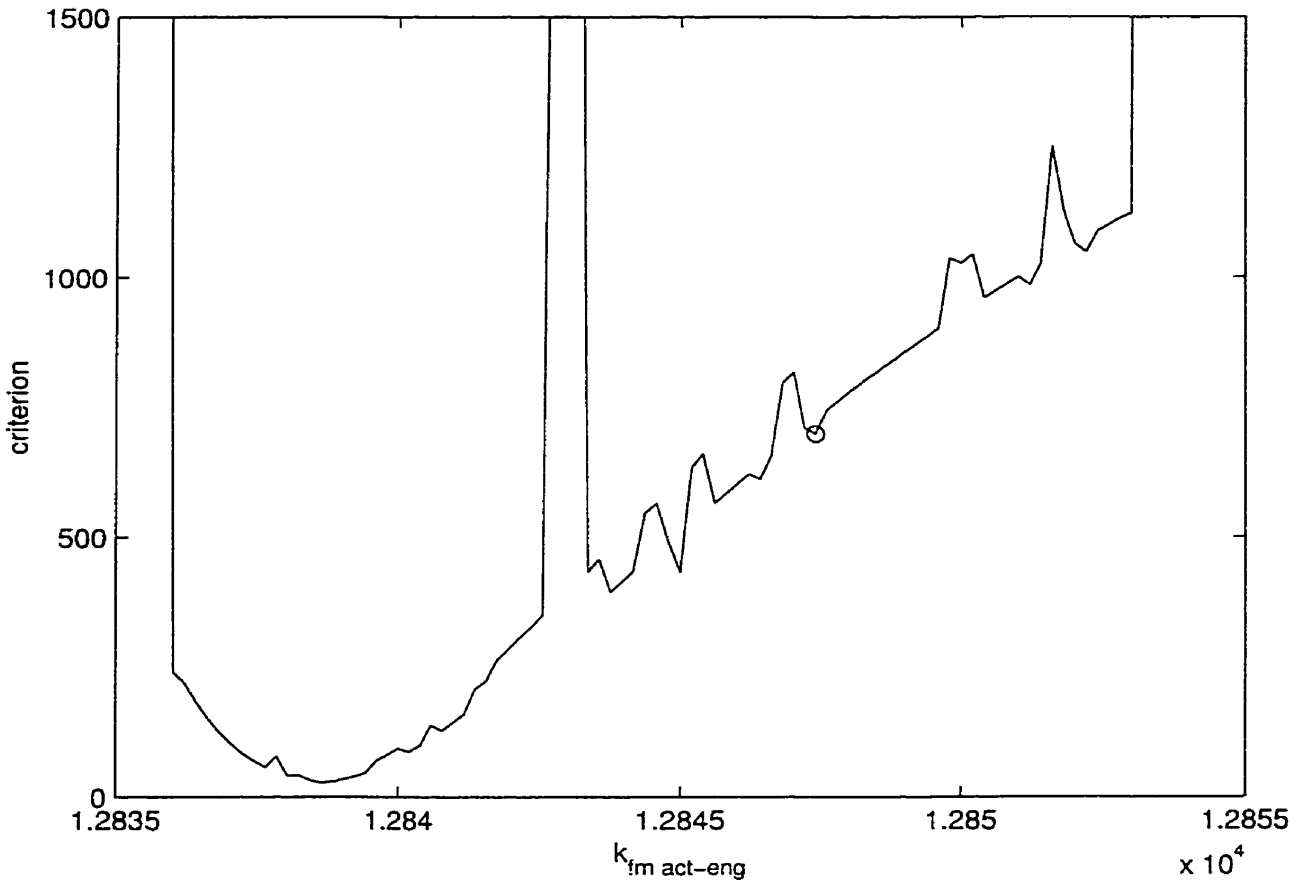


Figure 4.2: A cross section of the objective function surface, while varying $k_{fm\ act-eng}$

while in other cases it is not readily apparent. To deal with parameter correlation a number of options are available. If the correlation is due to model structure the model can be reparameterized. If it is due to the data, more data at different locations or from different responses can usually reduce the amount of correlation, (Box and Draper, 1965).

While parameter correlation can make it difficult to obtain independent parameter estimates, it should be noted that if only a good prediction from the model is desired, then the correlation of the parameter estimates may not be a significant problem.

To illustrate how parameter correlation can affect the parameter estimation process, an example will be given. This example is based on the estimation of the parameters $k_{p\,pre-exp}$ and $k_{p\,act-eng}$, within the Watpoly model for the simulation of Styrene homopolymerization. A section of the objective function surface to estimate the parameters is shown in Figure 4.3. In this figure, the horizontal axes are the values of the parameters and the vertical axis is the log of the criterion value. The log of the criterion was taken to compress the vertical axis to show more of the surface features. An inspection of Figure 4.3 reveals two characteristics that will make the parameter estimation process very difficult. The first is the very steep curved valley, and the second are the multiple rows of peaks parallel to the steep valley (these rows of peaks are actually ridges and appear as peaks due to the resolution of the data used to generate the surface). The rows of peaks will produce local optima where the optimization algorithms can get stuck.

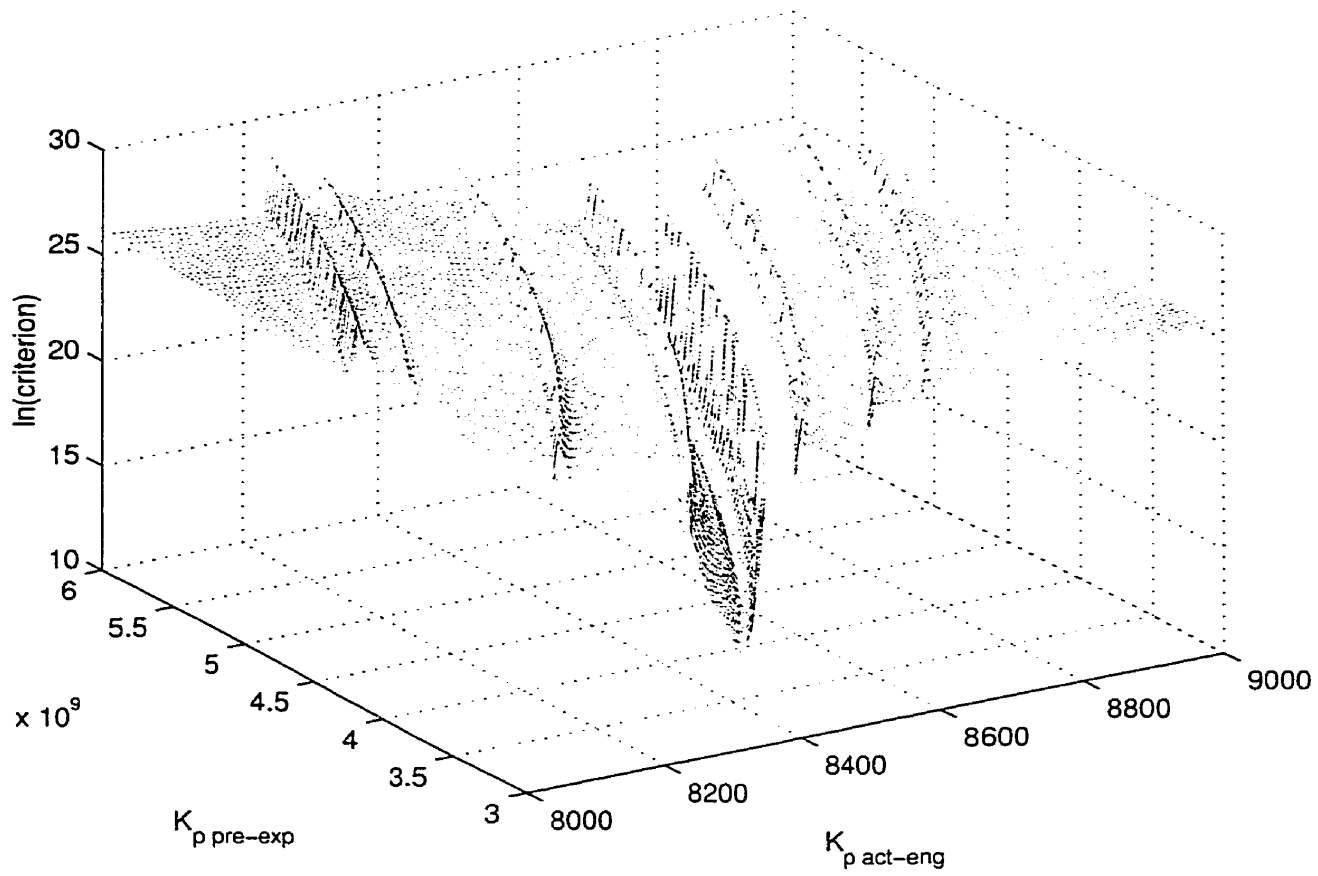


Figure 4.3: A plot of the objective function surface, while varying $k_{p\,pre-exp}$ and $k_{p\,act-eng}$

Chapter 5

Confidence Regions

Whenever parameter estimates are obtained there is a given amount of uncertainty in these estimates. Some of the possible reasons for this uncertainty are, the experiment design, the measurement error, the number of data points and the observability of the parameters.

To examine the uncertainty in a given set of parameter estimates, the joint confidence regions of the parameter estimates must be obtained. In analysing the amount of uncertainty, the use of joint confidence regions is recommended over individual parameter confidence bounds, since the confidence bounds will in general overestimate the uncertainty present in the parameter estimates (Draper and Guttman, 1995).

When nonlinear models are used, the joint confidence regions can be obtained in a number of different ways. These are:

- linear approximation (i.e. ellipsoidal regions)
- true shape, approximate confidence level
- true shape, true confidence level

The above methods are listed in increasing order of accuracy which also corresponds to an increasing amount of computation required.

Section 5.1 will discuss the different methods available for obtaining confidence regions and review the theory. A discussion of the issues in using true shape approximate confidence level confidence regions with multiple responses is given in Section 5.2. This section will present the case when the joint confidence regions obtained are at a confidence level that is very different than the expected confidence level. To illustrate this, two examples will be presented and discussed. Obtaining the joint confidence regions can be a difficult task. Section 5.3 will discuss the implementation of methods for obtaining confidence regions.

5.1 Background Theory

This section will review the theory of confidence regions. Only an overview will be presented, as the detailed proofs and discussion can be found in various sources such as, Seber and Wild (1989) and Bates and Watts (1988), for example.

As an introduction to confidence regions, the linear case will be presented. The nonlinear and multiresponse cases are direct extensions of it. Given the linear model

$$y = X\beta + \epsilon \tag{5.1}$$

where y is a vector of measured responses, X is the design matrix, β is the vector of parameters to be estimated and ϵ is a vector of *iid* normally distributed errors, the least squares estimate of β, $\hat{\beta}$, is then given by

$$\hat{\beta} = (X'X)^{-1}X'y \tag{5.2}$$

Under the above assumptions, the estimator of β, is normally distributed with a covariance matrix given by $\sigma^2(X'X)^{-1}$ and the $(1-\alpha)$ joint confidence region is then given by

$$(\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \leq p\, s^2\, F(p, n-p, \alpha) \tag{5.3}$$

where n is the number of trials, p is the number of parameters and $F(p, n-p, \alpha)$ is the value of the F distribution at a confidence level α with n and $n-p$ degrees of freedom. The estimate of the measurement error variance (i.e. s^2) is given by

$$s^2 = \frac{S(\hat{\beta})}{n-p} \tag{5.4}$$

where $S(\hat{\beta})$ is the residual sum of squares.

$$S(\hat{\beta}) = [y - x\hat{\beta}]'[y - x\hat{\beta}] \tag{5.5}$$

In equation 5.3 the F distribution is used because the terms $\{(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})\}$ and $\{s^2\}$ are both assumed to be χ^2 distributed with p and $n-p$ degrees of freedom, respectively. Therefore their ratio will be F distributed. To obtain the exterior surface of the joint confidence region at a given confidence level, the right hand side of equation 5.3 is fixed and the left hand side is solved for all possible parameter values. The joint confidence region that is produced will be a hyperellipsoid with dimensions equal to the number of parameters considered. Note that points on or inside the hyperellipsoid represent plausible values of β. Normally, if more than two parameters are estimated, then a two parameter conditional joint confidence region is calculated. This confidence region is obtained by fixing the remaining parameters at their point estimate values. Then the confidence contour is obtained by solving for the set of values of the two chosen parameters that satisfy the equality in equation 5.3.

The nonlinear case is a direct extension of the linear case and will be discussed next. Given the nonlinear model,

$$y = f(x, \theta^*) + \epsilon_i \tag{5.6}$$

where y is the measured value, $f(x, \theta^*)$ is the model response and ϵ is the measurement error, the estimate of $\hat{\theta}$ is obtained by minimizing the sum of squared deviations given by:

$$S(\hat{\theta}) = [y - f(x, \hat{\theta})]'[y - f(x, \hat{\theta})] \tag{5.7}$$

$S(\hat{\theta})$ defines a p dimensional surface and is referred to as the 'sum of squares surface'. If the assumption is made that in the neighbourhood of $\hat{\theta}$ a linear Taylor series expansion is an adequate approximation of the sum of squares surface, then the joint confidence region of the estimated parameters is given by:

$$(\theta - \hat{\theta})' V' V (\theta - \hat{\theta}) \leq p\, s^2\, F(p, n-p, \alpha) \tag{5.8}$$

Where V is the matrix of first derivatives, or Jacobian, of the model $f(x_i, \theta)$ with respect to the parameters, given by:

$$V_{ij} = \frac{\partial f(x_i, \theta)}{\partial \theta_j} \tag{5.9}$$

The above V matrix is analogous to the X matrix in the linear case. If the linear approximation is adequate then the ellipsoid generated by equation 5.8 will be an accurate representation of the true joint confidence regions. The linear approximation of the sum of squares surface may not be adequate due to the intrinsic curvature and/or parameter-effects curvature present. An analysis of this curvature is discussed by Clarke (1987), Cook and Goldberg (19986) and Cook and Witmer (1985). If the linear approximation is used, a test to determine if it is adequate should be carried out. To accomplish this, Bates and Watts (1980) propose the use of profile-t plots. These plots will show the curvature of the primary axis of the joint confidence region. If the linear approximation is adequate then the profile-t plot will not deviate significantly from a straight line. If the linear approximation is not adequate, as shown by the profile-t plots, then the joint confidence region is obtained using equation 5.10, (Beale, 1960).

$$S(\theta) - S(\hat{\theta}) \leq p\, s^2\, F(p, n-p, \alpha) \tag{5.10}$$

The above equation will produce the true shape of the confidence region but only at the approximate confidence level.

The nonlinear multiple response case is an extension of the single response case presented previously. If a multiresponse model with n trials and m responses is used,

$$y_{ui} = f_i(x_u, \theta) + \epsilon_{ui} \quad , i = 1 \ldots m,\ u = 1 \ldots n \tag{5.11}$$

the linear approximation of the joint confidence region is given by,

$$(\theta - \hat{\theta})' \frac{\Omega}{2} (\theta - \hat{\theta}) \leq p\, s^2\, F(p, n-p, \alpha) \tag{5.12}$$

where Ω is the Hessian of the objective function (i.e. $|Z'Z|$) with respect to θ evaluated at $\hat{\theta}$, and where Z is the matrix of deviations given by,

$$Z_{ui} = y_{ui} - f_i(x_u, \hat{\theta}) \quad , i = 1 \ldots m, \;\; u = 1 \ldots n \tag{5.13}$$

The exact shape approximate level joint confidence region is given by,

$$|Z'Z| - |\hat{Z}'\hat{Z}| \leq p\, s^2\, F(p, n-p, \alpha) \tag{5.14}$$

The true shape and true confidence level regions can be obtained using two methods. The first is based on the likelihood function and involves two steps. The first step is to integrate the area under the likelihood function from negative to positive infinity. This will allow the determination of the likelihood value that corresponds to the desired confidence level. The second step involves taking the contour of the likelihood function surface at the likelihood value determined in step one. A practical modification of this approach is to integrate from parameter values that produce very low likelihood values on each side of the parameter point estimate. A limitation of this approach to obtain the confidence regions is that is can easily become infeasible if the number of parameters is large. This is due to the amount of computation required.

The second method that can be used to obtain true shape, true level joint confidence regions is based on a Monte Carlo type approach like the Gibbs sampler (Cassela and George, 1992). A limitation of this approach is similar to that of the previous method in that the amount of computation required with large models may be too large to be practical. Although it can be significantly less than that required with the integration method.

A further point to consider when analysing confidence regions, is that confidence regions are a function of the quality of the data and the observability of the parameters.

For a given experiment design, the quality of the data obtained between experiments can vary. Therefore the shape of the joint confidence region will be different for different experiments. This effect is amplified if the number of sample points in the experiment is small, the parameters are not very observable, or the measurement error is large. To illustrate this, the joint confidence regions of the two parameters in the biological oxygen demand (BOD) model used in Bates and Watts (1988) will be shown. The BOD model is given below.

$$f(x, \theta) = \theta_1(1 - e^{-\theta_2 x}) \tag{5.15}$$

Where x is the time in days, and θ_1 and θ_2 are the parameters to be estimated. Based on the same experiment design, five different sample data sets with different measurement errors were generated. For each of these data sets, parameter estimates and their joint confidence regions were obtained. These confidence regions are shown in Figure 5.1. The true parameter values are indicated by the circle, and the stars are the parameter point estimates from each data set. As can be observed from Figure 5.1, the size and shape of the joint confidence region can vary by a large amount with different data sets. Therefore, if confidence regions are used as a measure of quality in a simulation analysis, a large number of confidence regions should be generated. If only a couple of confidence regions are used, due to chance the incorrect conclusion may be derived.

5.2 Issues in Using True Shape Confidence Regions With Multiple Responses

When parameter estimates are obtained, joint confidence regions are used to determine the quality of the estimates. For nonlinear models it is recommended that the true shape approximate level joint confidence region be obtained. This approach is recommended as the true shape will take into account the nonlinearity of the model and the approximate level is used to significantly reduce the computation required if large models are used.

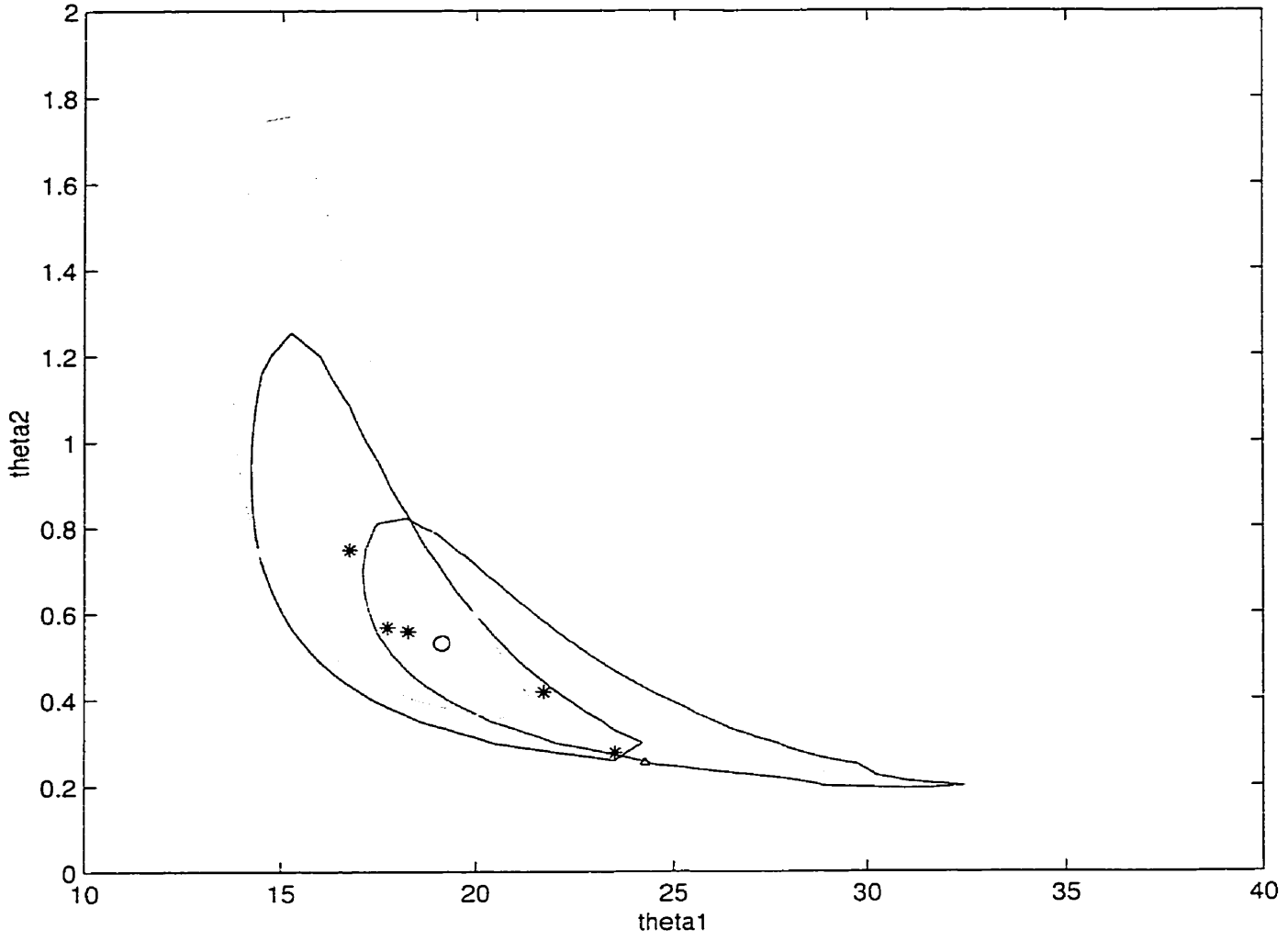


Figure 5.1: Joint confidence regions of five different cases, of the parameters θ_1 and θ_2 in the BOD model.

A limitation of the true shape approximate level joint confidence region method when multiple responses are used, is that the confidence level of the joint confidence region obtained can be significantly different than expected when the sample size is small. This difference can be up to 70 percent as observed in case studies. Therefore, there is a failure of the confidence region theory to accurately represent the uncertainty present in the parameter estimates.

To explain why the confidence region theory fails for small sample sizes when multiple responses are used a description of the steps in the derivation of equation 5.14, the equation that defines the parameter confidence regions, will be presented. This will show the assumption that is made which was found to be not valid in the examples considered for small sample sizes.

As a starting point in the derivation of equation 5.14, the derivation of the determinant criterion will be presented as outlined in Box and Tiao (1973). Given the nonlinear model.

$$y_{ui} = f_i(x_u, \theta) + \epsilon_{ui} \quad , i = 1 \ldots r, u = 1 \ldots n \tag{5.16}$$

A Bayesian argument can be used to obtain the posterior distribution of (θ, Σ^{-1}), where noninformative priors are used for θ and Σ. The resulting posterior distribution is shown below.

$$p(\theta, \sigma^{ij}|y)d\theta \prod d\theta^{ij} \propto (2\pi)^{-\frac{1}{2}nr}|\Sigma^{-1}|^{\frac{1}{2}(n-r-1)} exp\left(-\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} v_{ij}\,\sigma^{ij}\right) d\theta \prod d\sigma^{ij} \tag{5.17}$$

To remove the Σ^{-1} term, the above equation is compared to the Wishart distribution. This simplifies the above equation to the following.

$$p(\theta|y) = C\ |V|^{-\frac{1}{2}n} = C\ |Z'Z|^{-\frac{1}{2}n} \tag{5.18}$$

The matrix V in the above equation corresponds to the $Z'Z$ matrix and C is a normalizing constant so that the sum of the probabilities is one. In deriving the above equation the assumption that Σ^{-1} is Wishart distributed is made. Therefore the joint confidence

region of the parameter estimates is given by.

$$ln[p(\hat{\theta}|y)] - ln[p(\theta|y)] = (|\hat{Z}'\hat{Z}| - |Z'Z|)/p \tag{5.19}$$

which is χ^2 distributed with p degrees of freedom. If the measurement error is estimated then the above becomes.

$$\frac{[|Z'Z| - |\hat{Z}'\hat{Z}|]/p}{s^2} \tag{5.20}$$

which is F distributed with p and $n-p$ degrees of freedom.

The reason why the level of the confidence region is not near the expected level when multiple responses are used, is because equation 5.20 is not F distributed for small sample sizes. Small here was found to be when n is less than or equal to $2p$, where p is the number of parameters estimated. Equation 5.20 is assumed to be F distributed because both the numerator and denominator are assumed to be χ^2 distributed. This assumption is made because the measurement error in the model is assumed to be normally distributed. Monte Carlo studies showed equation 5.20 to deviate from being F distributed as the sample size decreased. Though if only one response was used then equation 5.20 was found to be F distributed for all sample sizes. The Monte Carlo studies also showed that for large sample sizes, the numerator and denominator in equation 5.20 were not χ^2 distributed but their ratio was F distributed. Why this occurs and why equation 5.20 in not F distributed with a small sample size in not clear.

To illustrate the failure of the joint confidence region formula (i.e. equation 5.14), two examples will be presented. The first will consider the classic $A \rightarrow B \rightarrow C$ reaction, used by various authors such as Box and Draper (1965) for example. This example was used as it is similar to the type of models considered in this work but is much smaller, therefore allowing various Monte Carlo studies to be easily performed. The second example is based on case study two as outlined in Section 6.4. This example will show how the calculation of the joint confidence region fails with a large model.

A further consideration when generating confidence regions for parameter estimates obtained with the determinant criterion with multiple responses is the degrees of freedom

of the parameter estimates. In this case we have n observations per response, m responses and the number of parameters is p. Different values for the degrees of freedom have been proposed. Bard (1974) states that $nm - p$ should be used, while Bates and Watts make a case for $n - p$. Oxby (1997) addresses this issue and discusses the limitations of the values proposed by Bard and by Bates and Watts. He states that any simple expression will be an approximation and therefore its validity will be case dependent. Of the two possibilities he recommends the use of Bard's (1974) $nm - p$ value over Bates and Watts $n - p$ value, which he considers too conservative. In examining the failure of the approximate level true shape joint confidence region theory, both measure of degrees of freedom were considered and both showed similar results.

5.2.1 ABC Example

The ABC model has been used many times in the literature and represents the general sequence of first order irreversible chemical reaction as shown in equation 5.21, where A, B and C are chemical compounds and k_1 and k_2 are the reaction rate constants.

$$A \overset{k_1}{\Rightarrow} B \overset{k_2}{\Rightarrow} C \tag{5.21}$$

The A to B to C reaction can be described by the following set of differential equations.

$$\frac{d[A]}{dt} = -k_1[A]exp\frac{-E_1}{RT} \tag{5.22}$$

$$\frac{d[B]}{dt} = k_1[A]exp\frac{-E_1}{RT} - k_2[B]exp\frac{-E_2}{RT} \tag{5.23}$$

$$\frac{d[C]}{dt} = k_2[B]exp\frac{-E_2}{RT} \tag{5.24}$$

The ABC model was used in the simulation study since it is nonlinear, has multiple responses and is small enough so that various Monte Carlo studies can be easily run. A

further advantage of the ABC model is that an analytical solution exists to the differential equations, and this makes its use in this context much simpler.

To better understand the behaviour of the joint confidence region of the parameter estimates when multiple responses are used a number of Monte Carlo simulation studies were performed. These examined the effect of the number of trials and the number of responses used on the joint confidence regions of the parameters.

Each Monte Carlo study generated 10000 sample data sets. Each data set was generated by adding a normally distributed error with a mean of 0 and standard deviation of 0.05 to the true response values. Using this sample data the parameters were estimated using the determinant criterion. It is was then determined if the true parameter values occurred within the true shape joint confidence region of the parameter estimates, obtained using equation 5.20, at different levels of confidence. This procedure was repeated 10000 times in each Monte Carlo study and various Monte Carlo studies were run with different numbers of trials. For the simulation the parameter values of k_1 and k_2 and the experiment design, are the same as those used by Bates and Watts (1988), and are given below. The Bates and Watts design is an empirical design and the four trial design is obtained by removing two of the points.

- parameter values, $k_1 = 0.5$ and $k_2 = 0.2$
- sampling times for the 6 trial Bates and Watts experiment, 0.5, 1, 2, 4, 8 and 16 minutes
- sampling times for the 4 trial Bates and Watts experiment, 0.5, 2, 8 and 16 minutes

Table 5.1 shows how often the true parameter values were in the joint confidence region, expressed as a percentage, for the cases of four and six trials. The confidence regions were generated at three different confidence levels, 99, 95 and 50 percent. It is expected that with repeated sampling the true values of the parameter would fall inside the joint confidence regions at a percentage equal to the confidence level. From the results

number of trials	confidence level		
	99	95	50
4	28.0	25.4	11.2
6	92.3	81.6	32.2

Table 5.1: Percent of the time in 10000 trials, that the true parameter values of the ABC model were within the calculated confidence region of the parameter estimates, when the standard deviation of the measurement error was 0.05.

in the table we can see that the true confidence level of the joint confidence region in the four trial case is much smaller than the expected value. To determine if the observed results were due to the measurement error, the case study was repeated with different magnitudes of the measurement error and similar results were found. The only observed effect was that an increase in the error, increased the difference between the observed confidence level and the expected confidence level. In the above Monte Carlo studies the experiment design used to estimate the parameters was similar to the one used by Bates and Watts (1988). This design is not optimal. If a D-optimal experiment with two support points at 2.2 and 6.6 minutes is used, these points are replicated two and three times for the four and six trial experiments, and the Monte Carlo study is repeated using an optimal experiment design based on the true parameter values. The results obtained are similar to those obtained using the design from Bates and Watts (1988).

To gain an understanding of why the results in Table 5.1 were obtained, the distribution of the terms given by equations 6.25, 6.26 and 6.27 were generated for the four response and six response cases.

$$|\hat{Z}'\hat{Z}|/(n-p) \tag{5.25}$$

$$(|Z'Z| - |\hat{Z}'\hat{Z}|)/p \tag{5.26}$$

$$\frac{(|Z'Z| - |\hat{Z}'\hat{Z}|)/p}{|\hat{Z}'\hat{Z}|/(n-p)} \tag{5.27}$$

The first two terms are expected to be χ^2 distributed while the third is expected to be F distributed.

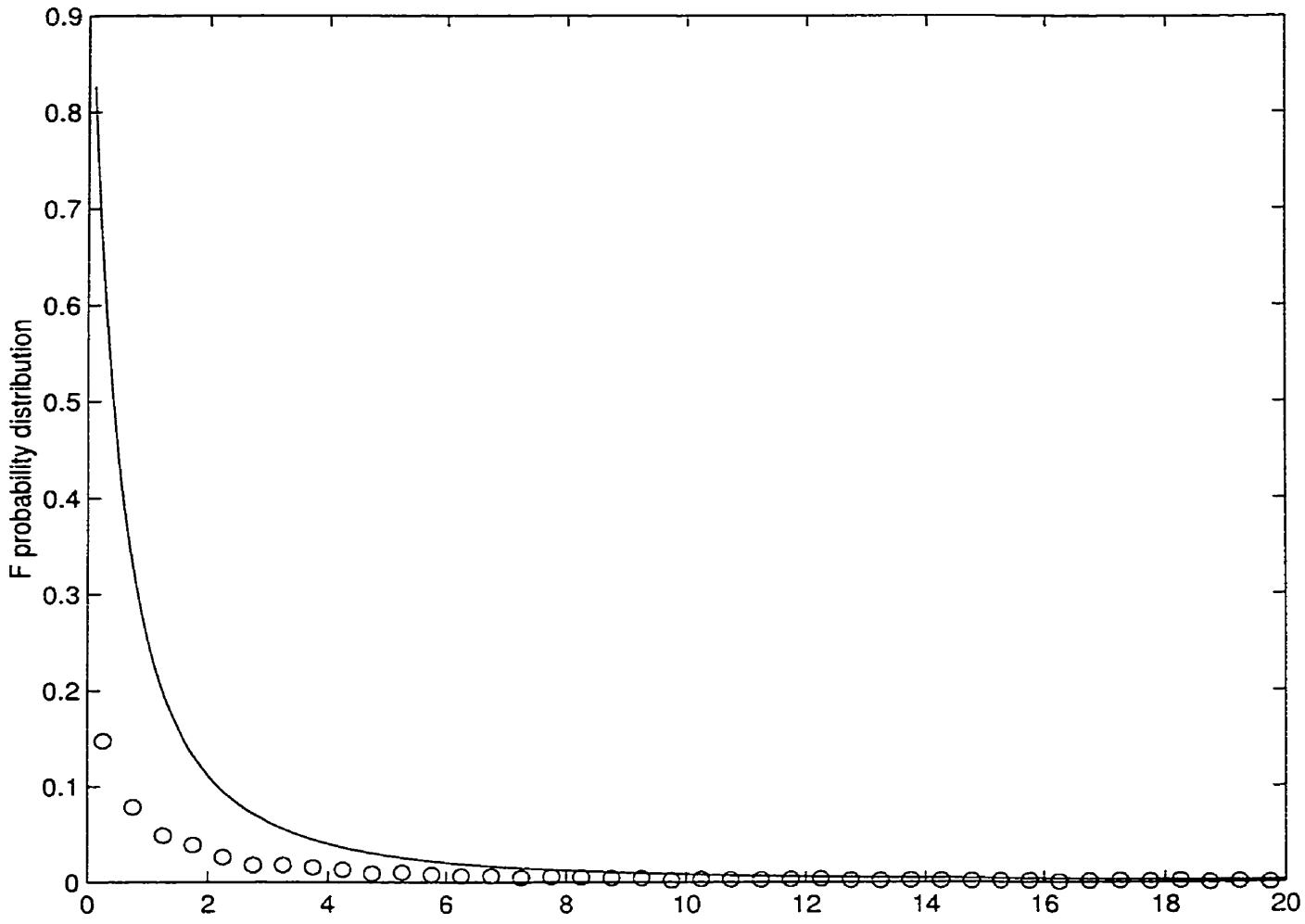


Figure 5.2: Distribution of the $\frac{(|Z'Z|-|\hat{Z}'\hat{Z}|)/p}{|Z'Z|/(n-p)}$ term for the four trial case of the ABC example (circles), and the F distribution with (2,2) degrees of freedom (solid line).

Figure 5.2 shows the distribution of equation 6.27 for the four trial case. The circles are the distribution that was obtained based on the Monte Carlo study and the solid line

is the F distribution with (2,2) degrees of freedom. The vertical axis is the probability density of the F distribution and the horizontal axis is the F value. The difference in the two distributions is what caused the discrepancy obtained in Table 5.1.

The above analysis was repeated for the six trial case. It was found that the ratio of the terms $|Z'Z|/(n-p)$ to $(|Z'Z|-|\hat{Z}'\hat{Z}|)/n$ is closer to being F distributed. Therefore the observed confidence level is closer to the expected confidence level. When the distribution of the individual terms was examined, it was found that they were not χ^2 distributed. Why the ratio is then F distributed is unclear.

5.2.2 Case Study 2

This example is based on case study two as described in section 6.4. It involved the estimation of five parameters using five responses in the Watpoly model. The system that was simulated is the copolymerization of Styrene and Methyl Methacrylate.

This example will show that the failure of the joint confidence region theory also occurred in the large model used in this thesis. Due to the size of the model, only 177 cases were run in the Monte Carlo study. This study as in the previous example involved generating sample data sets and obtaining parameter estimates from them. Then it was determined if the true parameter values were within the true shape joint confidence region obtained using equation 5.20. The results obtained are shown in Table 5.2, and are similar to those obtained in the ABC example. As in the ABC example, the level of the joint confidence region is significantly different than the expected level.

	confidence level		
expected	99	95	50
observed	72	60	31

Table 5.2: Percent of the time in 177 trials, that the true parameter values in Case Study 2 were within the calculated confidence region of the parameter estimates.

5.2.3 Discussion

The failure of the joint confidence region theory when the number of trials is small and multiple responses are used has to be taken into account when parameter estimation is being carried out. When confidence regions are used as a measure of the uncertainty in the parameter estimates, the situation can occur that a greater amount of confidence may be associated with the parameter estimates than is appropriate. This can detrimentally affect any later actions that depend on these results such as model discrimination or the confidence region of the model predictions.

From the Monte Carlo studies in both examples it was found that using multiple responses will only affect the confidence region. The parameter estimates are not affected, and as expected better parameter estimates are obtained if multiple responses are used. Therefore it is still recommended to use multiple responses for parameter estimation, but caution should be exercised in the interpretation of the joint confidence regions obtained.

5.3 Contouring algorithm

In the previous section a number of different methods to obtain the joint confidence region of a set of parameter estimates were discussed. It was found that the true shape confidence regions approach was the best choice due to the nonlinearity that is present in the type of models considered.

This section will discuss the practical aspect of how to obtain the joint confidence regions. Two possible methods will be presented, the first is based on the algorithm proposed by Dhib and Oxby (1998) which will produce a very accurate joint confidence region. The second was developed by the author and is an adaptation of the simplex optimization method.

The true shape joint confidence region is obtained using equation 5.10, which is re-

produced below for the readers convenience.

$$S(\theta) - S(\hat{\theta}) \leq p\, s^2\, F(p, n-p, \alpha) \tag{5.28}$$

The above equation when used is usually rearranged into the form shown below with the right hand side being constant.

$$S(\theta) = S(\hat{\theta}) + ps^2 F(p, n-p, \alpha) \tag{5.29}$$

If the single response case is used, then $S(\hat{\theta})$ is equal to the sum of the squared deviations, while for the multiresponse case $S(\hat{\theta})$ is equal to $|Z'Z|$. In using equation 5.29 to generate a two parameter joint confidence region, the parameter values that represent the contour are obtained by fixing the first parameter and solving for the second. The first parameter is then incremented and the procedure is repeated. This method can become numerically unstable if a region of the contour is reached where a small change in the first parameter results in a large change in the second parameter. Normally when this region is reached the second parameter is held constant while the first is solved for. While this is a possible solution, it has the limitations that operator intervention is required and the process can be slow due to the optimization that is required to obtain each point on the contour. This problem is compounded if the objective function surface is convoluted and the function evaluations are very expensive.

The method proposed by Dhib and Oxby (1998) is an adaptive algorithm that follows the contour. This is achieved by taking a step from the current point and searching for the contour along a path that is perpendicular to the step taken. Their algorithm works well since it will adapt its step size based on the curvature of the contour and use a search direction that is perpendicular to the arc created by the last three points along the contour. Its limitation is that it can be computationally expensive if a large number of contour searches are required (i.e. if a large number of points are required to define the contour, such as in an area of high curvature). The author has also found that it can be

numerically unstable in certain case studies performed. This was due to the significant amount of convolution present in the objective function surface.

The method proposed by the author will follow the contour thus minimizing the number of function evaluations required. In general it will usually require less function evaluations than the method proposed by Oxby and Dhib. The only limitation is that the resolution of the confidence region and the limits on the parameter values have to be selected a priori.

The algorithm is similar to the simplex optimization algorithm in that a shape is flipped in the required direction to achieve the objective (minimization in the simplex and following the contour in the contouring algorithm). The algorithm will start at the point estimate values, move horizontally until it finds the contour, and then move a box along the contour until all of the contour is found. The algorithm was used to generate all of the confidence regions in the case studies given in Chapter 6.

The steps of the contouring algorithm are as follows:

1. define the parameter ranges and the desired resolution of the contour
2. create a matrix to store a grid based on the parameter ranges, where the step size in the grid is equal to the contour resolution
3. starting at the grid square that contains the point estimate, move in a horizontal direction until the contour is crossed
4. calculate the points of the grid-square and store in the grid matrix
5. determine the direction in which to flip the square
6. if we are at the starting grid-square stop, else go to step 4
7. based on the data stored in the grid matrix generate the contour, (the `contour.m` function in Matlab can be used to extract the contour from the grid matrix)

The contouring algorithm was written as a Matlab function, and the Matlab source code is provided in Appendix B. Documentation in the sourcecode will describe the details of the algorithm such as, how it chooses direction in which to flip the square, or what to do if the grid boundary is encountered.

Chapter 6

Case Studies

6.1 Introduction

This chapter will give a description of the polymerization model used and discuss three case studies. The case studies presented will highlight the four principal steps of the parameter estimation process and the different problems that may be encountered when estimating parameters within large models. While each of the four parameter estimation steps will be described in each case study, each case study will focus on one or two of the steps. In the case studies there are a number of common aspects to all, such as the generation of sample data and optimization algorithm used. These will be discussed below.

The polymerization model used in the case studies has been extensively tested with experimental data, and has been found to perform well. Therefore the simulated data obtained from it is very representative of the data that would be obtained from a laboratory experiment. In the case studies, the data was simulated by adding an error to the response values given by the model. This error was normally distributed with a mean of zero and a response measurement error standard deviation as specified in each of the case studies. The magnitude of the measurement error for each response was determined

from experimental studies reported in the literature (Dube and Penlidis, 1996; D'Agnillo et al. 1998), and from discussions with other experienced researchers (McManus, 1999).

The sensitivity analysis in each of the case studies was carried out by an analysis of the gradient values as outlined in Chapter 2. In the interpretation of the gradient plots, only a sample of the plots generated will be presented in each case study, and for each case study the sensitivity analysis is summarized in a table in the sensitivity analysis subsection. The plots that are not shown in the case studies sections are included in Appendix C.

The D-optimality criterion and the `fmins` algorithm, as discussed in Chapter 3, were used to design the experiments in all of the case studies. The parameter estimates were obtained using the Determinant criterion and the `fmins` optimization algorithm as discussed in Chapter 4.

Joint confidence regions of the parameter estimates were obtained as outlined in Chapter 5. If p parameters are estimated, the joint confidence region is a p dimensional space. Since it is not possible to view this space if p is greater than three, two parameter joint confidence regions are generated in the case studies. A limitation of these regions is that they are conditional confidence regions and thus are a function of the other parameter estimates, where a change in the other parameter estimates may result in a change in the conditional confidence region.

A general description of the model used in the case studies is given in the next section. This is followed by the three case studies and a summary of the observations from the case studies.

6.2 Model Description

The model used in the case studies is the Watpoly polymerization model, (Gao and Penlidis, 1996 and 1998). This is a mechanistic model for the simulation of free radical homopolymerization and copolymerization, that has been tested extensively by compar-

ison with experimental studies. As an introduction to the model an overview of polymerization reaction kinetics is presented and is followed by a general discussion of the model. This includes a listing of the input variables and responses, and a description of a selection of equations from the model. This description will illustrate how the estimated parameters and responses used in the case studies are integrated within the model and how they are related to each other.

6.2.1 Polymerization Reaction Kinetics

In free radical polymerization there are three principal steps; initiation, propagation and termination. In the following description of these steps the assumption is made that only the terminal monomer unit on the growing polymer chain affects the reaction rate. The polymerization model based on this assumption is referred to as the 'terminal' model. In the description that follows of the three polymerization steps, the notation $R_{n,i}\bullet$ indicates a radical chain n monomer units long and ending with monomer type i. The principal steps in polymerization kinetics are:

Initiation; decomposition of the initiator I into two primary radicals $R_{in}\bullet$, that further react with the monomer to form a radical of chain length one, $R_{1,i}\bullet$, where i is the type of monomer.

$$I \xrightarrow{k_d} 2R_{in}\bullet \tag{6.1}$$

$$R_{in}\bullet + M_i \xrightarrow{k_{pi}} R_{1,i}\bullet \tag{6.2}$$

Propagation; growth of the radical chain by the addition of monomer units.

$$R_{n,i}\bullet + M_j \xrightarrow{k_{p,ij}} R_{n+1,j}\bullet \tag{6.3}$$

Termination; the two radical chains form either one polymer chain of their combined length or two polymer chains of the respective radical chain lengths.

$$R_{m,i}\bullet + R_{n,k}\bullet \xrightarrow{k_{t,ik}} P_{m+n} \quad or \quad P_m + P_n \tag{6.4}$$

Termination can also occur by transfer to monomer as shown below.

$$R_{m,i}\bullet + M_j \overset{k_{fm_{ij}}}{\rightarrow} P_m + R_j\bullet \tag{6.5}$$

6.2.2 Model Equations

The following is a listing of the relationships that are part of the Watpoly model (Gao 1999). Only a subset of the relationships and equations are presented here, as the objective is to provide the reader with an overview of the model and the parameters estimated in the case studies.

Initiation

Rate of initiation

$$R_I = 2fk_d[I] \tag{6.6}$$

where $[I]$ is the initiator concentration, k_d is the rate constant of initiator decomposition and f is the initiator efficiency, which is a function of temperature and is expressed using an Arrhenius relationship:

$$f = f_{pre-exp}e^{[-f_{act-eng}/RT]} \tag{6.7}$$

where R is the gas constant and T is the reaction temperature. The parameters $f_{pre-exp}$ and $f_{act-eng}$ will be estimated in Case Study 1.

When the radical steady-state hypothesis is used, the total radical concentration, $[R\bullet]$, is given by,

$$[R\bullet] = \left(\frac{2fk_dN_I}{k_tV}\right)^{1/2} \tag{6.8}$$

Propagation

Rate of copolymerization

$$R_p = k_p[M][R\bullet] \tag{6.9}$$

where $[M]$ is the total concentration of monomers and $[R\bullet]$ is the concentration of radicals. The pseudo-k_p in the above equation is given by,

$$k_p = k_{p11}\phi_1 \bullet f_1 + k_{p12}\phi_1 \bullet f_2 + k_{p21}\phi_2 \bullet f_1 + k_{p22}\phi_2 \bullet f_2 \tag{6.10}$$

where k_{p11} and k_{p22} are the homopolymerization rate constants, k_{p12} and k_{p21} are the cross propagation rate constants, f_1 and f_2 are the mole fractions in the reacting mixture of monomer 1 and 2, respectively, and $\phi_1\bullet$ and $\phi_2\bullet$ are the mole fractions of radical type 1 and 2, respectively, given by,

$$\phi_1\bullet = \frac{k_{p21}f_1}{k_{p21}f_1 + k_{p12}f_2} \tag{6.11}$$

$$\phi_2\bullet = \frac{k_{p12}f_2}{k_{p21}f_1 + k_{p12}f_2} \tag{6.12}$$

If the total free volume is less than the critical free volume for propagation the $k_{p_{ij}}$ values are adjusted using the following relationship,

$$k_{p_{ij}} = k_{p_{ij}} e^{\left[-BBm_i\left(\frac{1}{V_f} - \frac{1}{V_{f_{crit}}}\right)\right]} \tag{6.13}$$

where BBm_i is the rate of decrease of k_p with free volume for monomer i, V_f is the total free volume and $V_{f_{crit}}$ is the critical free volume for propagation. The parameter BBm_i will be estimated in Case Studies 2 and 3.

Termination

Rate of termination

$$R_t = k_{t11}[R_1\bullet]^2 + 2k_{t12}[R_1\bullet][R_2\bullet] + k_{t22}[R_2\bullet]^2 \tag{6.14}$$

where k_{t11} and k_{t22} are the homo-termination rate constants, k_{t12} is the cross-termination rate constant and $[R_i\bullet]$ is the concentration of radical type i.

Transfer to monomer

The pseudo-rate constant for chain transfer to monomer k_{fm} is given by,

$$k_{fm} = k_{fm_{11}}\phi_1 \bullet f_1 + k_{fm_{12}}\phi_1 \bullet f_2 + k_{fm_{21}}\phi_2 \bullet f_1 + k_{fm_{22}}\phi_2 \bullet f_2 \tag{6.15}$$

where $k_{fm_{11}}$ and $k_{fm_{22}}$ are the rate constants for chain transfer to monomer 1 and 2, respectively, and $k_{fm_{12}}$ and $k_{fm_{21}}$ are the rate constants for cross-chain transfer to monomer. The $k_{fm_{ij}}$ parameters are a function of temperature and determined using an Arrhenius relationship.

$$k_{fm_{ij}} = k_{fm_{ij}\,pre-exp}e^{[-k_{fm_{ij}\,act-eng}/RT]} \tag{6.16}$$

The parameters $k_{fm_{ij}\,pre-exp}$ and $k_{fm_{ij}\,act-eng}$ will be estimated in Case Study 1.

Free volume

The monomer free volume is given by,

$$V_{Fm_i} = V_{fom_i} + \alpha_{m_i}(T - T_{gm_i})(V_{m_i}/V_T) \tag{6.17}$$

where V_{Fm_i} is the free volume of monomer i, V_{m_i} is the total volume of monomer i in the reaction mixture, V_T is the total volume of the reaction mixture, α_{m_i} is the variation of free volume with temperature for monomer i, T is the reaction temperature, Tgm_i is the glass transition temperature of monomer i and V_{fom_i} is the constant of free volume with temperature for monomer i. The parameter α_{m_i} will be estimated in Case Studies 1, 2 and 3.

The polymer free volume is given by,

$$V_{Fp} = V_{fop} + \alpha_p(T - T_{gp})(V_p/V_T) \tag{6.18}$$

where V_{Fp} is the polymer free volume, V_p is the total volume of polymer in the reaction mixture, V_T is the total volume of the reaction mixture, α_p is the variation

of free volume with temperature for polymer, T is the reaction temperature, Tgp is the glass transition temperature of polymer and V_{fop} is the constant of free volume with temperature for polymer and will be estimated in Case Study 1.

Tgp is obtained from the following relationship,

$$\frac{1}{Tgp} = \frac{W_{M_1} P_{M_1 M_1}}{Tg_{M_1}} + \frac{W_{M_2} P_{M_2 M_2}}{Tg_{M_2}} + \frac{W_{M_1} P_{M_1 M_2} + W_{M_2} P_{M_2 M_1}}{Tg_{M_1 M_2}} \tag{6.19}$$

where $P_{M_1 M_1}$, $P_{M_2 M_2}$ and $P_{M_1 M_2}$ are the probabilities of the corresponding sequences of monomer 1 (M_1) and monomer 2 (M_2), W_{M_1} and W_{M_2} are the weight fractions of monomer 1 and 2, and $Tg_{M_1 M_2}$ is the glass transition temperature of the alternating copolymer. The $Tg_{M_1 M_2}$ parameter is estimated in Case Study 3 and the Tg_{M_1} parameter which is equal to the Tgp parameter in a homopolymerization is estimated in Case Study 2.

Composition

The instantaneous copolymer composition is given by,

$$F_1 = \frac{r_1 f_1^2 + f_1 f_2}{r_1 f_1^2 + 2 f_1 f_2 + r_2 f_2^2} \tag{6.20}$$

where r_1 and r_2 are the reactivity ratios and f_1 and f_2 are the mole fractions of free monomer 1 and monomer 2.

The mole fractions are a function of the moles of monomer, which is a function of the rate of propagation, the rate of termination and the radical concentration.

Triad fractions

The instantaneous triad fractions are given by,

$$A_{111} = \frac{r_1^2 f_1^2}{r_1^2 f_1^2 + 2 r_1 f_1 f_2 + f_2^2} \tag{6.21}$$

$$A_{211} = A_{112} = \frac{r_1^2 f_1 f_2}{r_1^2 f_1^2 + 2 r_1 f_1 f_2 + f_2^2} \tag{6.22}$$

$$A_{212} = \frac{f_2^2}{r_1^2 f_1^2 + 2r_1 f_1 f_2 + f_2^2} \tag{6.23}$$

Molecular weight distribution

The moments of the molecular weight distribution are given by,

$$\frac{1}{V}\frac{dVQ_0}{dt} = \frac{1}{V}\frac{dN_p}{dt}\left(\tau + \frac{\beta}{2}\right) \tag{6.24}$$

$$\frac{1}{V}\frac{dVQ_1}{dt} = \frac{1}{V}\frac{dN_p}{dt} \tag{6.25}$$

$$\frac{1}{V}\frac{dVQ_2}{dt} = \frac{1}{V}\frac{dN_p}{dt}\left(\frac{2\tau + 3\beta}{(\tau + \beta)^2}\right) \tag{6.26}$$

where τ and β are given by,

$$\tau = \frac{(\gamma k_t [R\bullet]^2 + k_{fM}[M][R\bullet])}{\frac{1}{V}\frac{dN_p}{dt}} \tag{6.27}$$

$$\beta = \frac{(1-\gamma) k_t [R\bullet]^2}{\frac{1}{V}\frac{dN_p}{dt}} \tag{6.28}$$

In using the Watpoly model only part of its capabilities were used. All of the simulations studies were carried out in batch reactor mode at isothermal conditions. Of the available responses from the Watpoly model, the following responses were used in the case studies.

- conversion
- polymer composition
- radical concentration
- number average molecular weight
- weight average molecular weight
- polymerization rate
- triad fractions

6.3 Case Study 1

6.3.1 Description

This case study will describe the estimation of five parameters within the Watpoly model for the homopolymerization of Styrene using five responses. A listing of the estimated parameters is given in Table 6.1. The responses used and their respective measurement error standard deviations (σ_{resp}) are given in Table 6.2.

6.3.2 Sensitivity Analysis

The sensitivity analysis was performed with respect to the five parameters listed in Table 6.1, and is summarized in Table 6.3. In this table the comments column is a summary of the observability of each parameter, where any areas of large or small gradient values (i.e. good or poor observability) are identified. As an example, at the top of Table 6.3, for the parameter $k_{fm\ act-eng}$ and the response of conversion the comments column states '65 °C , max. at 820 min., 75 °C , max. at 450 min., better observability at 65 °C '. This area corresponds to the area of large gradient values as shown in Figures 6.1 and 6.2.

To illustrate the sensitivity analysis procedure, the analysis with respect to the parameter $k_{fm\ act-eng}$ will be described. Figures 6.1 and 6.2 show the gradient values for all of the responses with respect to $k_{fm\ act-eng}$ at 65 °C and 75 °C respectively, plotted versus time in minutes. These temperatures were chosen, as they represent a range of typical operating temperatures. In these plots the horizontal axis is time and the vertical axis is the normalized gradient value. The gradient values were normalized by dividing by the standard deviation of the response measurement error as given in Table 6.2 and discussed in Chapter 2.

Good parameter observability is determined by locating areas where the absolute normalized gradient values are large. Figure 6.1 shows the gradient plots for $k_{fm\ act-eng}$ at 65 °C . The gradient values have been plotted on three separate graphs to allow for

Parameter	Description
$k_{fm\ act-eng}$	activation energy in Arrhenius expression for the rate of radical transfer to monomer (cal/mol)
$k_{fm\ pre-exp}$	pre-exponential factor in Arrhenius expression for the rate of radical transfer to monomer (L/mol min)
$f_{act-eng}$	activation energy in Arrhenius expression for initiator efficiency (cal/mol)
$f_{pre-exp}$	pre-exponential factor in Arrhenius expression for initiator efficiency (-)
V_{fop}	constant of free volume with temperature for polymer (free volume units/K)

Table 6.1: Parameters estimated in case study 1

Response	Measurement error standard deviation (σ_{resp})
conversion	0.025
M_n	10000
M_w	10000
rate	0.001
radical conc.	25%

Table 6.2: Measurement error standard deviation (σ_{resp}) of the responses used in case study 1

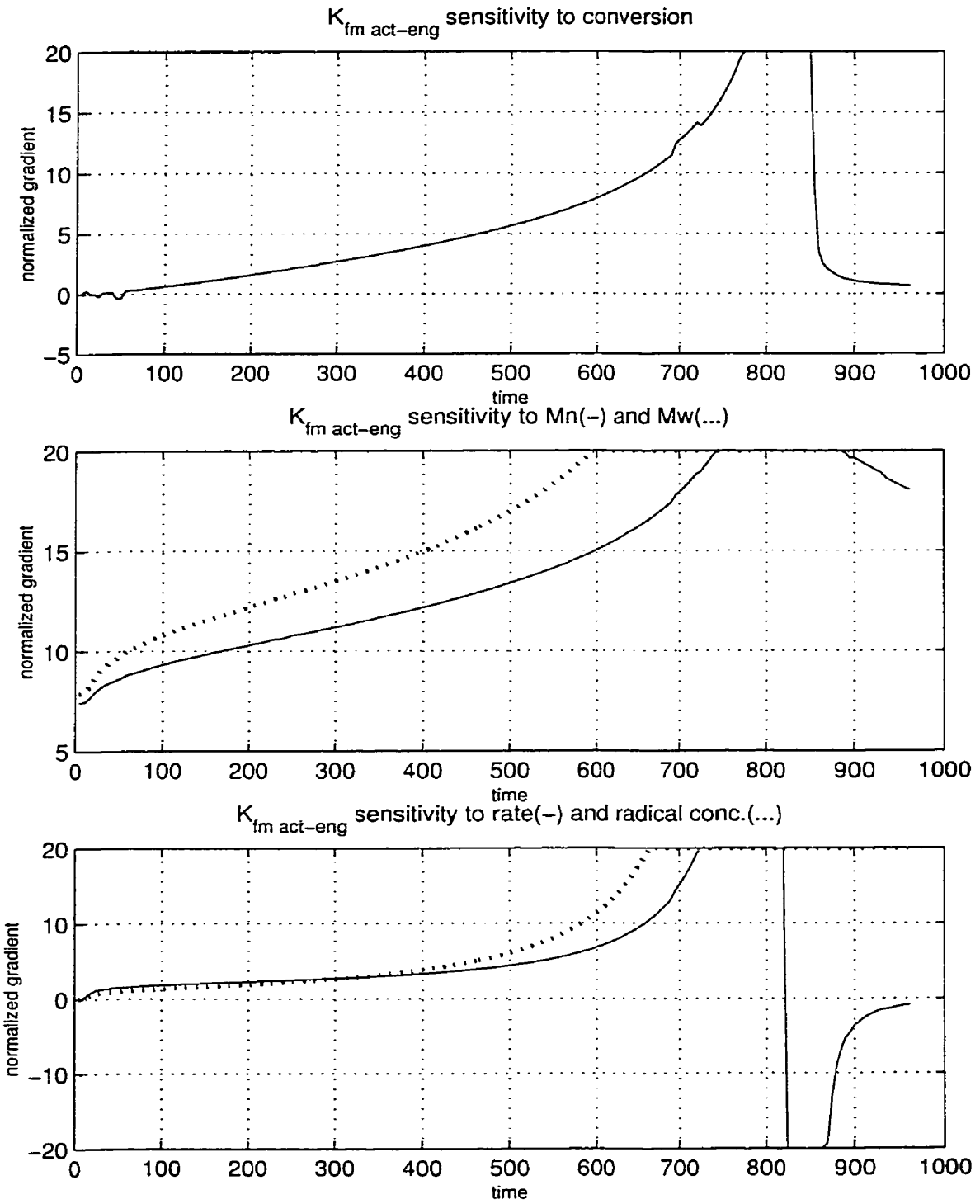


Figure 6.1: Gradient plots with respect to the parameter $k_{fm\ act-eng}$ at 65 °C .

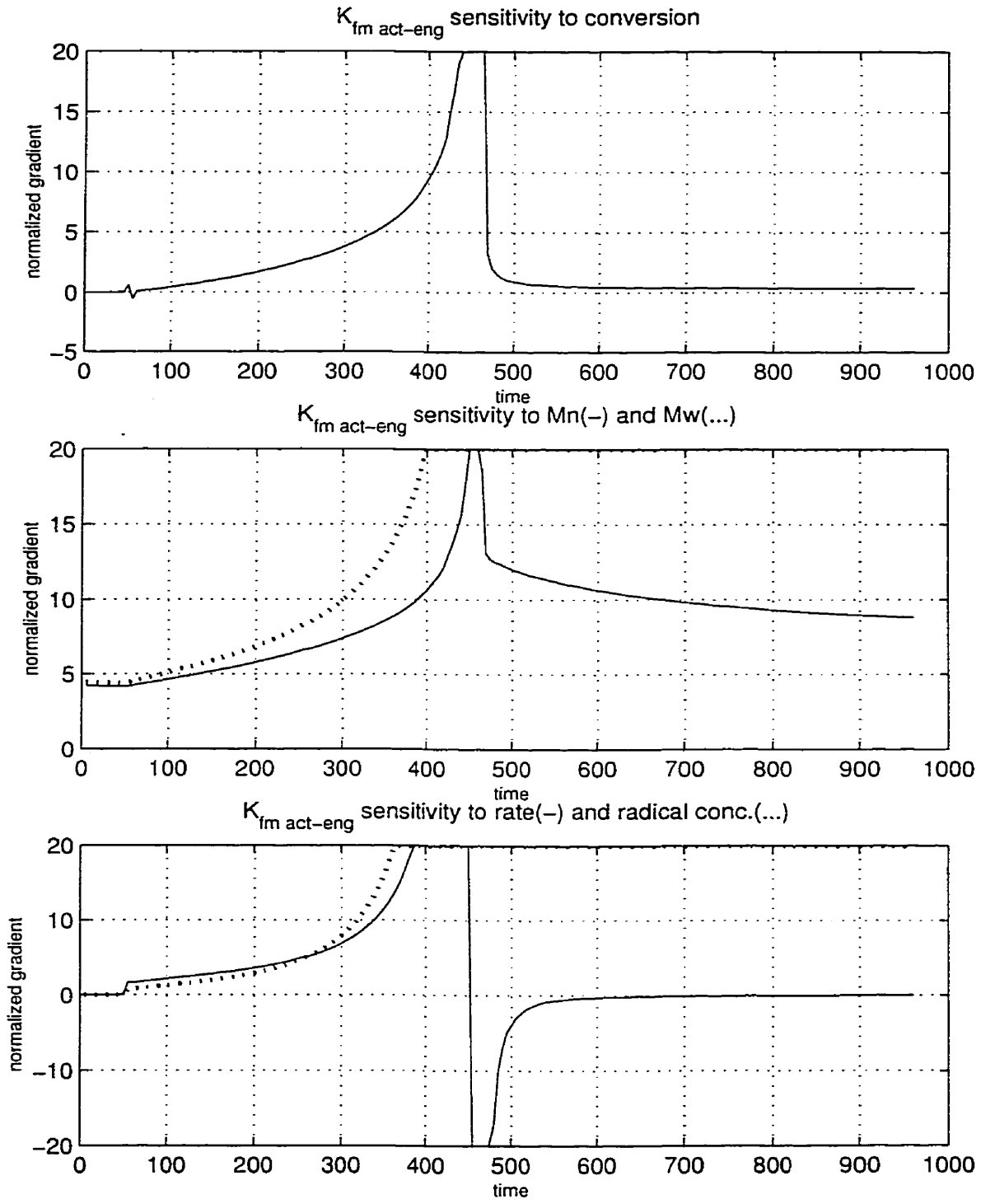


Figure 6.2: Gradient plots with respect to the parameter $k_{fm\ act-eng}$ at 75 °C .

easier interpretation. The top graph contains the response of conversion, the middle graph contains the responses of number average molecular weight (M_n) and weight average molecular weight (M_w), and the bottom graph contains the responses of rate and radical concentration. In all of the graphs the normalized gradient values have been clipped at a value of 20. This is performed to maintain the resolution of the graphs and to aid interpretability. An examination of all the graphs indicates the presence of a peak of varying width near 800 minutes. Therefore all five responses will provide a good level of information about the parameter at that point. A comparison of all of the responses indicates that the M_n and M_w are the best responses to use. This is due to the broadness of the peaks. By broadness it is meant that the peak width is significant, so that large gradient values occur over a wide range of sampling times. All of the responses maximize the amount of information they contain in the area when time is equal to 800 minutes. This time corresponds to the auto-accelaration phase of the polymerization. This can be seen by examining the plot of conversion versus time shown in Figure 6.3. In this plot the vertical axis is conversion, where 1.0 indicates 100% conversion, the horizontal axis is time and the solid and dashed lines represent the conversion profiles at 65 °C and 75 °C , respectively. The auto-acceleration section is the steep section of the conversion versus time curve (i.e. from 420 to 480 minutes for the 75 °C curve).

An analysis of the gradient plots at 75 °C shown in Figure 6.2 indicates similar results as at 65 °C . The differences are that the gradient peaks have moved to 450 minutes, which corresponds to the auto-accelaration section at 75 °C as shown in Figure 6.3. A second difference is that the magnitude of the observability has decreased, due to the smaller peak size.

The sensitivity analysis was performed with the assumption that the initial values of the parameters were adequate. This assumption may not be valid and the second part of the sensitivity analysis is to test it by determining the effect a change in the parameter values will have on the gradient plots. Figure 6.4 for example shows the effect

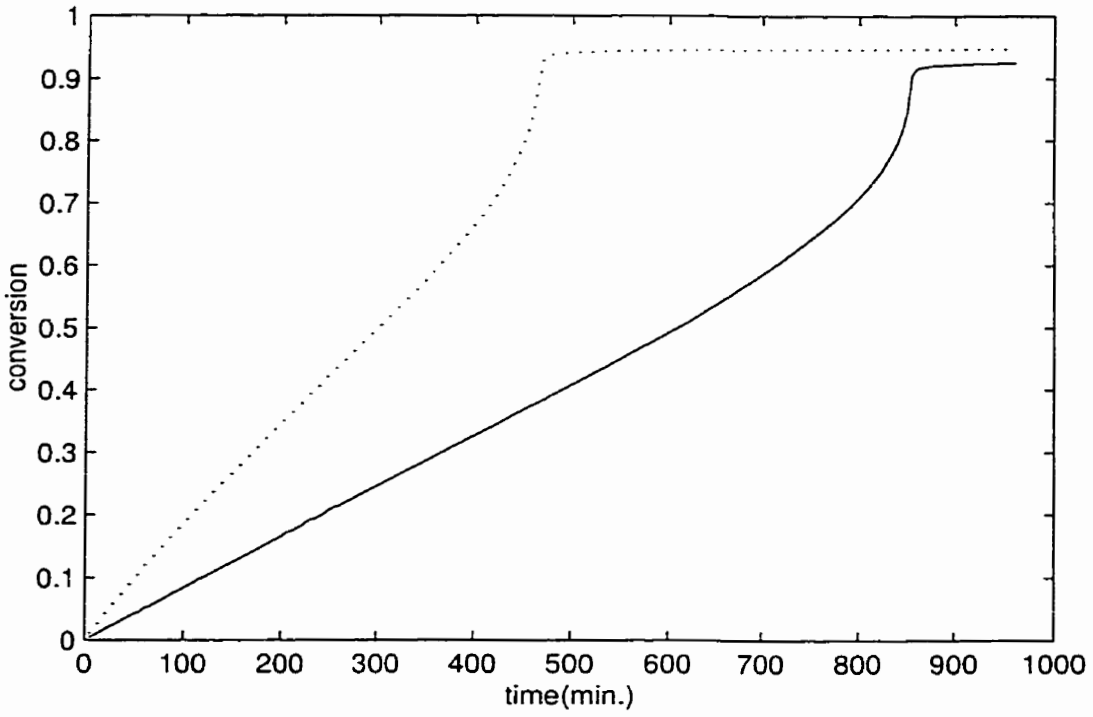


Figure 6.3: Plot of conversion versus time for the homopolymerization of Styrene, solid line 65 °C , dashed line 75 °C .

of changing the value of $k_{fm\ act-eng}$ on the gradient of the conversion response at 65 °C . The axes are similar to the axes of the gradient plots describes earlier. The three gradient curves that are shown were calculated using three different values of $k_{fm\ act-eng}$ (12000, 13426 and 15000, labelled A, B and C, respectively). These values represent the possible range of $k_{fm\ act-eng}$, and were chosen based on an examination of the value of $k_{fm\ act-eng}$ for other similar systems. An inspection of the graph reveals that the observability decreases significantly as the value of the parameter increases. Although the general area of maximum observability (i.e. where the absolute gradient value is largest) does not move, it becomes smaller. Therefore a measurement taken at 450 minutes will provide less information about the parameter as the parameter value increases, but will be the best location to take samples nonetheless.

Figure 6.5 shows the affect of changing the value of V_{fo} on the gradient of the conversion response at 65 °C . The three curves that are shown were calculated using three different values of V_{fo} (0.020, 0.025 and 0.030, labelled A, B and C, respectively). These values represent the possible range of V_{fo} and were chosen based on an examination of the value of V_{fo} for other similar systems. An examination of the graph reveals that as the value of the parameter changes the peak size and its location change. Good observability is available with all three peaks due to their large size, but the location of maximum observability moves as the parameter values change. This indicates that the designed experiment will change as different parameter values are used. Since the area of maximum observability occurs in a very sharp peak, this will result in a rapid decrease in the amount of information provided by the conversion response if the sampling time deviates by a small amount from the optimal value. This deviation from the optimal sampling time can occur if the parameter values used to design the experiment are different from the estimated parameter values. Therefore the designed experiment will be very sensitive to the values of V_{fo} used (i.e. the experiment will not be robust to the V_{fo} parameter uncertainty).

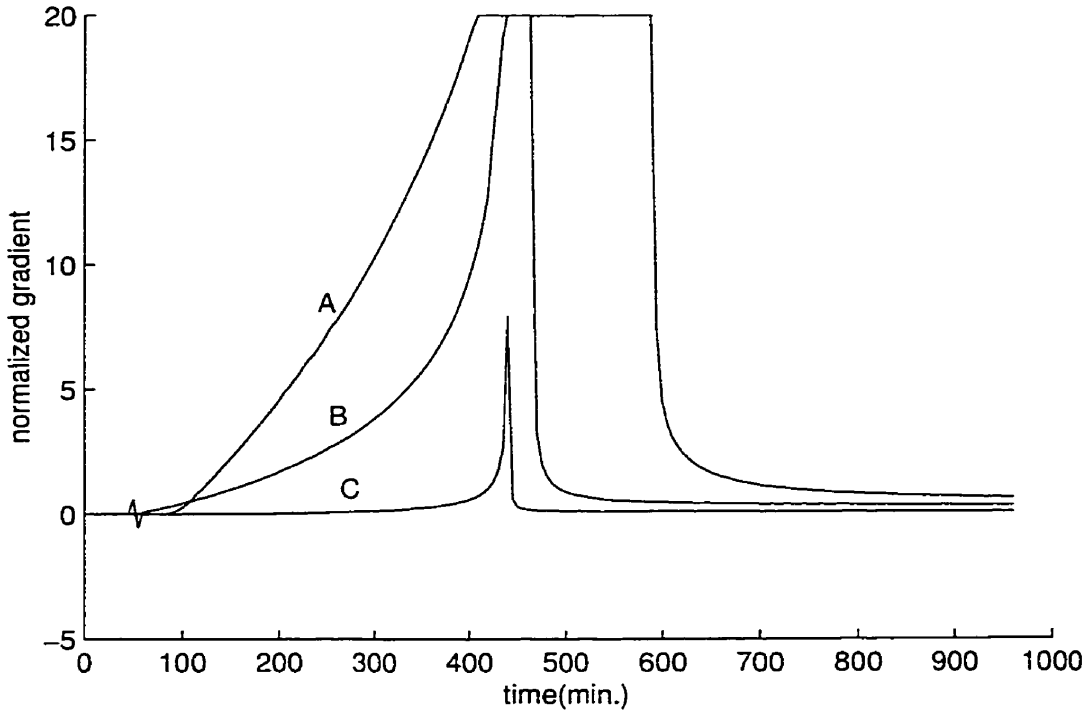


Figure 6.4: Effect of changing the value of $k_{fm\ act-eng}$ on the gradient plot of conversion at 65 °C , where $k_{fm\ act-eng}$ equals 12000(A), 13426(B) and 15000(C)

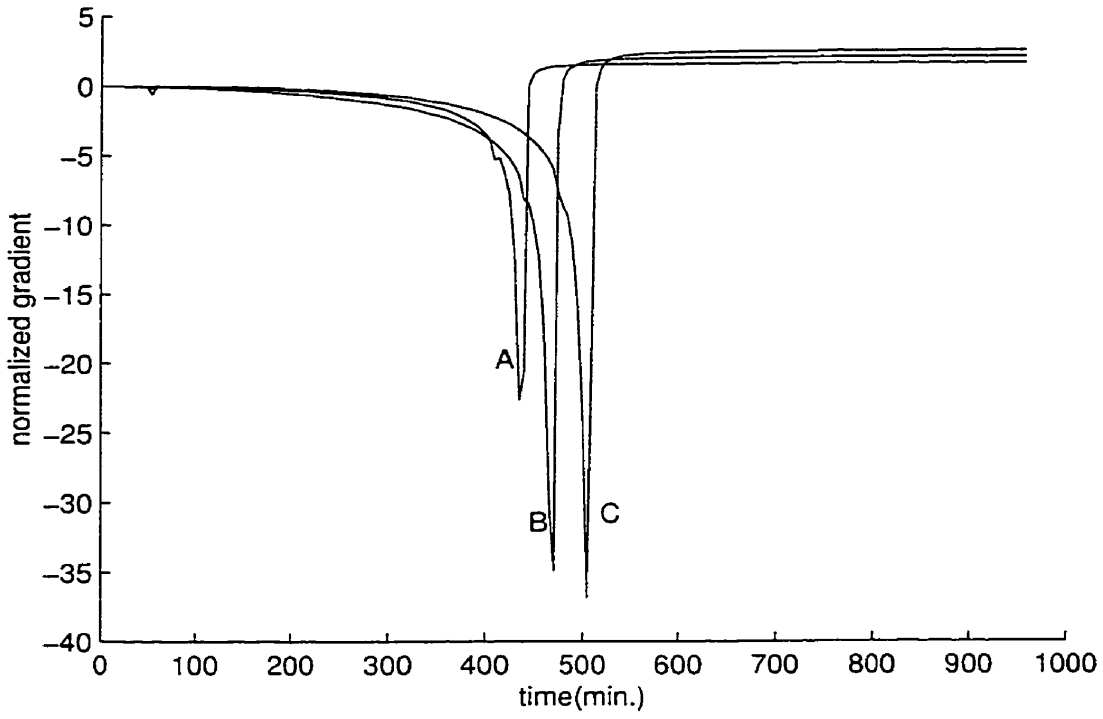


Figure 6.5: Effect of changing the value of V_{fo} on the gradient plot of conversion at 65 °C , where V_{fo} equals 0.020(A), 0.025(B) and 0.030(C)

Parameter	Response	Comments
$K_{fm}act - eng$	conversion	65 °C , max. at 820 min., 75 °C , max. at 450 min., better observability at 65 °C
	M_n	65 °C , max. at 820 min., 75 °C , max. at 450 min., very good observability at both temperatures, better at 65 °C
	M_w	65 °C , max. at 820 min., 75 °C , max. at 450 min., very good observability at both temperatures, better at 65 °C
	rate	65 °C , max. at 800 min., 75 °C , max. at 450 min.
	radical conc.	65 °C , max. at 800 min., 75 °C , max. at 450 min.
$k_{fm\ pre-exp}$	conversion	65 °C , max. at 850 min., 75 °C , max. at 460 min., sharp peak at both temperatures
	M_n	65 °C , max. at 850 min., 75 °C , max. at 460 min., sharp peak at both temperatures
	M_w	65 °C , max. at 850 min., 75 °C , max. at 460 min., sharp peak at both temperatures
	rate	65 °C , max. at 850 min., 75 °C , max. at 460 min., sharp peak at both temperatures
	radical conc.	65 °C , max. at 850 min., 75 °C , max. at 460 min., sharp peak at both temperatures
$fact - eng$	conversion	65 °C , max. at 850 min., 75 °C , max. at 460 min., sharp peak at both temperatures and very poor observability
	M_n	65 °C , max. at 850 min., 75 °C , max. at 460 min., sharp peak at both temperatures and very poor observability
	M_w	65 °C , max. at 850 min., 75 °C , max. at 460 min., sharp peak at both temperatures and very poor observability
	rate	65 °C , max. at 850 min., 75 °C , max. at 460 min., sharp peak at both temperatures and very poor observability
	radical conc.	65 °C , max. at 850 min., 75 °C , max. at 460 min., sharp peak at both temperatures and very poor observability
$fpre - exp$	conversion	65 °C , max. at 850 min., 75 °C , max. at 460 min.
	M_n	65 °C , max. at 850 min., 75 °C , max. at 460 min.
	M_w	65 °C , max. at 850 min., 75 °C , max. at 460 min.
	rate	65 °C , max. at 850 min., 75 °C , max. at 460 min.
	radical conc.	65 °C , max. at 850 min., 75 °C , max. at 460 min.
V_{fo}	conversion	65 °C , max. at 850 min., 75 °C , max. at 460 min., poor observability at low conversion
	M_n	65 °C , max. at 850 min., 75 °C , max. at 460 min., poor observability at low conversion
	M_w	65 °C , max. at 850 min., 75 °C , max. at 460 min., poor observability at low conversion
	rate	65 °C , max. at 850 min., 75 °C , max. at 460 min., poor observability at low conversion
	radical conc.	65 °C , max. at 850 min., 75 °C , max. at 460 min., poor observability at low conversion

Table 6.3: Summary of the sensitivity analysis

6.3.3 Experiment Design

The objective of experiment design was to find an experiment that would maximize the amount of information provided about the parameter values. The experiment was designed using the D-optimality criterion as discussed in Chapter 3. Since a subset of the parameters to be estimated are within an Arrhenius expression, two different temperatures were used in the experiment. The size of the experiment was chosen to be 10 trials, with 5 trials at 65 °C and 75 °C , respectively. The value of 10 was chosen, as it is twice the number of parameters.

In designing the optimal experiment it was found that a large number of local optima existed, as a different experiment was obtained with each different initial starting point used. To deal with the local optima three different approaches were tried. The first was based on information provided by the sensitivity analysis, the second was to randomly choose 10 points within the time range of 100 to 900 minutes, and the third, included as a comparison, was to try evenly spaced points. Table 6.4 shows a sample of initial points used for the optimization from the three approaches used and the criterion value of the experiment that was obtained from each set of initial points. Using the information provided by the sensitivity analysis produced the experiments with the largest criterion values (i.e. the best experiments). Random initial points produced experiments of a wide range in quality. When evenly spaced points were used the optimization algorithm did not converge, even though a number of different attempts were made by decreasing convergence tolerance and/or changing the optimization algorithm parameters. The best experiment found is shown in Table 6.5.

To determine how sensitive the experiment design is to the initial parameter values a sensitivity analysis of the designed experiment, shown in Table 6.5, was performed. This was accomplished by calculating the criterion value of the designed experiment using parameter values that span the feasible parameter space. To span this space the parameter values were perturbed according to a factorial experiment arrangement. The

source of initial values	initial values	designed experiment criterion value
sensitivity analysis information	430, 440, 450, 460, 470, 820, 830, 840, 850, 860	3.28e37
	450, 450, 450, 450, 450, 830, 830, 830, 830, 830	2.32e37
	500, 500, 500, 400, 400, 800, 800, 900, 900, 900	1.11e33
random	114, 284, 455, 465, 488, 585, 709, 757, 813, 860	6.85e31
	214, 424, 428, 592, 690, 733, 814, 833, 837, 848	1.37e35
evenly spaced	100, 200, 300, 400, 500, 600, 700, 800, 900, 900	did not converge

Table 6.4: Initial values used by the optimization algorithm and the criterion of the experiment obtained

sampling time (min.)	criterion value
415, 431, 446, 459, 473 840, 850, 855, 860, 865	3.28e37

Table 6.5: The best experiment design obtained

parameter	Low level	High level
$k_{fm\ act-eng}$	8000	18000
$k_{fm\ pre-exp}$	6.00e5	6.00e11
$f_{act-eng}$	-1000	0
$f_{pre-exp}$	0.1	1.0
V_{fop}	0.020	0.030

Table 6.6: High and low levels of the parameters used in the experiment sensitivity analysis.

high and low levels of the parameter values used to span the parameter space are given in Table 6.6. These values were obtained based on prior experience and an examination of the parameter values of similar systems.

To determine how the quality of the experiment changes over the parameter space the criterion values or the radius measure as described in Chapter 3 can be used. While the radius measure is an approximation, it is easier to perform a comparison using it than the criterion value. This is because the radius measure is linearly proportional to the average uncertainty in the individual parameter estimates (i.e. if the radius measure decreases by half, the uncertainty in the individual parameter estimates will decrease by half). This quality makes the radius a more intuitive measure that simplifies the experiment design sensitivity analysis. It is important to remember that due to the assumptions made in its development, discussed in Chapter 3, the radius is intended to be used only as a qualitative measure when comparing designed experiments.

The results of the experiment sensitivity analysis are given in Table 6.7. This table shows the perturbed parameter values, in a factorial format, and the corresponding values of the radius of the confidence region spheroid. The experiments are sorted in decreasing order of quality (i.e. increasing radius or standard deviation of the parameter estimates). The last six values are assigned values of '*inf*' as the D-optimality criterion was evaluated

to be zero. The original experiment as designed is indicated by a star on the right hand side of its radius value. An examination of the values in Table 6.7 indicates that given the chosen parameter ranges, the quality of the experiment will decrease by more than two orders of magnitude for more than half of the test points. This may be due to any of the following reasons:

- the parameter ranges chosen for the sensitivity analysis are too large
- the experiment design is very sensitive to the parameter values used
- one or more of the parameters has become 'practically' unobservable (i.e. its confidence region is very large)

To determine the effect of the chosen parameter ranges, the sensitivity analysis was repeated with smaller parameter ranges, a plus/minus ten percent change in all of the parameter values. This decreased the sensitivity of the design but there were still three orders of magnitude difference between the value of the radius at the initial parameter values and at the worst location. Further information about how sensitive the design is to the parameter values can be inferred by examining how the gradient plots change with a change in the parameter values, as shown in the previous section with the analysis of Figures 6.4 and 6.5. To determine if one of the parameters has become 'practically' unobservable and is biasing the experiment sensitivity analysis, the analysis should be repeated with the chosen parameter kept constant at the initial value used. If the criterion values change significantly when the analysis is repeated, then the chosen parameter is causing the experiment design to significantly decrease in quality. Therefore the experiment is not robust to the uncertainty in the given parameter.

Based on the results of the designed experiment sensitivity analysis it is possible that the designed experiment will not be a very useful experiment if the true parameter values are very different from the values used to design the experiment. At this point two courses of action are available. The first is to redesign or modify the experiment so

Parameter values					Radius	
$k_{fm\ act-eng}$	$k_{fm\ pre-exp}$	$f_{act-eng}$	$f_{pre-exp}$	V_{fop}		
6.00E+05	8000	0.03	1	-1000	0.00009	
6.00E+11	18000	0.03	1	-1000	0.00013	
6.58E+08	13426.8	0.025	0.6	0	0.00018	*
6.00E+05	8000	0.02	1	-1000	0.00113	
6.00E+11	18000	0.02	1	-1000	0.00185	
6.00E+11	18000	0.03	1	0	0.01091	
6.00E+05	8000	0.03	1	0	0.02416	
6.00E+11	18000	0.02	0.1	-1000	0.05707	
6.00E+05	8000	0.02	0.1	-1000	0.11868	
6.00E+05	18000	0.03	1	-1000	0.13634	
6.00E+11	18000	0.02	1	0	0.18124	
6.00E+11	18000	0.03	0.1	-1000	0.21363	
6.00E+05	8000	0.02	1	0	0.27330	
6.00E+05	8000	0.03	0.1	-1000	0.28605	
6.00E+05	18000	0.02	1	-1000	0.34334	
6.00E+11	8000	0.03	1	-1000	3.30209	
6.00E+11	8000	0.02	1	-1000	5.93090	
6.00E+11	18000	0.03	0.1	0	6.85350	
6.00E+11	18000	0.02	0.1	0	6.94367	
6.00E+05	18000	0.02	0.1	-1000	7.55549	
6.00E+05	8000	0.03	0.1	0	8.98383	
6.00E+05	8000	0.02	0.1	0	9.21568	
6.00E+05	18000	0.03	1	0	18.68176	
6.00E+05	18000	0.02	1	0	42.46608	
6.00E+05	18000	0.03	0.1	-1000	238.99123	
6.00E+05	18000	0.03	0.1	0	3812.12697	
6.00E+05	18000	0.02	0.1	0	4874.87872	
6.00E+11	8000	0.02	0.1	-1000	inf	
6.00E+11	8000	0.03	0.1	-1000	inf	
6.00E+11	8000	0.02	0.1	0	inf	
6.00E+11	8000	0.03	0.1	0	inf	
6.00E+11	8000	0.02	1	0	inf	
6.00E+11	8000	0.03	1	0	inf	

Table 6.7: Results of the designed experiment sensitivity analysis

that it is more robust, and the second is to perform the experiment as designed and then to perform further experiments once better parameter estimates have been obtained, i.e. to follow a sequential, iterative approach. Which approach to take would be decided by an examination of the amount of risk that the researcher is willing to accept in the performing the experiment and at what stage of the overall parameter estimation process is at (i.e. at the initial exploratory stage or near the end and just improving the parameter values).

6.3.4 Parameter Estimation

For the estimation of the parameters the determinant criterion was used as discussed in Chapter 5. This criterion was used due to its favorable properties, and the optimization algorithm used was the `fmins` algorithm. To estimate the parameters various initial estimates were used for the optimization algorithm. Unfortunately each of the estimates resulted in a local optimum that was very near the initial guess. It was suspected that this was due to insufficient trials in the experiment. To test this the parameters were estimated with two and five times the number of trials in the experiment. With the increased number of trials the local optima were still present and the optimization problem was similar to the original case. A different optimization algorithm was then investigated. The algorithm that was tried is the genetic optimization algorithm, (Yao, 1994; Moros, 1996), though it too found local optima. The MWLS criterion as discussed in Section 5.1.2 was also implemented. Unfortunately it did not produce results that were different from the determinant criterion.

To obtain a better understanding why the parameter estimation was not working it was attempted to estimate different subsets of the five parameters. What was discovered from this part of the investigation is that the parameters within the two Arrhenius relationships (i.e. $k_{fm\ act-eng}$, $k_{fm\ pre-exp}$, $f_{act-eng}$ and $f_{pre-exp}$) were the source of the problem. When the Arrhenius parameters for either k_{fm} or f were estimated, very

different local optima would still be found. Although the model predictions using the parameter values from the local optima were reasonable.

The general Arrhenius expression is shown in equation 6.29.

$$k = A * e^{-E/RT} \tag{6.29}$$

where A is the pre-exponential factor, E is the activation energy, R is the ideal gas constant and T is the temperature. Inspection of the equation reveals that for a given temperature there is a set of activation energies and pre-exponential factors that will produce the same rate value. Therefore it was suspected that the experiments as run did not have a sufficiently large temperature difference to adequately estimate the parameters. To examine this possibility, Figure 6.6 shows a plot of $k_{fm\ act-eng}$ versus $k_{fm\ pre-exp}$. The horizontal axis is the activation energy and the vertical axis is the log of the pre-exponential factor. The two lines represent the set of activation energy and pre-exponential factor values that produce the same rate value at a fixed temperature. The solid line is for 65 °C and the dashed line is for 75 °C . The two lines cross at the true value of the activation energy and the pre-exponential factor, shown by a circle on the plot. Included in the plot are all of the point estimates from the local optima obtained in the simulation studies, shown as stars. An inspection of the plot reveals that the temperature difference used in the experiment resulted in the two temperature lines being close together. This may contribute to the problem of local optima that was found.

6.3.5 Confidence Regions

To determine the quality of the parameter estimates, confidence regions need to be generated. In general the joint confidence region is a p dimensional space, where p is equal to the number of parameters estimated. Since it is not possible to view this space if p is greater than three, joint confidence regions of two parameters are usually obtained. It should be remembered that these regions are conditional confidence regions at fixed

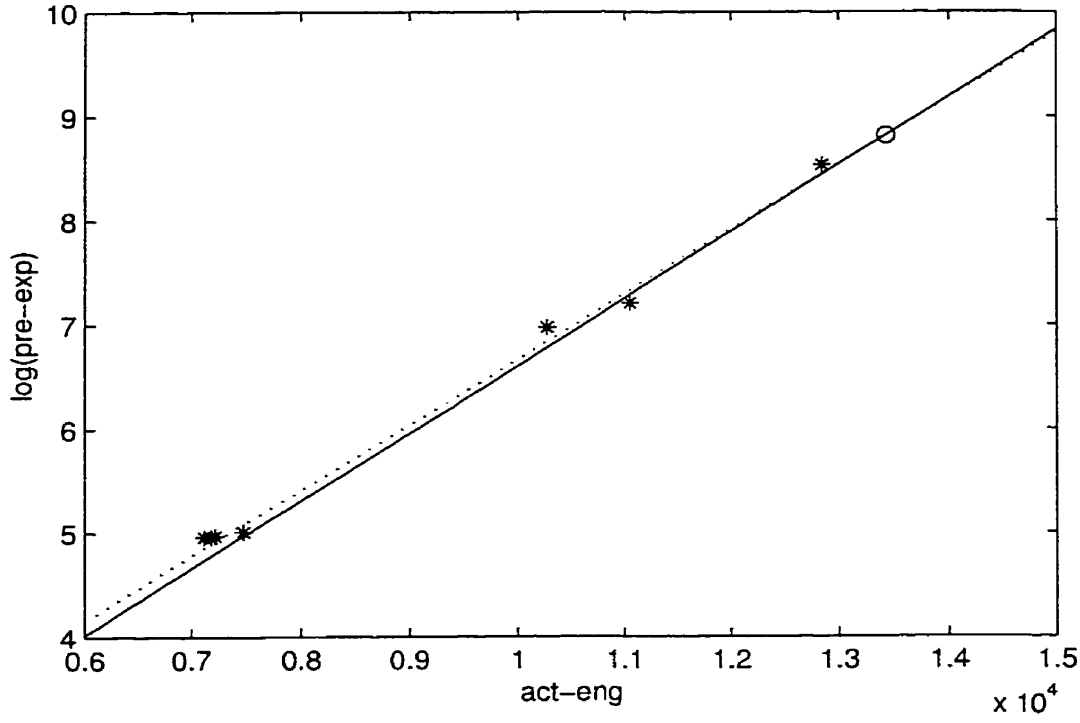


Figure 6.6: Plot of the estimated parameter values of the k_{fm} Arrhenius expression; stars are locally optimal estimates and the circle indicates the true values

$k_{fm\ act-eng}$	12682
$k_{fm\ pre-exp}$	3.311E8
$f_{act-eng}$	0.0114
$f_{pre-exp}$	0.503
V_{fop}	0.0196

Table 6.8: Estimated parameter values

temperature	sampling time
65	840 850 855 860 865
75	415 431 446 459 473

Table 6.9: Experiment design used in the parameter estimation

values of the other parameters, and that they will change with a change in the values of the other parameters.

The confidence regions are based on the parameter estimates given in Table 6.8, obtained from the experiment given in Table 6.9. The values that correspond to the 95 percent confidence contour were obtained using equation 5.14 from Chapter 5.

A set of conditional joint confidence regions of the estimated parameters are given in Figures 6.9, 6.10 and 6.11. These figures are only a subset of all of the possible joint confidence regions that can be generated. If the parameter values are not changed from the point estimates then there are five choose two (i.e. 10) possible parameter combinations for which confidence regions can be generated.

Figure 6.9 shows the 95% joint confidence region of the Arrhenius parameters for k_{fm}. The star indicates the point estimate. The confidence region is banana shaped and is not closed at the top of the plot. To determine if the joint confidence region is closed or open for the $k_{fm\ pre-exp}$ parameter, a cross section of the confidence region was obtained at increasing values of the $k_{fm\ pre-exp}$ parameter. The joint confidence region was found

to still be open at a value of 100e9, therefore it may be considered to be open, or very large, indicating a very poor parameter estimate for the $k_{fm\ pre-exp}$ parameter.

Figure 6.10 shows the 95% joint confidence region of the Arrhenius parameters for f. The star indicates the point estimate. The confidence region indicates a good parameter estimate for $f_{pre-exp}$, but a poor estimate for $f_{act-eng}$, as the contour appears to be open on the left and right hand sides of the plot. Negative values of $f_{act-eng}$ were not considered feasible and the contour was still open at a value of 100 for $f_{act-eng}$. This poor estimate of $f_{act-eng}$ is a reflection of the low level of information provided by the responses used in the gradient plots as shown in Figures 6.7 and 6.8. An examination of these figures indicates that the normalized gradient values are very small for most of the responses. The radical concentration provides most of the information about the parameter where the gradient value has a very sharp peak at 850 min. at 65 °C and at 480 min. at 75 °C . The largest normalized gradient value is -1.5 for the response of radical concentration at 850 min. and 65 °C . While this value is largest for this parameter in general it is a small normalized gradient value, and an indication that a large number of trials will be required to obtain good estimates of the parameter.

Figure 6.11 shows the 95% joint confidence region of the parameters for V_{fop} and $k_{fm\ act-eng}$. The star indicates the point estimate. The confidence region is very nonlinear and appears closed except for the area at the bottom of the graph.

An important point to remember is that joint confidence regions will only indicate the amount of uncertainty present in the parameter estimates based on the point estimates obtained and the shape of the objective function surface in the neighborhood of those estimates. Therefore, if a local optimum is found, the confidence region is for the local optimum. To explore this, a cross section of the objective function surface was generated. This was obtained by calculating the objective function values over a range of values of the $k_{fm\ act-eng}$ parameter. A plot of this cross section is shown in Figure 6.12. The horizontal axis is the parameter value, the vertical axis is the objective function value,

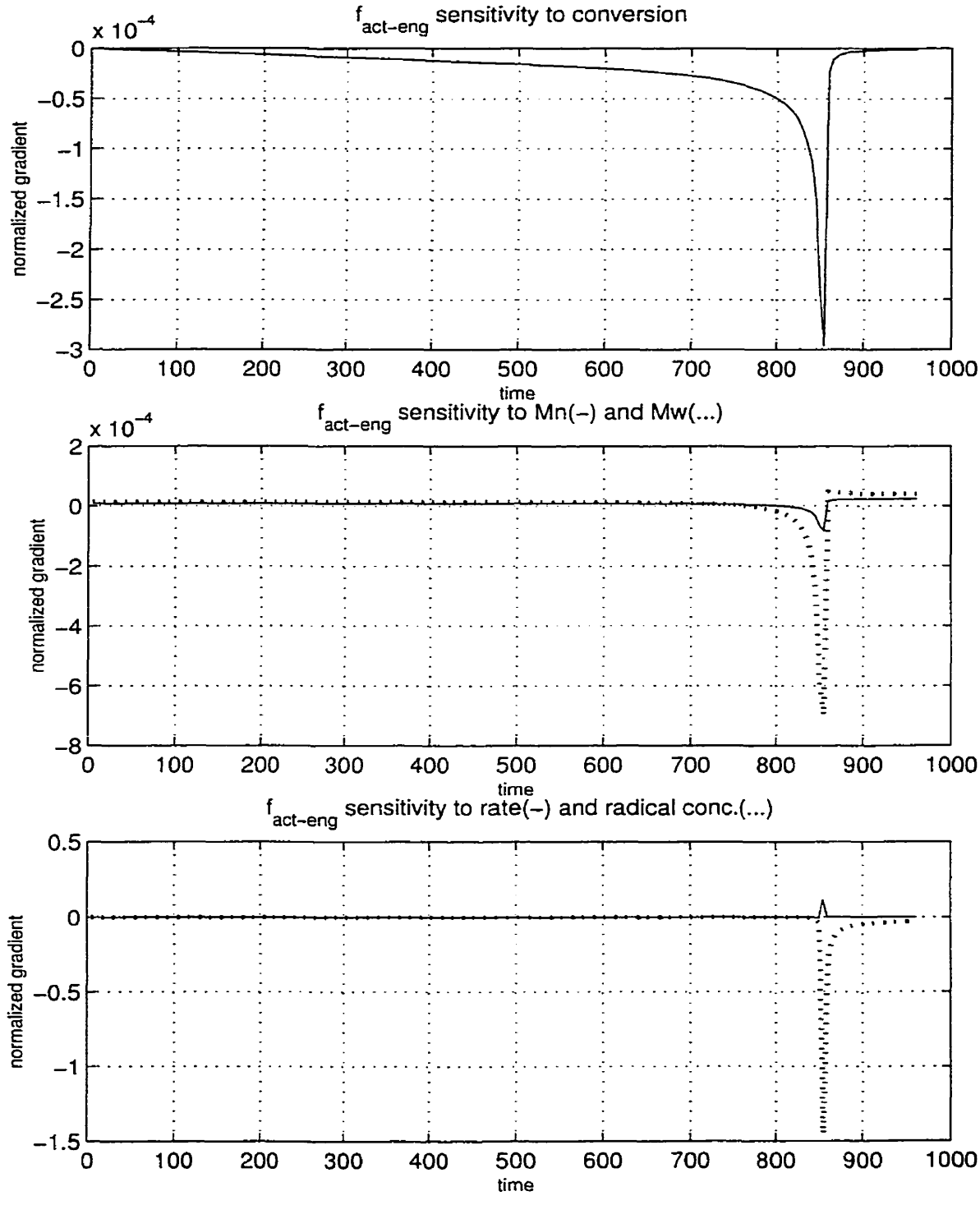


Figure 6.7: Gradient plots with respect to the parameter $f_{act-eng}$ at 65 °C .

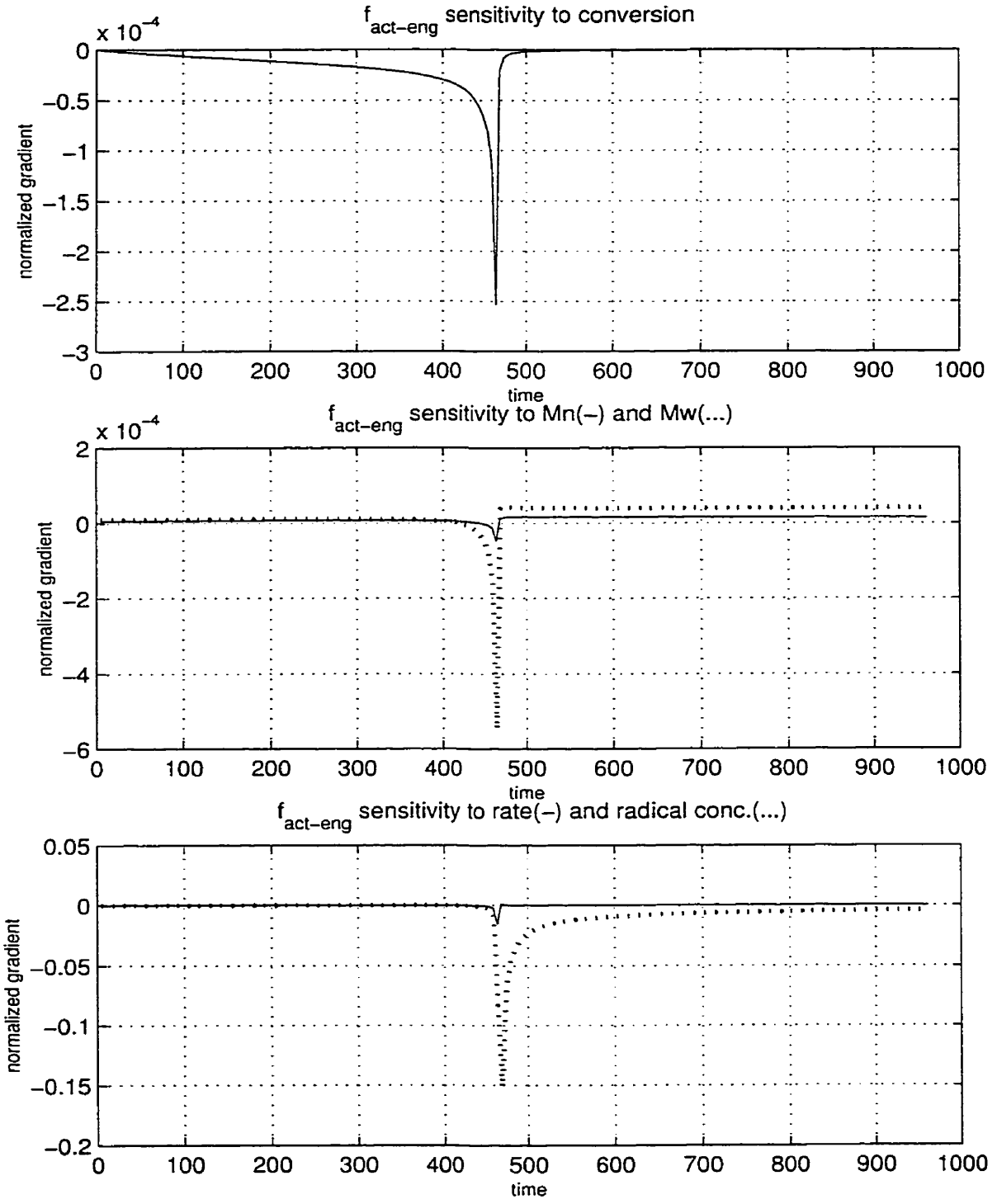


Figure 6.8: Gradient plots with respect to the parameter $f_{act-eng}$ at 75 degrees Celsius.

the star is the point estimate and the horizontal dotted line indicates the criterion value that corresponds to a 95 percent confidence level. This plot revealed that the point estimate obtained is a local minimum and that a number of other local minima exist, illustrating the difficulty of the optimization problem.

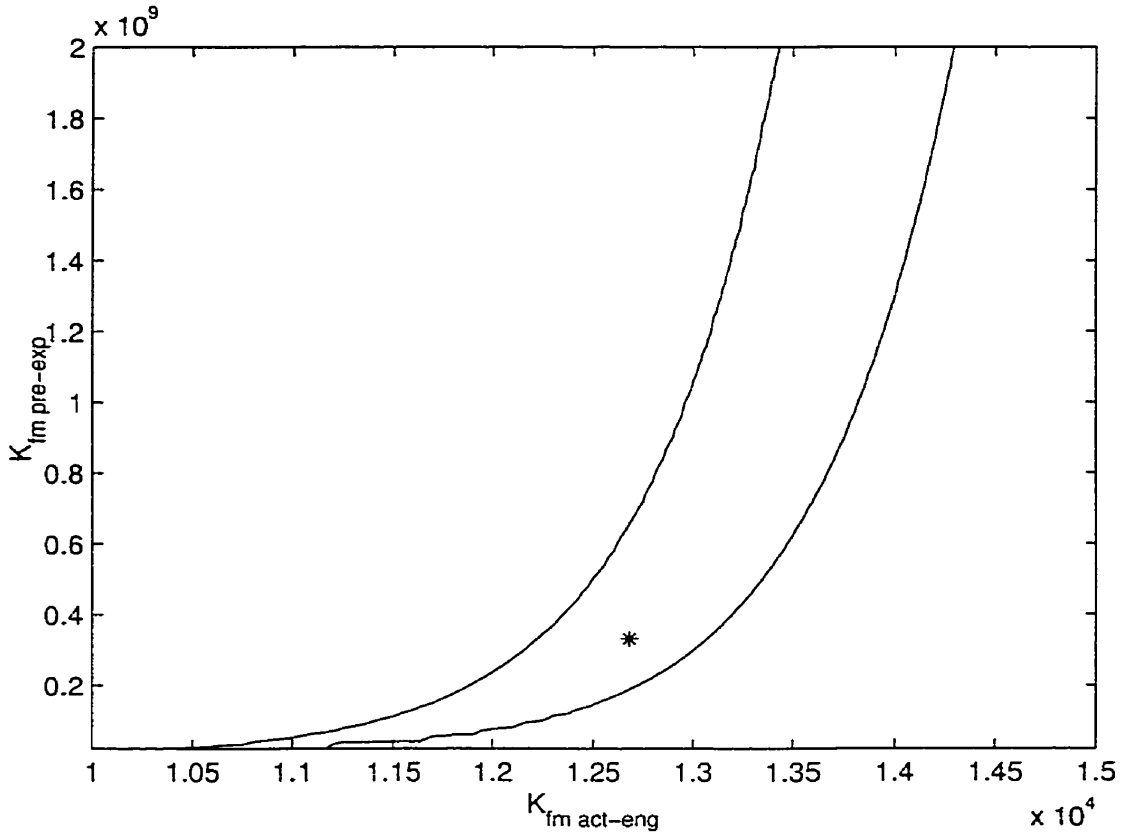


Figure 6.9: 95 percent confidence region for the parameters $k_{fm\ act-eng}$ and $k_{fm\ pre-exp}$

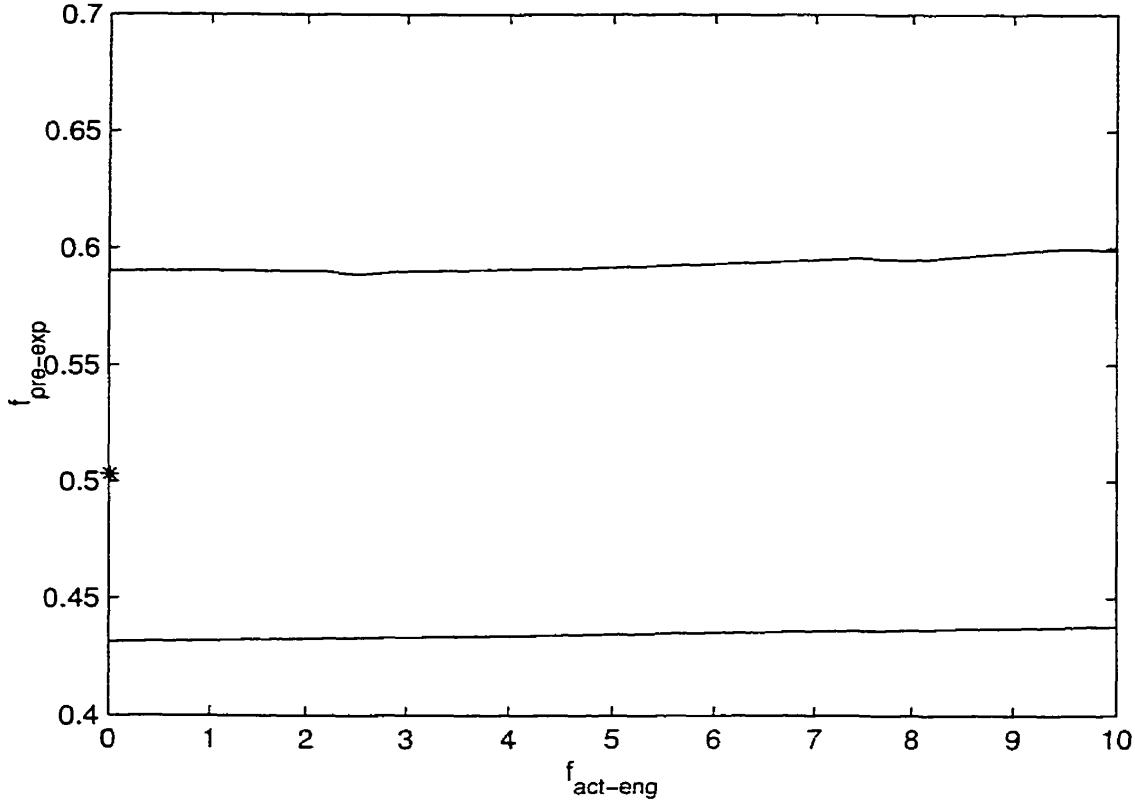


Figure 6.10: 95 percent confidence region for the parameters $fact - eng$ and $fpre - exp$

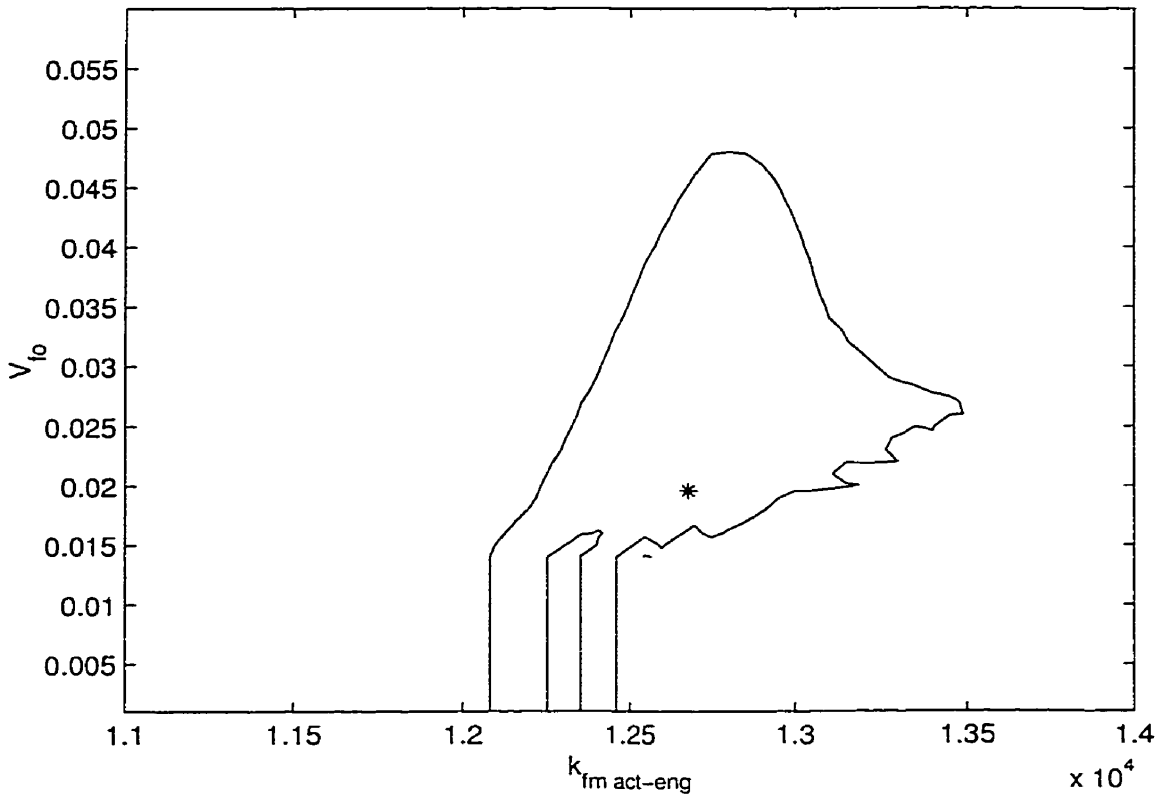


Figure 6.11: 95 percent confidence region for the parameters $k_{fm\ act-eng}$ and V_{fo}

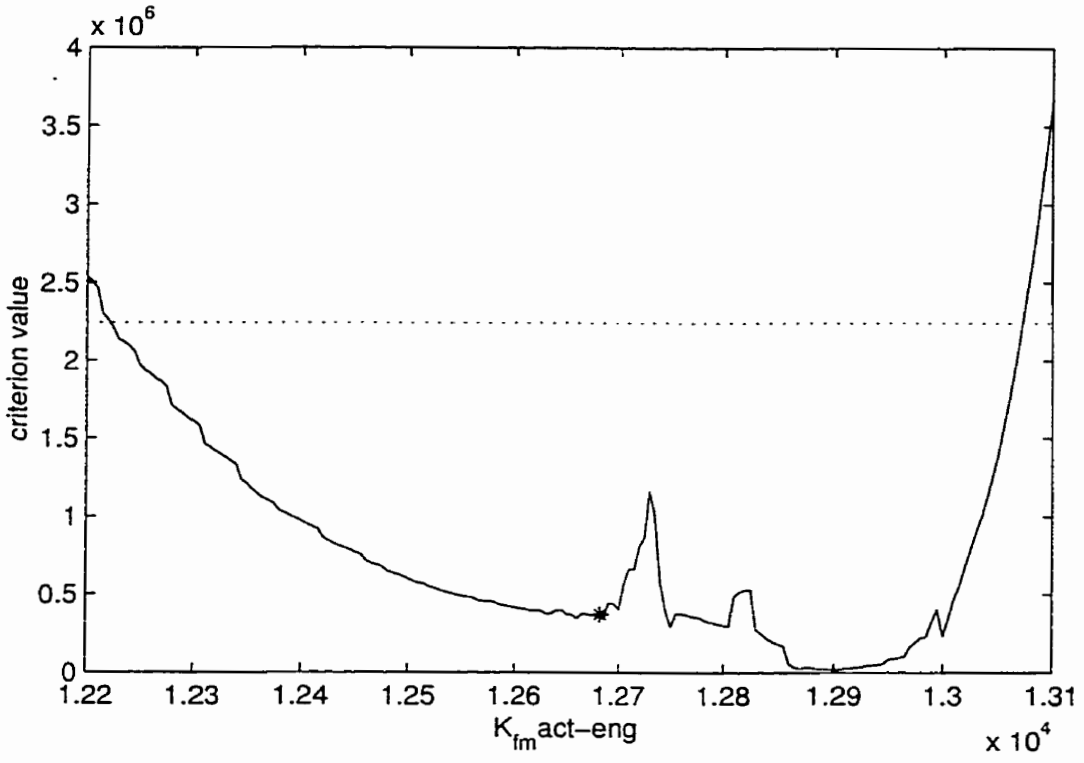


Figure 6.12: Plot of the criterion value versus $k_{fm\ act-eng}$

6.4 Case Study 2

6.4.1 Description

This case study will describe the estimation of three parameters within the Watpoly model for the homopolymerization of Styrene using five responses. A listing of the estimated parameters is given in Table 6.10. The responses used and their respective measurement error standard deviations (σ_{resp}) are given in Table 6.11.

6.4.2 Sensitivity Analysis

The sensitivity analysis was performed with respect to the three parameters listed in Table 6.10, and is summarized in Table 6.12. In this table the comments column is a summary of the observability of each parameter, where any areas of large or small gradient values (i.e. good or poor observability) are identified. As an example, at the top of Table 6.12, for the parameter α_m and the response of conversion the comments column states 'broad peak at 850 min.'. This area corresponds to the area of large gradient values as shown in Figure 6.13.

To illustrate how parameter observability can differ between two parameters the sensitivity analysis of the parameters α_m and BBm will be compared. The gradient plots with respect to these parameters are shown in Figures 6.13 and 6.14, respectively. In these plots the horizontal axis is time and the vertical axis is the normalized gradient value. The gradient values were normalized by dividing by the standard deviation of the response measurement error, as given in Table 6.11, and by multiplying by a percentage of the parameter value, as discussed in Chapter 3.

The level of a parameter's observability is determined by the magnitude of the normalized gradient values. To maximize parameter observability, trials should be performed in the area where the absolute normalized gradient values are largest. An examination of the gradient plots for the parameter α_m, shown in Figure 6.13, indicates that the gradient

Parameter	Description
α_m	variation of free volume with temperature for monomer (free volume units/K)
BBm	rate of decrease of k_p with free volume (L/mol min per free volume unit)
Tg_{M_1} †	glass transition temperature of the polymer (K)

† For a homopolymerization $Tg_{M_1} = Tgp$

Table 6.10: Parameters estimated in case study 2

Response	Measurement error standard deviation (σ_{resp})
conversion	0.025
M_n	10000
M_w	10000
rate	0.001
radical conc.	25%

Table 6.11: Measurement error standard deviation (σ_{resp}) of the responses used in case study 2

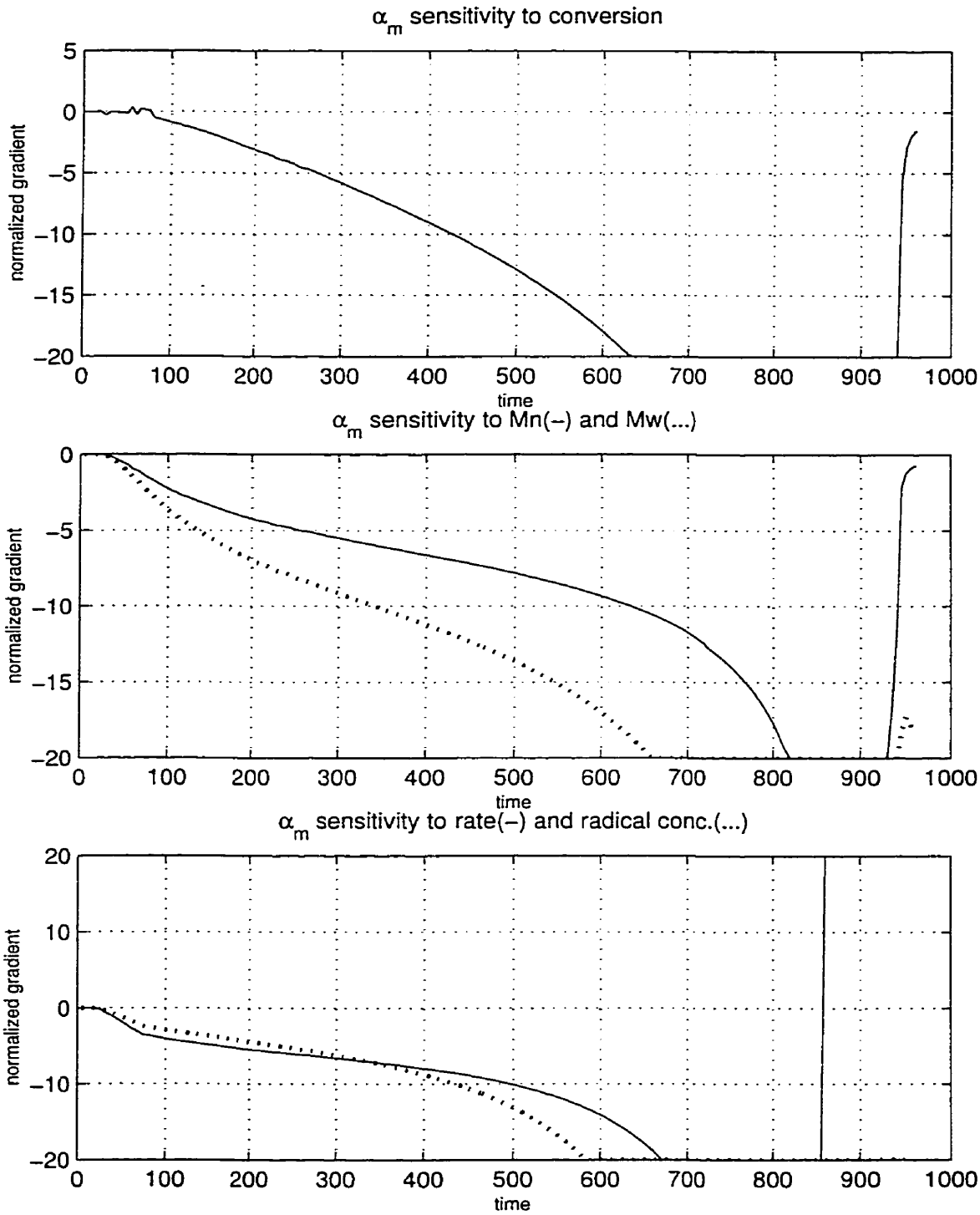


Figure 6.13: Gradient plots with respect to the parameter α_m at 65 degrees Celsius.

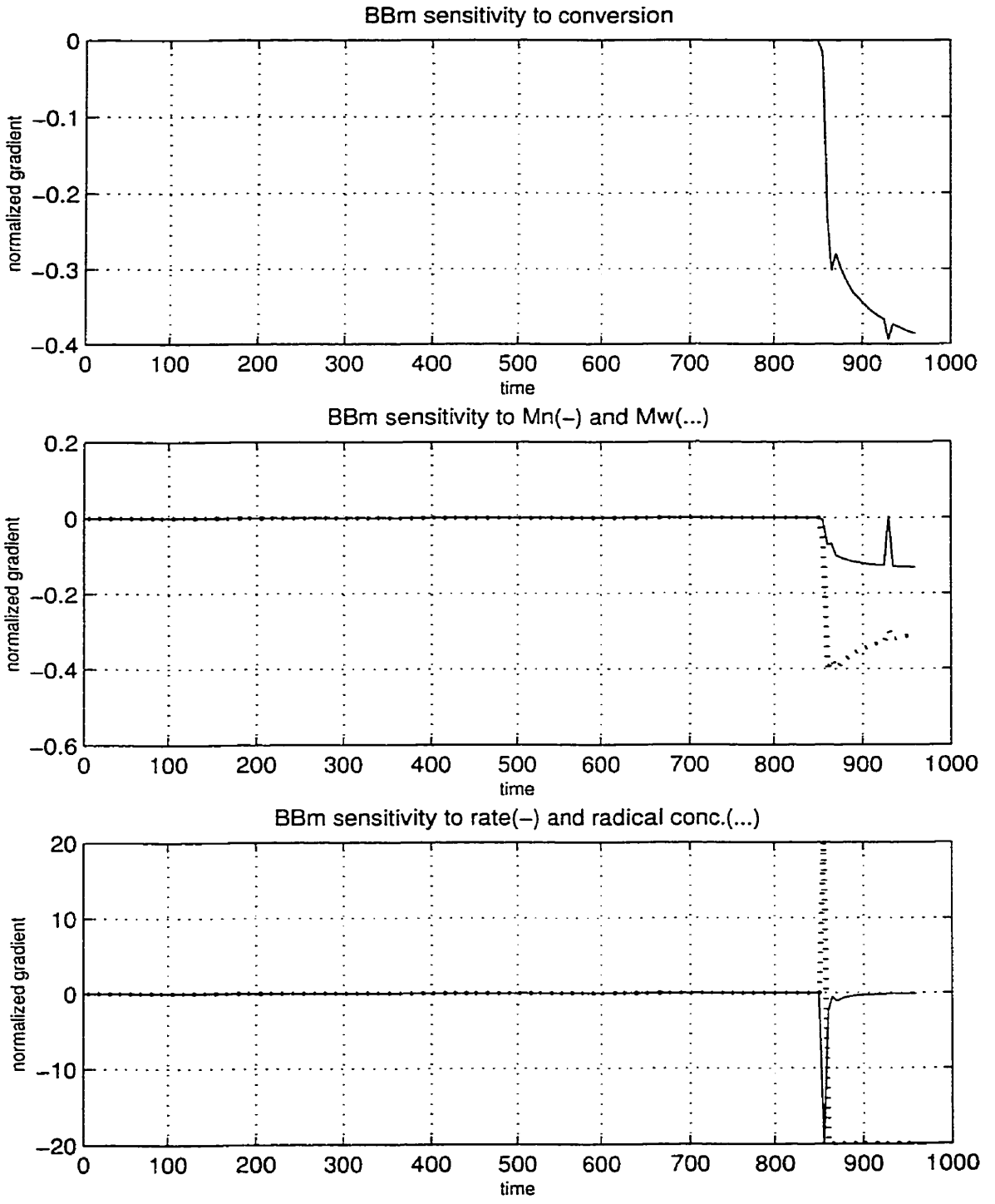


Figure 6.14: Gradient plots with respect to the parameter BBm at 65 degrees Celsius.

the gradient plots for the parameter α_m, shown in Figure 6.13, indicates that the gradient values are maximized between 800 and 900 minutes. The gradient values are also greater than 20 for a wide range of sampling times. This indicates that there is a large region of time at which a large amount of information about the α_m parameter is available.

The gradient plots of the BBm parameter are given in Figure 6.14. An inspection of Figure 6.14 reveals that the gradient values are almost zero for all of the low and medium conversion phase of the polymerization, less than 850 minutes. Low and medium conversion is defined as conversion before the autoacceleration effect, and is identified from a plot of conversion versus time for the polymerization, as shown in Figure 6.15. In Figure 6.15 the vertical axis is conversion, where 1.0 indicates 100% conversion, and the horizontal axis is time. The auto-acceleration section is the steep section of the conversion versus time curve (i.e. from 800 to 850 minutes). The BBm parameter does become observable once high conversion is reached. The lack of observability of the BBm parameter at the low and medium conversion is expected, as BBm is a parameter related to a phenomenon that only takes effect once high conversion is reached and the rate of polymerization becomes diffusion controlled.

To determine the effect of changing the BBm parameter values on the sensitivity analysis, gradient plots were generated for different values of BBm. Figure 6.16 shows the gradient plots for the M_w response at values of 0.9, 1.0 and 1.1 for BBm. An inspection of the plot indicates that the location of observability will not change as the value of BBm changes, only the level of observability will change. Therefore the experiment design using the response of M_w will not be affected by uncertainty in the value of the BBm parameter. This is contrast to the parameter V_{fo} in case study one, where the location and magnitude of the observability changed with a change in the parameter values (Figure 6.5).

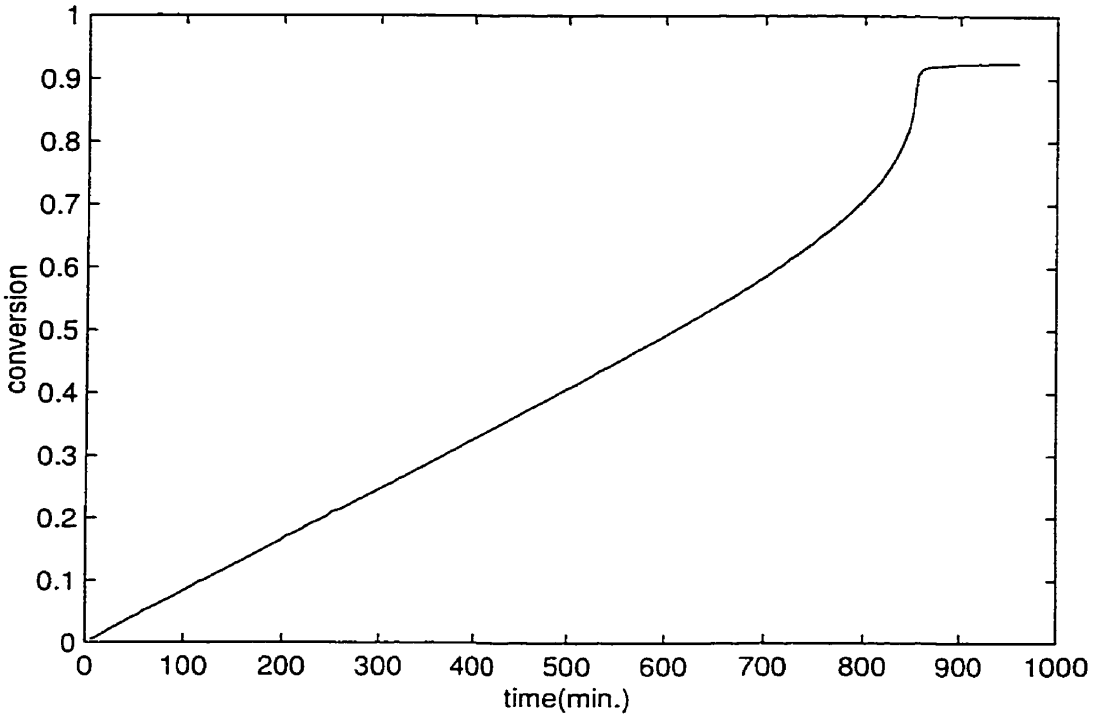


Figure 6.15: Plot of conversion versus time for the homopolymerization of Styrene, at 65 degrees Celsius

Parameter	Response	Comments
α_m	conversion	broad peak at 850 min
	M_n	broad peak at 850 min
	M_w	broad peak at 850 min
	rate	very large gradient values at time > 800 min
	radical conc.	very large gradient values at time > 700 min
BBm	conversion	no observability at < 850 min, poor observability at > 850 min
	M_n	no observability at < 850 min, poor observability at > 850 min
	M_w	no observability at < 850 min, poor observability at > 850 min
	rate	no observability at < 850 min, poor observability at > 850 min
	radical conc.	no observability at < 850 min, very good observability at > 850 min
Tg_{M_1}	conversion	broad peak at 850 min
	M_n	peak at 850 min
	M_w	broad peak at 850 min
	rate	broad peak at 850 min
	radical conc.	very large gradient values at time > 700 min

Table 6.12: Summary of the sensitivity analysis

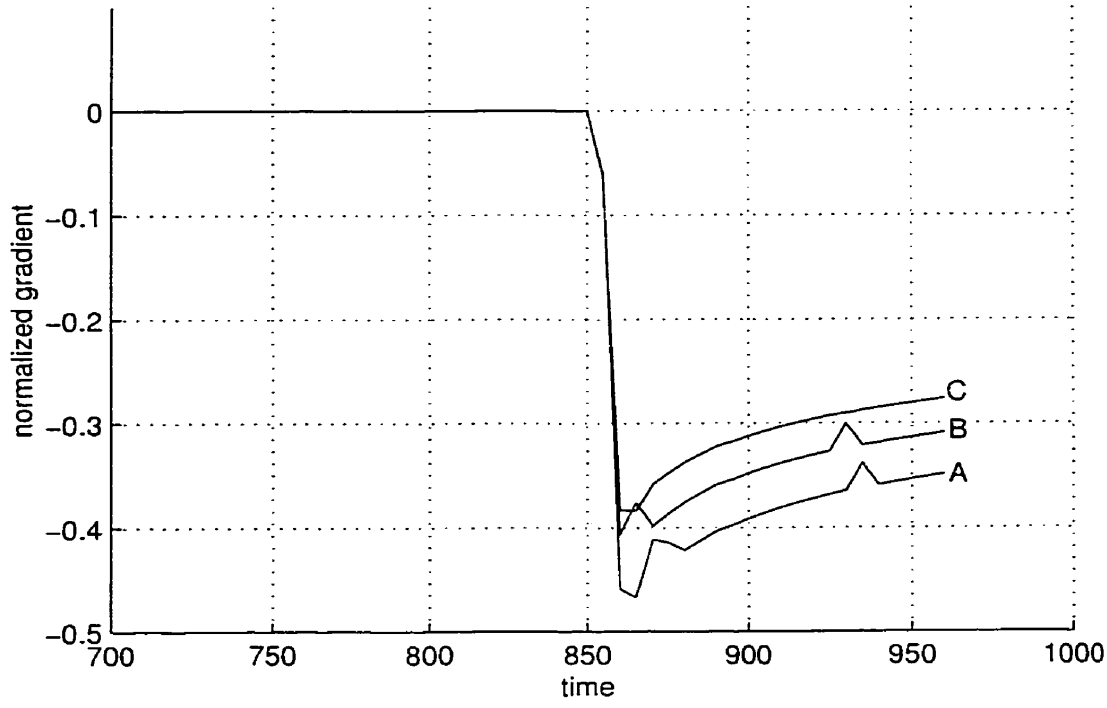


Figure 6.16: Effect of changing the value of BBm on the gradient plot of M_w at 65 degrees Celsius, where BBm equals 0.9(A), 1.0(B) and 1.1(C)

6.4.3 Experiment Design

The objective of experiment design was to find an experiment that would maximize the amount of information provided about the parameter values. The experiment was designed using the D-optimality criterion as discussed in Chapter 3. The size of the experiment was chosen to be 10 trials.

In designing the optimal experiment, as expected, local optima were found. To deal with this, experiments were designed using the information provided in the sensitivity analysis to obtain the initial values for the optimization algorithm. A sample of the designed experiments is shown in Table 6.13. Table 6.13 gives the initial values used in the optimization algorithm, the final experiment obtained, its criterion value and the radius. The radius is given as a qualitative measure to compare the experiment design, as discussed in Chapter 4. In the table three experiments are given, where the first two were designed using initial values based on the sensitivity analysis information, and as a comparison the third experiment designed by the author based solely on an inspection of the gradient plots. Once again, when comparing experiments it can be misleading to compare the design criterion values directly as they are proportional to the volume of the linearized joint confidence region. Therefore it would be misleading to say that the best experiment is 2.5 times better that the author's experiment. Using the radius approximation method the radius of the author's experiment and the best experiment are 50757 and 59371, respectively. Therefore this analysis illustrates that the best experiment is only approximately 20 percent better.

6.4.4 Parameter Estimation

For the estimation of the parameters the determinant criterion was used as discussed in Chapter 4 and the optimization algorithm used was the `fmins` algorithm. To estimate the parameters various initial estimates were used for the optimization algorithm. It was observed that each of the initial parameter values used in the optimization produced a

initial values used in the optimization (min.)	designed experiment		
	trials	criterion value	radius
600 700 800 850 900 600 700 800 850 900	584 588 849 843 960 621 692 850 885 924	2.93e28	55523
700 750 800 850 900 725 775 825 875 925	626 766 925 841 960 657 850 849 892 898	4.38e28	59371
	600 650 700 800 825 850 875 900 925 950	1.71e28	50757

Table 6.13: Initial values used by the optimization algorithm, the designed experiment and its criterion value

local optimum, where the parameter values obtained were close to the initial guess or were considered poor estimates. A set of parameters was considered poor if its criterion value was much larger than the best criterion values obtained to date and/or if the parameters are in a region determined to be not feasible based on prior knowledge. To illustrate the variability of the parameter estimates obtained due to the local optima, a sample of the initial and estimated parameter values as well as the criterion value obtained are given in Table 6.14.

		initial values	final values	criterion
simulation 1	α_m	0.001	0.001011	1.8142e-3
	BBm	1.0	1.02	
	Tg_{M_1}	383.15	383.49	
simulation 2	α_m	0.0015	0.00102	8.2264e0
	BBm	1.5	1.147	
	Tg_{M_1}	400	390.23	
simulation 3	α_m	0.001	0.00102	1.7217e1
	BBm	1.5	1.542	
	Tg_{M_1}	390	396.94	
simulation 4	α_m	0.0015	0.00107	2.7695e6
	BBm	0.5	0.536	
	Tg_{M_1}	350	418.71	

Table 6.14: Local optima from the parameter estimation

To illustrate why all of the local optima are occurring, a part of the criterion surface was generated. Since the criterion surface is three dimensional, to generate a 2D surface the Tg_{M_1} parameter was fixed at the point estimate and the remaining two parameters were varied. The generated surface is shown in Figure 6.17, where the horizontal axes are the parameters α_m and BBm and the vertical axis is the logarithm of the criterion values. The logarithm of the criterion values was used to compress the vertical scale to improve the presentation of the criterion surface. An examination of the criterion surface indicates two ridges that will result in local optima. These two ridges are parallel to the BBm parameter axis and occur near a value of 1.04e-03 and 1.02e-02 of the α_m parameter.

To deal with the problem of local optima two different approaches were attempted. The first was to use the optimization algorithm of simulated annealing that is robust to local optima, as discussed in Chapter 4. The second was to evaluate the quality of 500 starting points using the criterion value, and to use the point with the smallest criterion value as the starting point for the `fmins` algorithm. The parameter point estimates obtained from the two methods as well as the true parameter values are given in Table 6.15. The simulated annealing method obtained the best parameter estimates but it was more computationally expensive. The method required 8867 function calls versus a typical 150-300 for the `fmins` algorithm and 697 for the 500 guess method, 100 function evaluations required approximately 8 min to complete on a Pentium III, 500 MHz.

	simulated annealing	500 guess	true values
α_m	0.00100	0.00097	0.001
BBm	1.00103	0.95519	1.0
Tg_{M_1}	378.2	369.2	370.0
criterion	3.2795e-005	2.0499e-002	

Table 6.15: Parameter estimates and true values

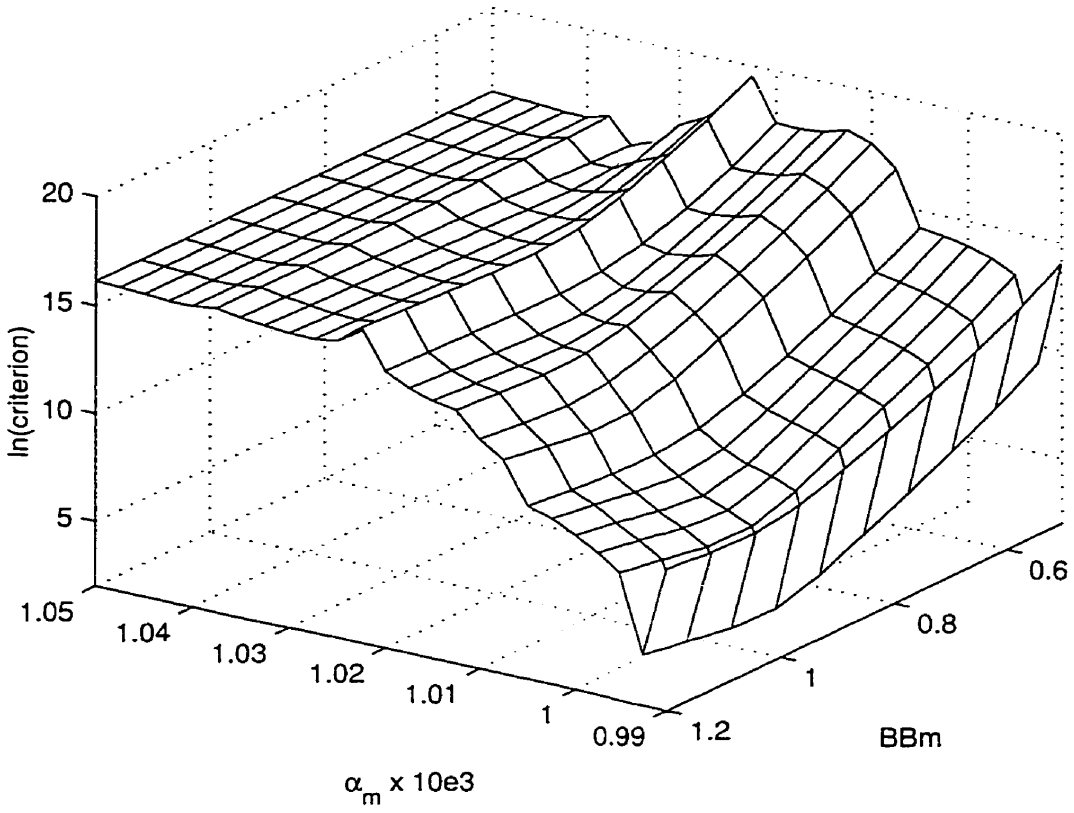


Figure 6.17: A sample of the criterion surface, varying the parameters α_m and BBm while Tg_{M_1} is fixed

6.4.5 Confidence Regions

The confidence regions obtained are based on the parameter estimates given in Table 6.16, which have been obtained from the experiment given in Table 6.17. The value that corresponds to the 95 percent confidence contour was obtained using equation 6.13 from Chapter 6.

α_m	BBm	Tg_{M_1}
0.00099	1.052	374.2

Table 6.16: Estimated parameter values

A sample of the conditional joint confidence regions of the estimated parameters are given in Figures 6.18, 6.19 and 6.20. These figures are only a subset of all of the possible joint confidence regions that can be generated.

Figure 6.18 shows the 95% joint confidence region of the parameters α_m and BBm. The star indicates the point estimate. The confidence region is not closed at the top of the plot. To determine if the joint confidence region is closed or open for the BBm parameter, a cross section of the confidence region space was obtained at increasing values of the BBm parameter. The joint confidence region was found to still be open at a value of 5000, therefore it may be considered to be open, or very large, indicating a poor parameter estimate for the BBm parameter.

Figure 6.19 shows the 95% joint confidence region of the parameters α_m and Tg_{M_1}. The star indicates the point estimate. The confidence region is not closed at the top of the plot. To determine if the joint confidence region is closed or open for the Tg_{M_1}

sampling time (min.)
626, 657, 766, 841, 849, 849, 891, 898, 925, 959

Table 6.17: Experiment design used in the parameter estimation

parameter, a cross section of the confidence region space was obtained at increasing values of the Tg_{M_1} parameter. The joint confidence region was found to still be open at a value of 1000, therefore it may be considered to be open, or very large, indicating a poor parameter estimate for the Tg_{M_1} parameter.

Figure 6.20 shows the 95% joint confidence region of the parameters α_m and BBm at different values of the Tg_{M_1} parameter. The star indicates the point estimate. The three confidence regions were obtained at values of 370.0, 374.2 and 380.0 for Tg_{M_1}, and are labeled A, B and C, respectively. The three confidence regions are similar in shape but are shifted with respect to the α_m parameter. This indicates that the marginal confidence region for the α_m parameter is significantly larger than the conditional confidence region indicates. Therefore to use the conditional confidence region as a representation of the overall α_m parameter uncertainty will be grossly misleading. For the BBm parameter there is not much change and the conditional confidence region may be representative of the marginal confidence region.

In the generation of the true shape approximate level confidence regions the assumption is made that the ratio given in equation (6.20), reproduced below, is F distributed.

$$\frac{[|Z'Z| - |Z'Z|]/p}{s^2} \tag{6.30}$$

As shown in Chapter 5 if multiple responses are used this assumption may not be correct. To test if this assumption is valid in this case study a Monte Carlo simulation was carried out. This simulation involved the generation of sets of data with 10 and 5 trials from which the parameters were estimated. In order to avoid the problem of local optima in the estimation, the true values were used as the initial values for the optimization algorithm.

Confidence region theory states that upon repeated sampling the true parameter values will occur within a 90 percent parameter estimate confidence region, 90 percent of the time. The results of the Monte Carlo study are shown in Table 6.18. This table indicates the percent of the time that the true parameter values were found within the true

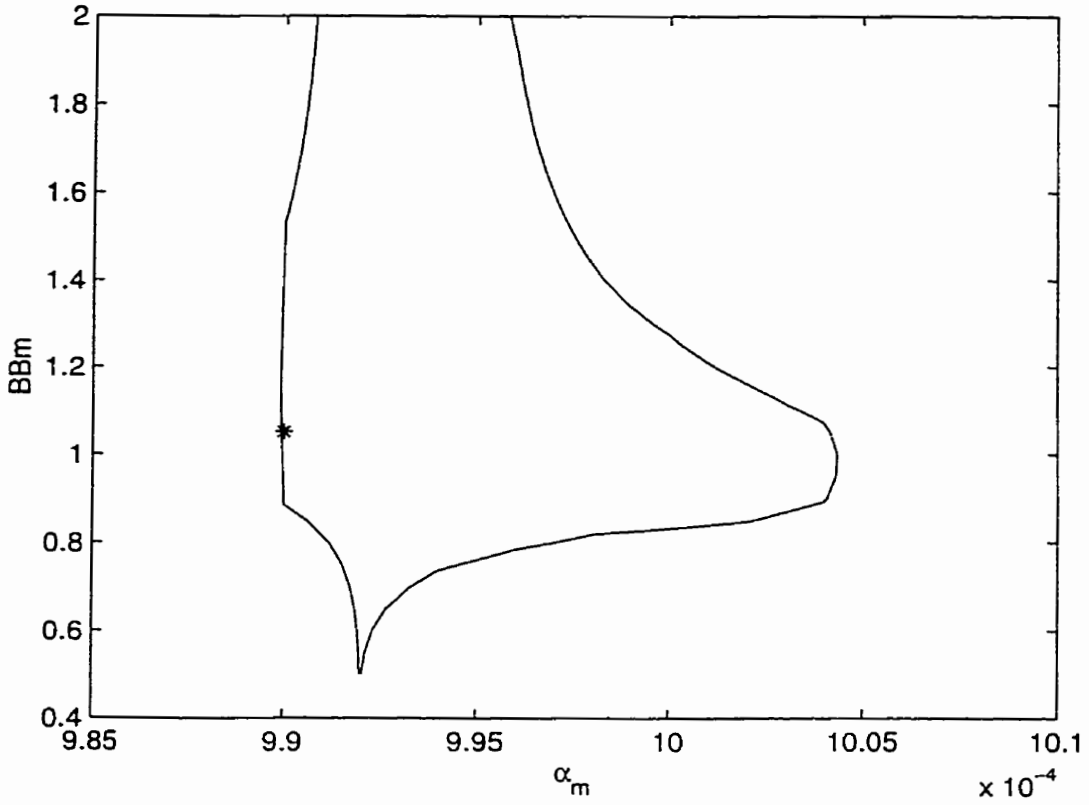


Figure 6.18: 95 percent confidence region for the parameters α_m and BBm

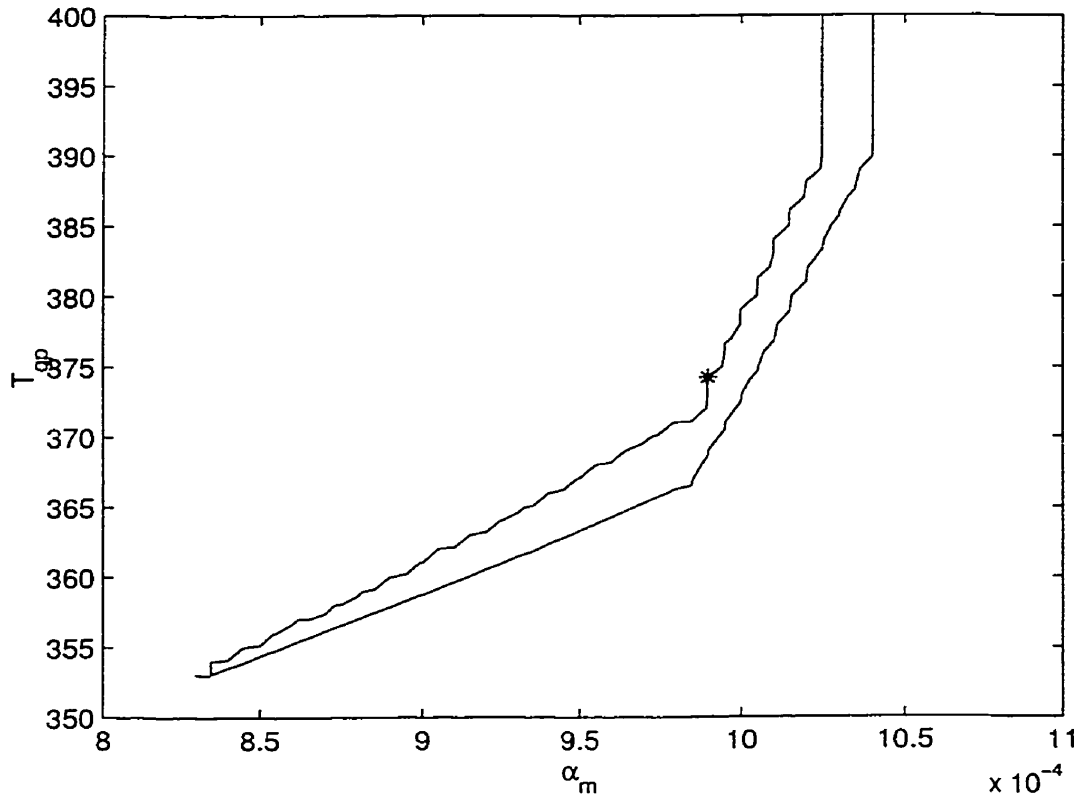


Figure 6.19: 95 percent confidence region for the parameters α_m and Tg_{M_1}

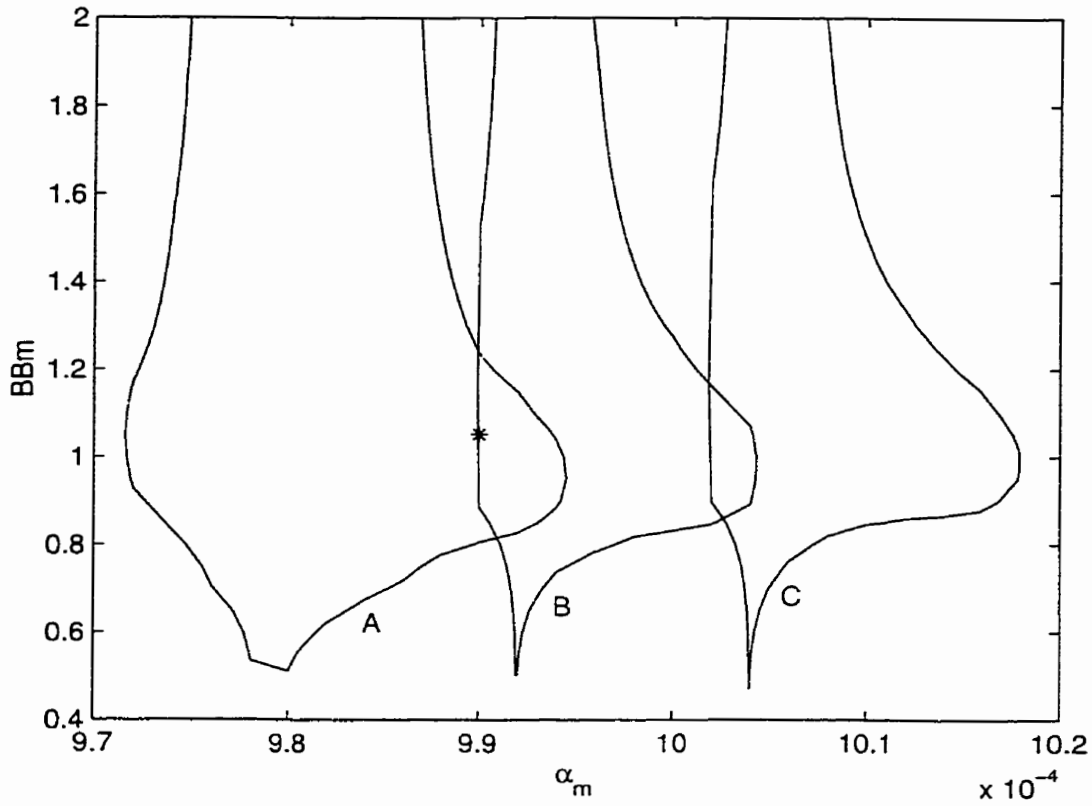


Figure 6.20: 95 percent confidence region for the parameters α_m and BBm at different values of Tg_{M_1}, where Tg_{M_1} equals 370.0(A), 374.2(B) and 380.0(C)

number of trials	theoretical confidence level	actual confidence level
5	50	14
	90	39
	95	48
10	50	80
	90	97
	95	99

Table 6.18: Estimated parameter values

shape approximate level confidence region of the parameter estimates. The theoretical confidence level is the expected percent of the time the true values will occur within the true shape approximate level confidence region, and the actual confidence level is the percent of the time that they did occur.

From Table 6.18 we can see that the difference between the theoretical and actual confidence level is very large when the sample size is five, thus indicating that the assumption that the ratio in equation 6.30 is F distributed is not valid. When the sample size is ten, the difference between the theoretical and actual is much smaller, and can be considered acceptable. As this small difference can be attributed to the confidence region being at an approximate confidence level.

Since the actual confidence level for the five trial case is much smaller than the theoretical, the true shape confidence regions obtained for any parameter estimates will be misleading as they will indicate a level of confidence that is much larger than is true.

6.5 Case Study 3

6.5.1 Description

This case study will describe the estimation of five parameters within the Watpoly model for the copolymerization of Styrene / Methyl Methacrylate using five responses. A listing of the estimated parameters is given in Table 6.19. The responses used and their respective measurement error standard deviations (σ_{resp}) are given in Table 6.20.

6.5.2 Sensitivity Analysis

In this case study the sensitivity analysis was performed with respect to two independent variables, time and initial feed composition. With two independent variables the gradient values produce a surface. This surface is illustrated in Figures 6.21 and 6.22 where the gradient surface of the conversion response with respect to the parameters $\alpha_{m\ M1}$ and $Tg_{M_1M_2}$ is shown. In these figures both a plot of the gradient surface and a contour of this surface is given. In the surface plots the horizontal axes are time and initial feed composition and the vertical axis is the normalized gradient value. The gradient values were normalized by dividing by the standard deviation of the response measurement error (σ_{resp}), as given in Table 6.20, and then multiplied by a percentage of the parameter value, as discussed in Chapter 2. In the contour plots the horizontal axis is time, the vertical axis is initial feed composition and the normalized gradient contour line spacing is equal to four normalized gradient units.

A parameter's observability is determined by locating areas where the absolute normalized gradient values are large. The gradient surface for the parameter $\alpha_{m\ M1}$ is shown in Figure 6.21. An examination of this figure reveals a valley (where the absolute gradient values are maximized) that runs diagonally from a feed concentration of 0.1 and a time of 300 minutes to a feed concentration of 0.8 and a time of 900 minutes. The start of this valley can be seen in the surface plot but the valley is more apparent in the contour

Parameter	Description
$\alpha_{m\ M1}$	variation of free volume with temperature for monomer 1 (free volume units/K)
$\alpha_{m\ M2}$	variation of free volume with temperature for monomer 2 (free volume units/K)
BBm_{M1}	rate of decrease of k_p with free volume of monomer 1 (L/mol min per free volume unit)
BBm_{M2}	rate of decrease of k_p with free volume of monomer 1 (L/mol min per free volume unit)
$Tg_{M_1M_2}$	glass transition temperature of the alternating copolymer (K)

Table 6.19: Parameters estimated in case study 3

Response	Measurement error standard deviation (σ_{resp})
conversion	0.025
composition	0.025
M_n	10000
M_w	10000
rate	0.001
radical conc.	25%
triad fractions	0.1
(sequence length distr.)	

Table 6.20: Measurement error standard deviation (σ_{resp}) of the responses used in case study 3

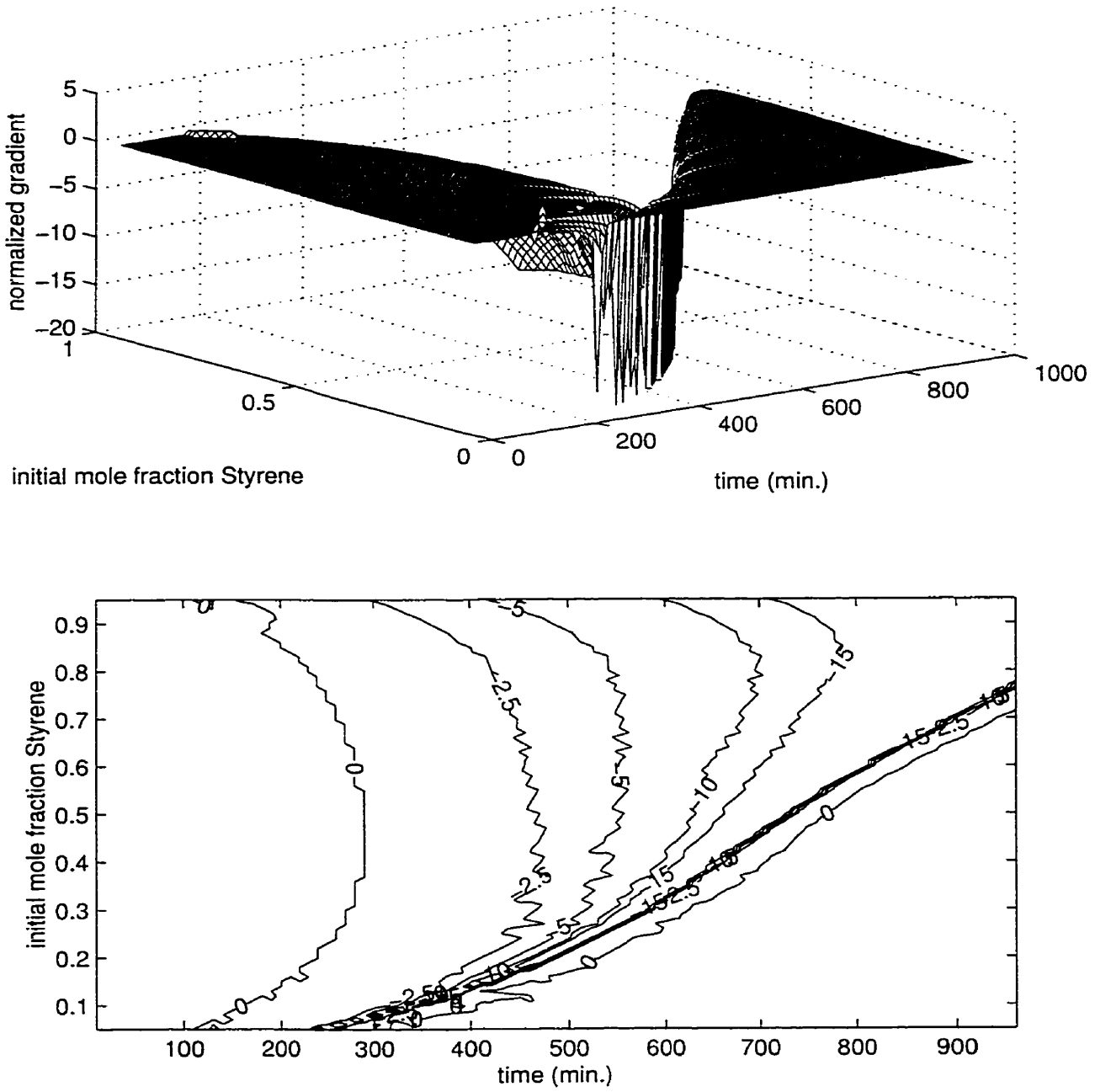
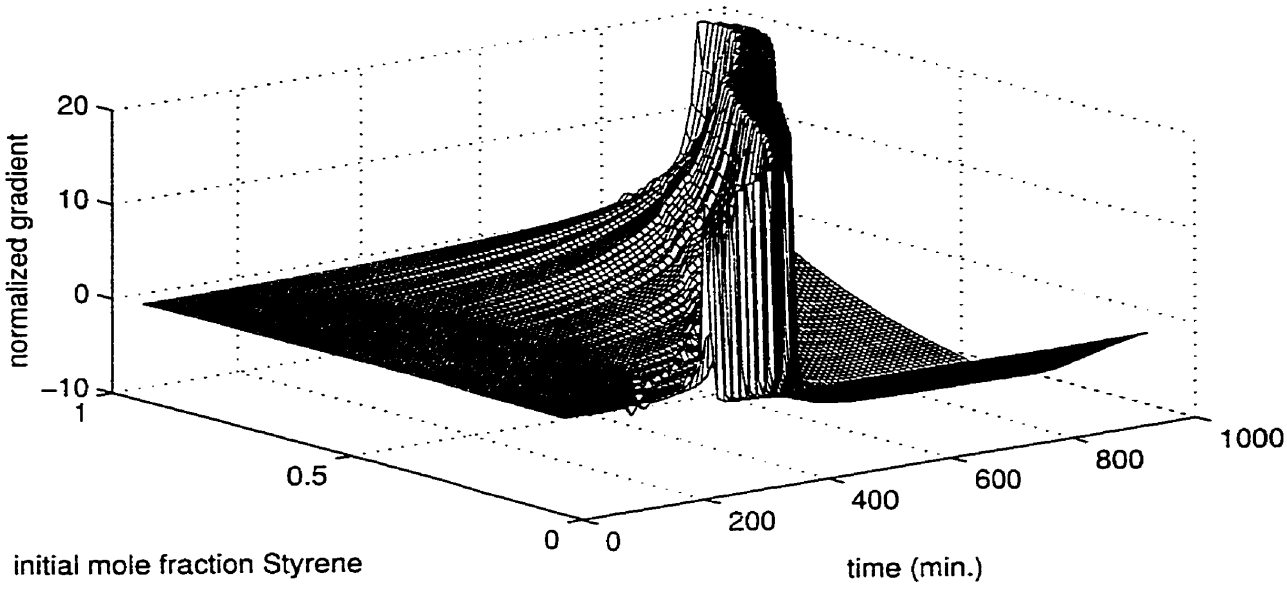


Figure 6.21: Gradient plots with respect to the parameter $\alpha_{m\ M1}$



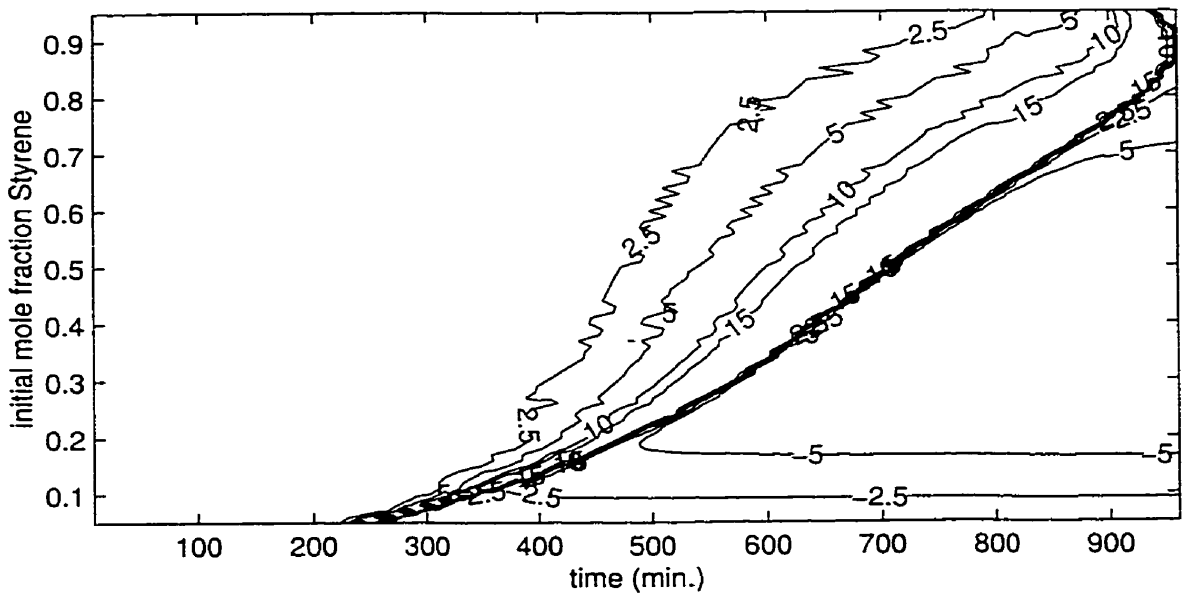


Figure 6.22: Gradient plots with respect to the parameter $Tg_{M_1M_2}$

plot. A further observation is that the absolute gradient values are largest in the area, as seen in the contour plot, above the valley where the initial feed is larger than 0.5 and the sampling time is greater than 700 minutes.

The gradient surface for the parameter $Tg_{M_1M_2}$ is shown in Figure 6.22. An examination of this figure reveals a ridge (where the absolute gradient values are maximized) that runs diagonally from a feed concentration of 0.1 and a time of 300 minutes to a feed concentration of 0.8 and a time of 900 minutes. A further observation, as seen in the contour plot, is that to the right of the ridge there is a valley, from an initial feed of 0.3 to 0.5 and a sampling time greater than 700 minutes, where the gradient values are large.

In both of the above sensitivity analysis there is an area of large gradient values that occur from a feed concentration of 0.1 and a time of 300 minutes to a feed concentration of 0.8 and a time of 900 minutes. This corresponds to the area of autoacceleration for the copolymerization carried out at different feed concentrations as shown in Figure 6.23. In this surface plot the vertical axis is conversion, where 1.0 indicates 100% conversion, and the horizontal axes are time and initial feed composition. The auto-acceleration section is the steep cliff face of the conversion versus time surface, from a composition of 0.1 and a time of 300 minutes to a feed composition of 0.9 and a time of 900 minutes.

A summary of the information obtained from the gradient plots is given in Table 6.21. In this table the comments column is a summary of the observability of each parameter, where any areas of large or small gradient values (i.e. good or poor observability) are identified. As an example, at the top of Table 6.21, for the parameter $\alpha_{m\ M1}$ and the response of conversion the comments column states 'ridge from 0.1 feed and 300 min. to 0.8 feed and 900 min., the ridge widens as the feed increases'. This area corresponds to the area of large gradient values as shown in Figure 6.21.

In using the information in Table 6.21 to generate an initial guess for the experiment design algorithm the best general area for each parameter was identified. This is shown in Table 6.22.

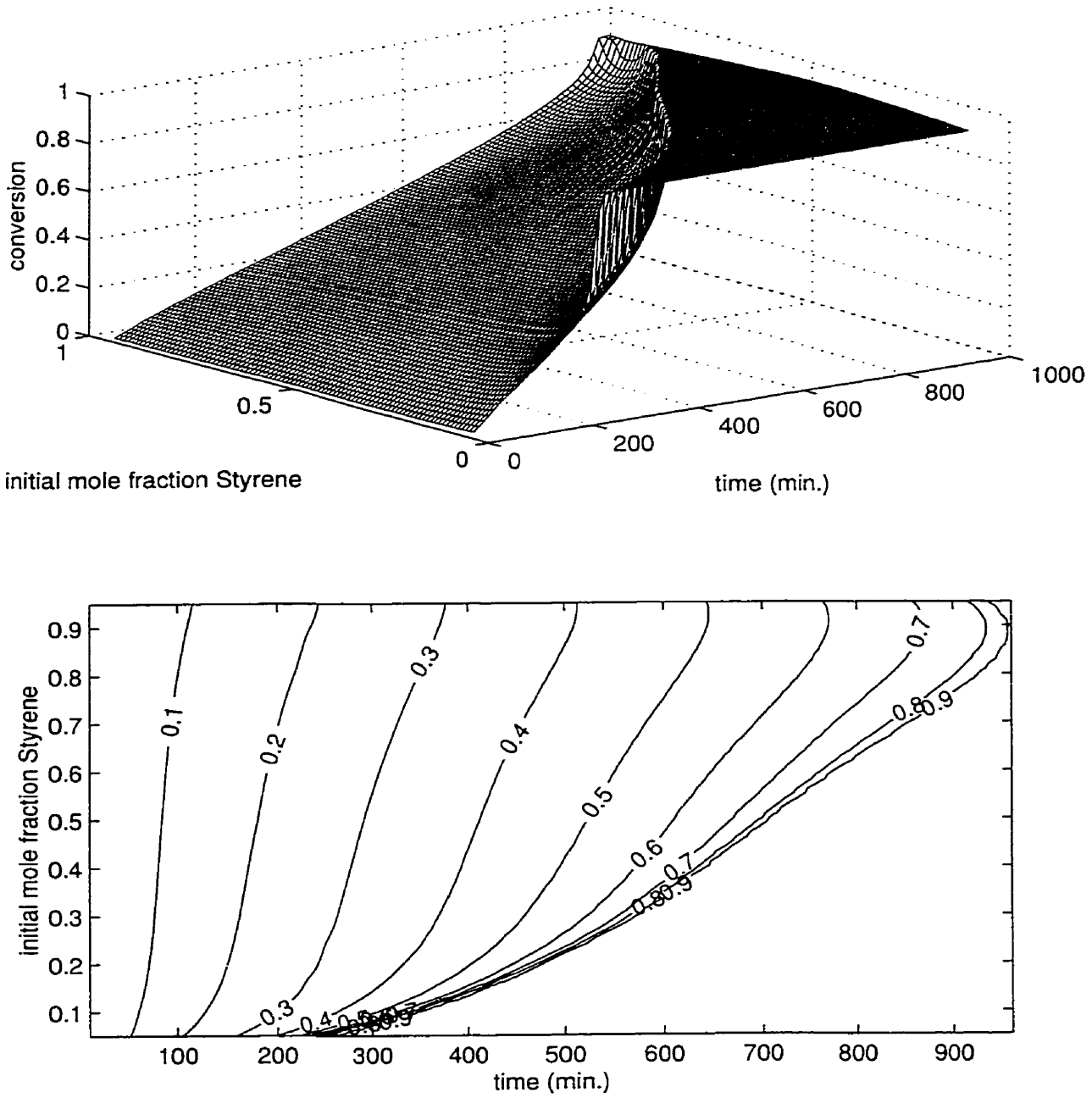


Figure 6.23: Plot of conversion versus time for the copolymerization of Styrene / Methyl Methacrylate

Parameter	Response	Comments
$\alpha_{m\ M1}$	conversion	ridge from 0.1 feed and 300 min. to 0.8 feed and 900 min., the ridge widens as the feed increases
	composition	sharp ridge from 0.2 feed and 500 min. to 0.6 feed and 800 min.
	M_n	ridge from 0.1 feed and 300 min. to 0.8 feed and 900 min., the ridge widens as the feed increases
	M_w	ridge from 0.1 feed and 300 min. to 0.8 feed and 900 min., the ridge widens as the feed increases and values are larger after the ridge (i.e. larger time)
	rate	ridge from 0.1 feed and 300 min. to 0.8 feed and 900 min., the ridge widens as the feed increases and values are larger before the ridge (i.e. smaller time)
	radical conc.	a rapid increase in values from 400 to 600 min. with very large values at time > 600 min for all feeds
	seq. length $M1$	sharp ridge from 0.3 feed and 600 min. to 0.8 feed and 900 min., poor observability elsewhere
	seq. length $M1M1$	same as for seq. length $M1$
	seq. length $M1M1M1$	same as for seq. length $M1$
	seq. length $M2$	sharp ridge from 0.2 feed and 500 min. to 0.6 feed and 800 min., poor observability elsewhere
	seq. length $M2M2$	sharp ridge from 0.3 feed and 600 min. to 0.7 feed and 900 min., poor observability elsewhere
	seq. length $M2M2M2$	sharp ridge from 0.3 feed and 600 min. to 0.6 feed and 800 min., poor observability elsewhere
$\alpha_{m\ M2}$	conversion	ridge from 0.1 feed and 300 min. to 0.8 feed and 900 min.
	composition	sharp ridge from 0.2 feed and 500 min. to 0.5 feed and 700 min.
	M_n	ridge from 0.1 feed and 300 min. to 0.8 feed and 900 min., values are larger with smaller feed values
	M_w	ridge from 0.1 feed and 300 min. to 0.8 feed and 900 min., the ridge widens as the feed increases and values are larger after the ridge (i.e. larger time)
	rate	ridge from 0.1 feed and 300 min. to 0.8 feed and 900 min., the ridge widens as the feed increases and values are larger before the ridge (i.e. smaller time)
	radical conc.	a rapid increase in values along the diagonal band of 0.1 feed and 200 min to 0.9 feed and 800 min. with very large values to the right of the band (i.e. larger time)
	seq. length $M1$	sharp ridge from 0.3 feed and 600 min. to 0.8 feed and 900 min., poor observability elsewhere
	seq. length $M1M1$	same as for seq. length $M1$
	seq. length $M1M1M1$	same as for seq. length $M1$
	seq. length $M2$	sharp ridge from 0.2 feed and 500 min. to 0.6 feed and 800 min., poor observability elsewhere
	seq. length $M2M2$	sharp ridge from 0.3 feed and 600 min. to 0.7 feed and 900 min., poor observability elsewhere
	seq. length $M2M2M2$	sharp ridge from 0.3 feed and 600 min. to 0.6 feed and 800 min., poor observability elsewhere

Table 6.21: Summary of the sensitivity analysis

Parameter	Response	Comments
BBm_{M1}	conversion	a band from 0.1 feed and 400 min. to 0.8 feed and 900 min., where there is no observability to the left (i.e. smaller time) and poor observability to the right (i.e. larger time)
	composition	same as conversion except very poor observability to the right of the ridge
	M_n	same as composition
	M_w	same as conversion except very marginal observability to the right of the ridge
	rate	very sharp ridge from 0.3 feed and 600 min. to 0.8 feed and 900 min., poor observability elsewhere
	radical conc.	a ridge from 0.1 feed and 400 min. to 0.8 feed and 900 min., with good observability to the right (i.e. larger time)
	seq. length $M1$	an increase in observability to the right of the line from 0.2 feed and 500 min. to 0.8 feed and 900 min., no observability to the left of the line , with a peak at 0.4 feed and 900 min
	seq. length $M1M1$	same as seq. length $M1$, except two smaller peaks at 0.3 feed and 900 min and 0.6 feed and 900 min
	seq. length $M1M1M1$	same as seq. length $M1$, except a smaller peak at 0.4 feed and 900 min and 0.7 feed and 900 min
	seq. length $M2$	an increase in observability to the right of the line from 0.1 feed and 400 min. to 0.7 feed and 900 min., no observability to the left of the line , with a peak at 0.3 feed and 900 min
	seq. length $M2M2$	same as seq. length $M2$, except two smaller peaks at 0.45 feed and 900 min
	seq. length $M2M2M2$	an increase in observability to the right of the line from 0.1 feed and 350 min. to 0.6 feed and 800 min., no observability to the left of the line , with a peak at 0.35 feed and 900 min
BBm_{M2}	conversion	very sharp small ridge from 0.1 feed and 400 min. to 0.9 feed and 900 min., poor observability in general
	composition	same as conversion
	M_n	same as conversion
	M_w	same as conversion except marginal observability to the right of the ridge and at low feed
	rate	same as conversion
	radical conc.	very sharp ridge from 0.1 feed and 400 min. to 0.9 feed and 900 min., with good observability to the right (i.e. larger time)
	seq. length $M1$	poor observability to the right of the line from 0.3 feed and 600 min. to 0.9 feed and 900 min. with a peak at 0.3 feed and 900 min , no observability to the left of the line ,
	seq. length $M1M1$	same as seq. length $M1$, except peak at 0.2 feed and 900 min
	seq. length $M1M1M1$	same as seq. length $M1$, except peak at 0.3 feed and 900 min
	seq. length $M2$	an increase in observability to the right of the line from 0.1 feed and 400 min. to 0.5 feed and 700 min. with a peak at 0.2 feed and 900 min , no observability to the left of the line ,
	seq. length $M2M2$	same as seq. length $M2$, except two smaller peaks at 0.45 feed and 900 min
	seq. length $M2M2M2$	a small ridge at 0.1 feed and 500 to 900 min with poor observability in general

Table 6.21 Cont'd. : Summary of the sensitivity analysis

Parameter	Response	Comments
$Tg_{M_1M_2}$	conversion	wide ridge from 0.1 feed and 400 min. to 0.8 feed and 900 min., a valley to the right of the ridge with a max. observability at 0.4 feed and 900 min
	composition	sharp ridge from 0.1 feed and 400 min. to 0.8 feed and 900 min., poor observability elsewhere
	M_n	ridge from 0.1 feed and 400 min. to 0.8 feed and 900 min., a small valley to the right of the ridge with an max. observability at 0.4 feed and 900 min
	M_w	ridge from 0.1 feed and 400 min. to 0.9 feed and 800 min., good observability to the right of the ridge with a peak at 0.6 feed and 900 min.
	rate	ridge from 0.1 feed and 400 min. to 0.9 feed and 900 min., the ridge widens as the feed increases
	radical conc.	a rapid increase in values in a band from 0.1 feed and 400 min to 0.9 and 700 min. with very large values to the right of the ridge
	seq. length $M1$	sharp ridge from 0.2 feed and 500 min. to 0.8 feed and 900 min., poor observability to the left, a mound to the left with a peak at 0.4 feed and 900 min.
	seq. length $M1M1$	same as $M1$ except two smaller peaks at 0.6 feed and 900min and 0.3 feed and 900 min
	seq. length $M1M1M1$	same as $M1$ except a smaller peak at 0.4 feed and 900min
	seq. length $M2$	sharp ridge from 0.1 feed and 400 min. to 0.7 feed and 800 min., poor observability to the left,a hollow to the right with a peak at 0.4 feed and 900 min
	seq. length $M2M2$	same as $M1$ except a smaller peak at 0.4 feed and 900min
	seq. length $M2M2M2$	same as $M1$ except a smaller peak at 0.4 feed and 900min

Table 6.21 Cont'd. : Summary of the sensitivity analysis

Parameter	Area of best observability
$\alpha_{m\ M1}$	feed > 0.6 time > 700 min
$\alpha_{m\ M2}$	feed < 0.4 time > 600 min
BBm_{M1}	feed < 0.5 time > 700 min
BBm_{M2}	feed < 0.4 time > 700 min
$Tg_{M_1M_2}$	feed > 0.5 time > 700 min

Table 6.22: General areas of best observability

Therefore based on this information the following 6 trial experiment or initial point for the experiment design is obtained; a feed of 0.3 mole fraction Styrene and a sampling time of 750, 800 and 850 minutes, and a feed of 0.8 mole fraction Styrene and a sampling time of 750, 800 and 850 minutes. A conservative approach was taken in the determination of areas of best observability in that the sharp peaks in the gradient plots were not considered. The author has found that sharp peaks in the gradient plots can be sensitive to the parameter values (i.e. their location can change with a change in the parameter values).

6.5.3 Experiment Design

Initially the size of the experiment to design was chosen to be 12 trials, but with 12 trials the optimization was found to be too slow to converge. To speed up convergence the design problem was simplified by reducing the experiment to 6 trials, that will be replicated to obtain a 12 or 18 trial experiment. This simplification decreased the optimization from 24 to 12 variables and significantly reduced the computation required.

The best experiment found using the optimization routine is given in Table 6.23 along with an experiment that was designed by the author from an inspection of the gradient plots. The design criterion values as well as the confidence region volume radius for each experiment are given. The radius is an approximation used to compare the experiments and is described in Chapter 3. A comparison of the radii indicates that the designed experiment is much better than the experiment designed by inspection.

An examination of the designed experiment reveals that the two points at a feed of 0.68 and 0.70 mol fraction Styrene occur very near the autoacceleration region, which corresponds to a sharp peak in the gradient values for most of the responses. The experiment sensitivity to these two points can be tested by moving the sampling time at these two compositions. If the sampling time for the two points is set to 700 minutes, the criterion decreases considerably and a radius of 1.28e-4 is obtained, which is much closer

	trials		criterion value	radius
	Sty feed (mol fr.)	time (min.)		
designed experiment	0.36	701	1.9478e52	5.9026e-6
	0.37	821		
	0.24	857		
	0.89	673		
	0.70	838		
	0.68	873		
inspection experiment	0.30	750	9.4773e33	4.0025e-4
	0.30	800		
	0.30	850		
	0.80	750		
	0.80	800		
	0.80	850		

Table 6.23: Designed experiments for case study 3

to the experiment designed by inspection. To further test the experiment sensitivity an induction time of 100 minutes was assumed. This would result in the autoacceleration occurring 100 minutes later than predicted and effectively decreasing all of the sampling times by 100 minutes. The criterion of this experiment also decreases considerable and a radius of 3.34e-5 is obtained. This drop in quality reveals the sensitivity of the design experiment to the location of the points near the autoacceleration curve. Since sample points near this curve may be difficult to obtain, the designed experiment may not be as good in application as its criterion value (or radius) would indicate. The effect of induction time on the inspection experiment was also examined. With an induction time of 100 minutes, the criterion value increased and a radius of 1.71e-4 was obtained, resulting in a better experiment.

6.5.4 Parameter Estimation

The parameters were estimated with the responses listed in Table 6.20 and the experiment designed by inspection given in Table 6.23. This experiment design was used because it

parameter	estimate	true value
$\alpha_{m\ M1}$	0.000994	0.001
$\alpha_{m\ M2}$	0.001018	0.001
BBm_{M1}	1.0052	1.00
BBm_{M2}	1.0506	1.00
$Tg_{M_1M_2}$	362.6	363

Table 6.24: Estimated parameter values

cause it is computationaly faster since for each criterion evaluation the simulation needs to be run at only two initial feed compositions versus six in the optimal experiment (i.e. this design is three time faster to simulate). As in the previous cases local optima were found to be a problem and the simulated annealing algorithm was found to work well, though it was very expensive computationally (10 hours on a Pentium III, 500). The parameter estimates obtained and the true values are given in Table 6.24.

The effect of using different responses and different numbers of trials on the quality of the parameter estimates was investigated. To perform this investigation the estimated parameter confidence regions obtained using different conditions cannot be used. Because as shown in Chapter 5, these confidence regions can vary significantly with the same experiment design but different sets of parameter data. Therefore a Monte Carlo approach is required to deal with the variance in the confidence regions.

The Monte Carlo study involved estimating the parameters using different responses and number of trials. Ten master sample data sets were generated that contained all of the responses used and 18 trials. The sample data sets used in each run were chosen as a subset of the master sample data sets. The sample data for each run was generated in this manner so that each of the runs would be identical in their area of overlap. By overlap it is meant that if an 8 and 12 trial experiment are compared, the 12 trial experiment will be the same as the 8 trial experiment plus 4 additional trials. Therefore , any differences

run name	number of trials	responses used†
case-1	18	1 2 3 4 5 6
case-2	18	1 2 3 4 5 6 7
case-3	12	1 2 3 4 5 6
case-4	12	1 2 3 4 5 6 7
case-5	12	1 2 3 6 7
case-6	8	1 2 3 4 5 6
case-7	8	1 2 3 4 5 6 7

† 1 conversion, 2 number average molecular weight, 3 weight average molecular weight, 4 polymerization rate, 5 radical concentration, 6 copolymer composition, 7 triad fractions

Table 6.25: Description of the runs used in comparing the effect of different responses and/or number of trials on the parameter estimates

in the parameter estimates obtained would be due to the extra responses or trials and not due to different error values within common sample points. The parameters in each case were estimated using the `fmins` optimization function and with the true parameter values as the initial values for the optimization. The true parameter values were used to overcome the problem of local optima and to speed up the estimation procedure.

The different cases that were considered are listed in Table 6.25. This table gives the case name, the number of trials in the case and the responses used. The responses are coded and an explanation of the response code value is provided at the bottom of the table.

The experiments used in the runs are based on the experiment designed by inspection and given in Table 6.23. For the 12 and 18 trial experiments the experiment was replicated two and three times respectively. For the 8 trial experiment the two trials at 800 minutes were replicated. In each of the Monte Carlo runs, 10 sets of parameter estimates were

obtained. To analyze the results of the different runs, the parameter values obtained were plotted and a visual inspection was performed. These plots are shown in Figures 6.24, 6.25 and 6.26. In these figures four subplots are shown, where in each of the subplots the parameter point estimates of two parameters are plotted versus each other. The subplots are analogous to a joint confidence region plot of the two parameters. The analysis is carried out by comparing the distribution of the point estimates, i.e. tightly grouped versus spread out. The plots as used are a crude measure due to only 10 points being used per run. For a more refined analysis, a much larger number of data sets would have to be used, so that the true shape and true level marginal confidence regions of the parameters are obtained.

Figure 6.24 is a comparison of the case-1(o), case-3(*) and case-6(◇). These cases have 18, 12 and 8 trials respectively and include all of the responses except triad fractions. There appears to be no clear difference between the 18 and 12 trial runs (case-1 and case-3). This may be due to a saturation of information at 12 trials where the extra trials (trials 13 to 18) provide little additional information about the parameters, when compared to the initial 12. A comparison of case-3 and case-6, with 12 and 8 trial respectively, indicates that a difference does exist. This difference is most apparent in the lower left subplot of $\alpha_{m\ M1}$ versus $Tg_{M_1M_2}$. Therefore the addition of the extra four trials is worthwhile with the chosen experiment, as they will significantly decrease the joint confidence region of the parameter estimates.

The effect of adding the triad fraction response is shown in Figure 6.25. This figure is a plot of case-6 and case-7. Both cases have 8 trials and case-7 includes the triad fraction response. An inspection of the subplots reveals that only the estimates of the BBm_{M1} and BBm_{M2} parameters are affected, shown in the top right subplot. This was expected as the sensitivity analysis indicated that very little information was provided about the other parameters by the triad fraction response. The BBm parameter estimates improved significantly with the addition of the triad fraction response. The improvement can

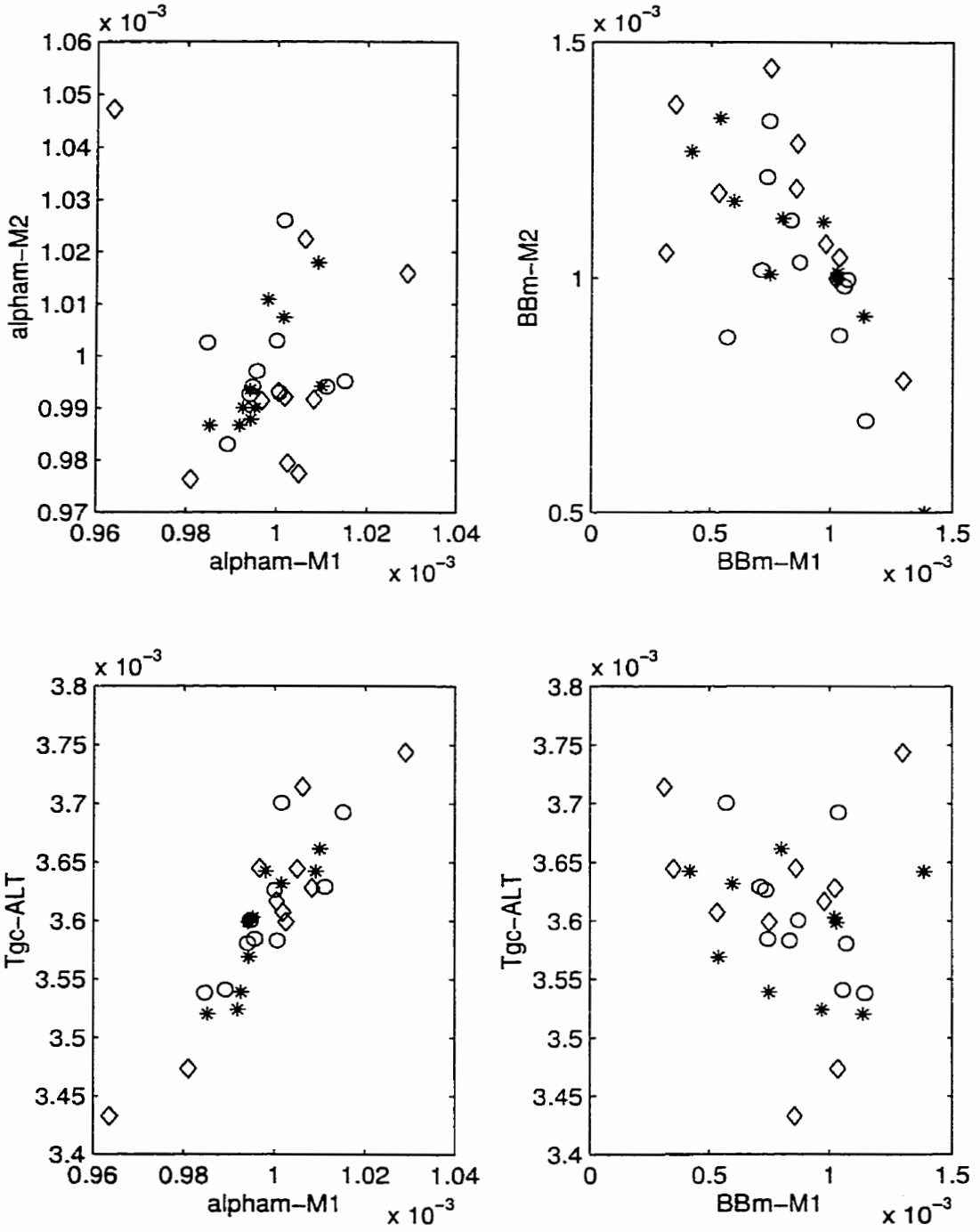


Figure 6.24: Examining the effect the number of trials has on the parameter estimates, with case-1(o) 18 trials, case-3(∗) 12 trials and case-6(⋄) 8 trials

be seen by the tightly grouped estimates for case-7 when compared to case-6. This improvement in the parameter estimates did not occur in the corresponding 12 and 18 trial runs, case-3 versus case-4 and case-1 versus case-2 respectively. Therefore, only if the number of trials is less than 12 will a benefit be obtained by the inclusion of the triad fraction response in the estimation of the BBm_{M1} and BBm_{M2} parameters.

Figure 6.26 is a plot of case-4 and case-5, where the effect of radical concentration and polymerization rate on the parameter estimates was examined. Both runs have 12 trials, case-4 has all of the responses and case-5 has all of the responses except radical concentration and polymerization rate. An inspection of the subplots indicates that the distribution of the parameter estimates is similar for both case-4 and case-5. Therefore the responses of radical concentration and rate did not improve the quality of the parameter estimates when the experiment has 12 trials. This was surprising as the sensitivity analysis indicated that a lot of information would be provided by these responses. Why this did not occur may be due to the experiment used or to the information saturation effect observed in the comparison of runs case-1, case-3 and case-6 (Figure 6.24).

The above examples have shown how parameter estimation simulation studies can be used to compare how different responses and different numbers of trials in the experiment will affect the quality of the parameter estimates.

6.5.5 Confidence Regions

The parameter estimate confidence regions are based on the parameter estimates obtained from an 18 trial experiment based on the design by inspection replicated three times. The parameter point estimates obtained are given in Table 6.26. The approximate level, true shape, conditional joint confidence regions for the parameter estimates were obtained using equation 5.13 from Chapter 5.

A sample of the conditional joint confidence regions of the estimated parameters are given in Figures 6.27 and 6.28. Figure 6.27 shows the 95% joint confidence region of the

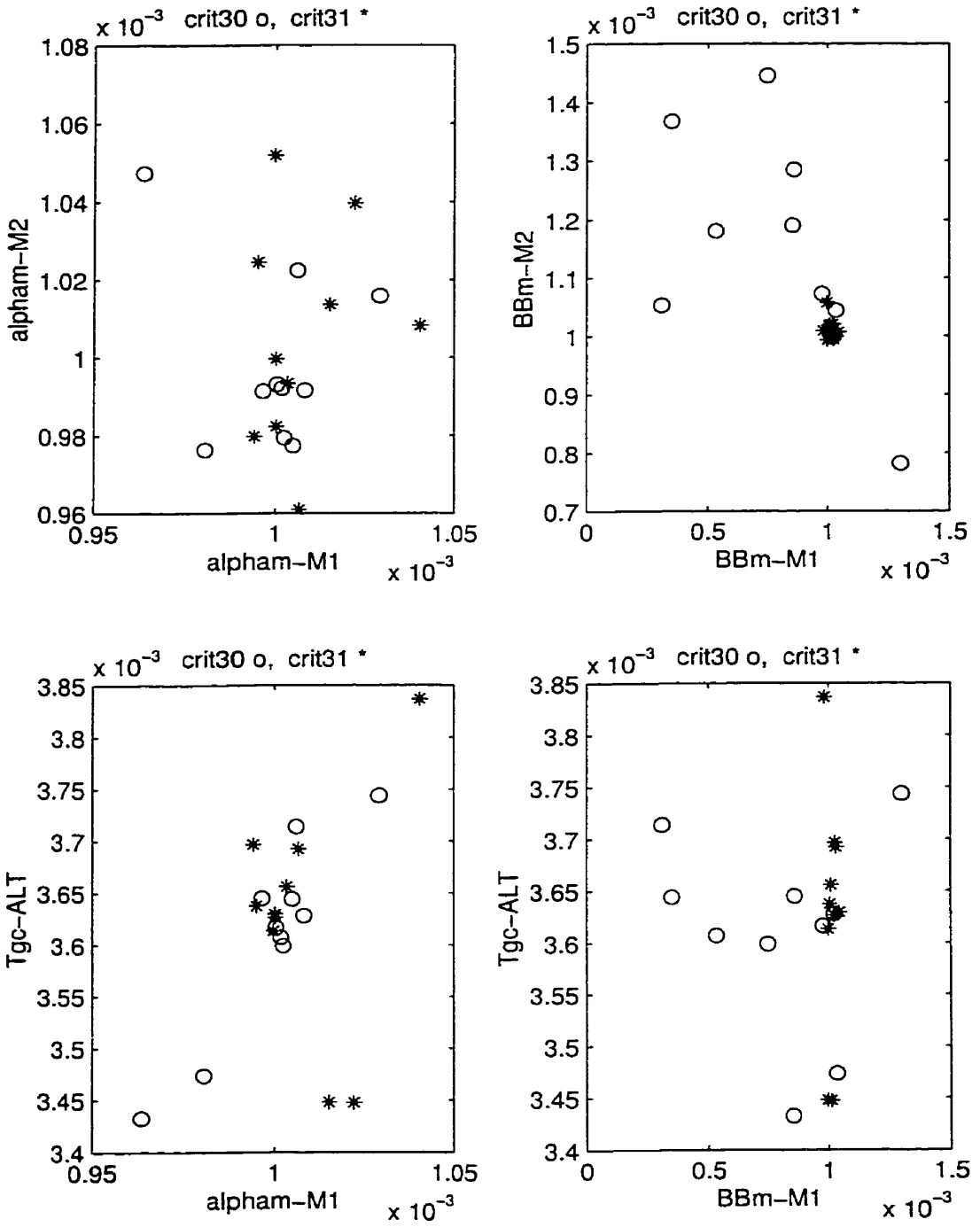


Figure 6.25: Examining the effect of including the triad fraction response on the parameter estimates, with case-6(circle) no triad fractions and case-7(∗) triad fractions

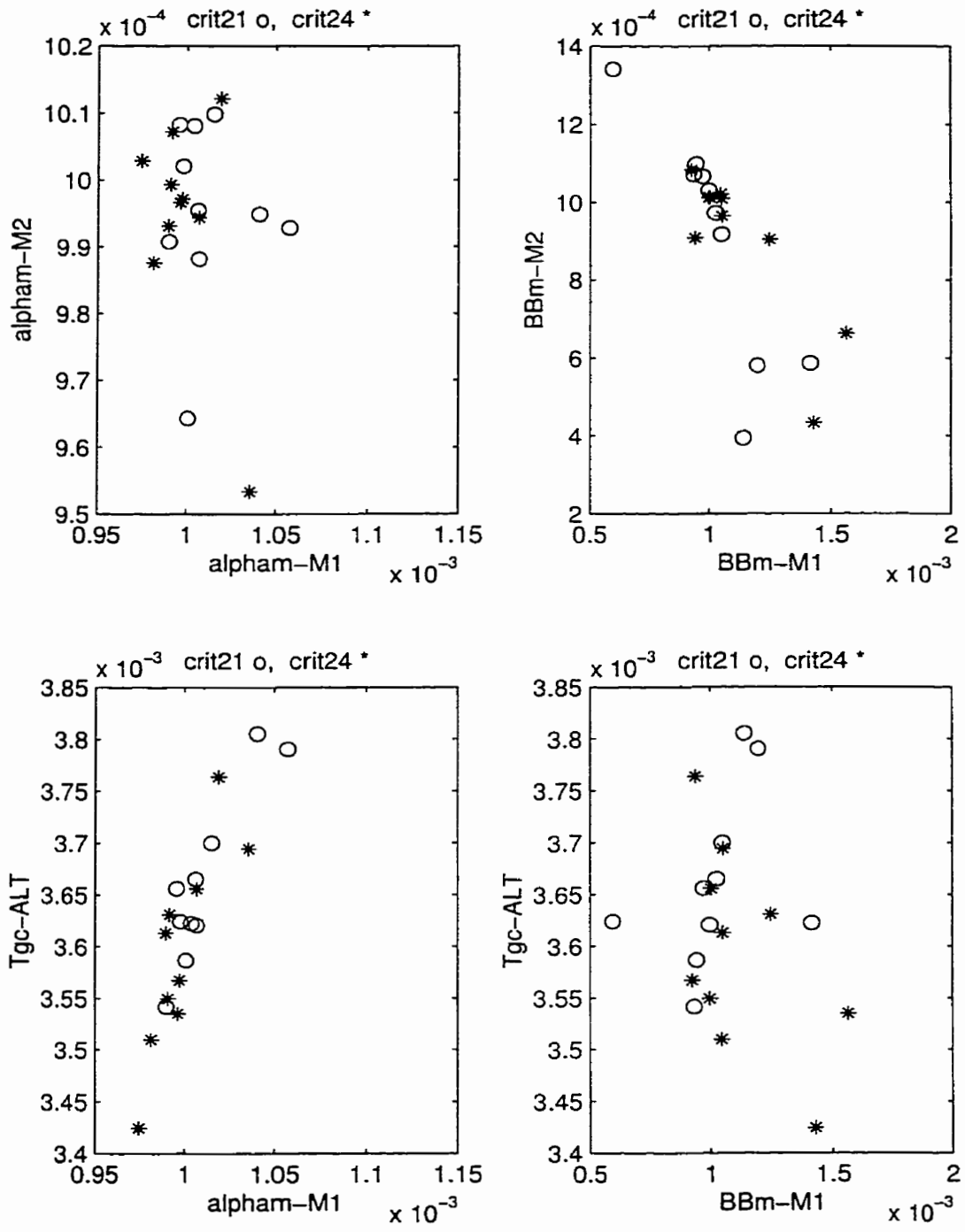


Figure 6.26: Examining the effect of including the radical concentration and polymerization rate on the parameter estimates, with case-4(circle) all responses and case-5(star) no radical concentration and polymerization rate responses

parameter	estimate	true value
$\alpha_{m\ M1}$	0.001006	0.001
$\alpha_{m\ M2}$	0.001010	0.001
BBm_{M1}	0.9183	1.00
BBm_{M2}	1.1606	1.00
$Tg_{M_1M_2}$	368.06	363

Table 6.26: Estimated parameter values

parameters $\alpha_{m\ M1}$ and $Tg_{M_1M_2}$, where the star indicates the point estimate. Figure 6.28 shows the 95% joint confidence region of the parameters BBm_{M1} and $Tg_{M_1M_2}$, where the star indicates the point estimate.

Figure 6.29 shows the 95% joint confidence region of the parameters $\alpha_{m\ M1}$ and $\alpha_{m\ M2}$ at different values of the $Tg_{M_1M_2}$ parameter. The star indicates the point estimate. The three confidence regions were obtained at values of 360.0, 368.1 and 375.0 for $Tg_{M_1M_2}$, and are labeled A, B and C, respectively. The three confidence regions are similar in shape but are shifted with respect to the $\alpha_{m\ M1}$ parameter. This indicates that the marginal confidence region for the $\alpha_{m\ M2}$ parameter is significantly larger than the conditional confidence region indicates. Surprisingly though, there is little shift of the confidence region with respect to the $\alpha_{m\ M2}$ parameter.

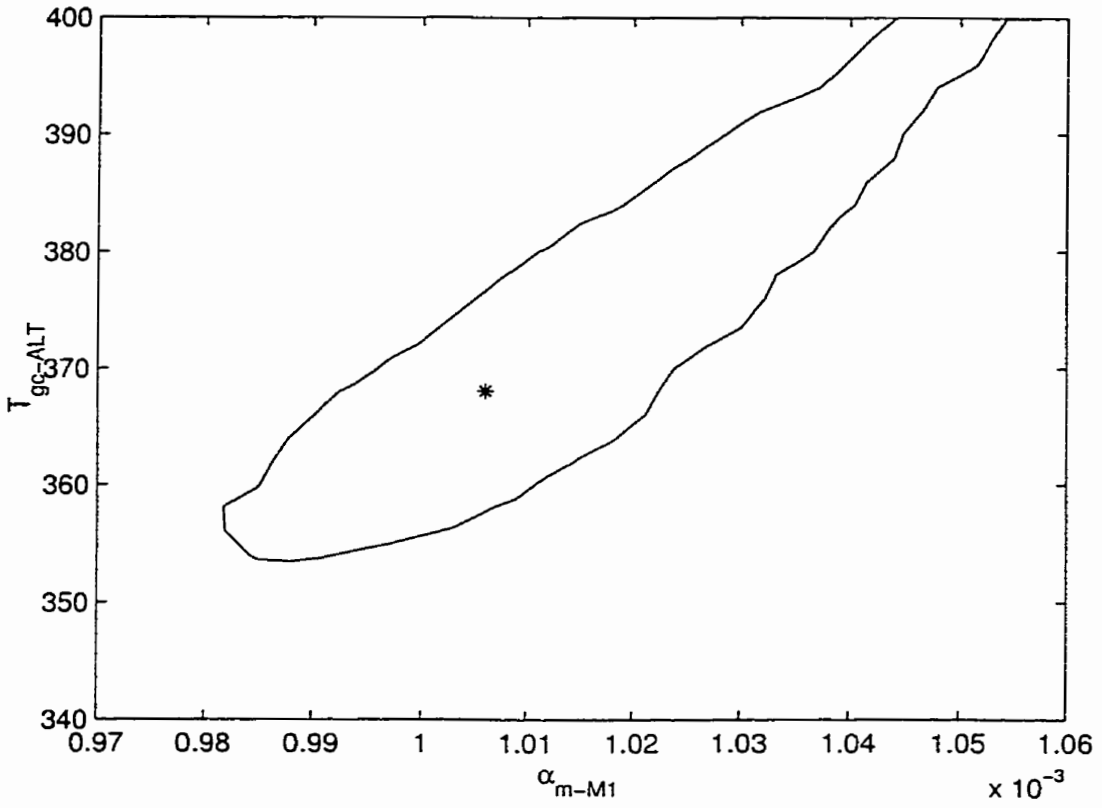


Figure 6.27: 95 percent confidence region for the parameters $\alpha_{m\ M1}$ and $Tg_{M_1M_2}$

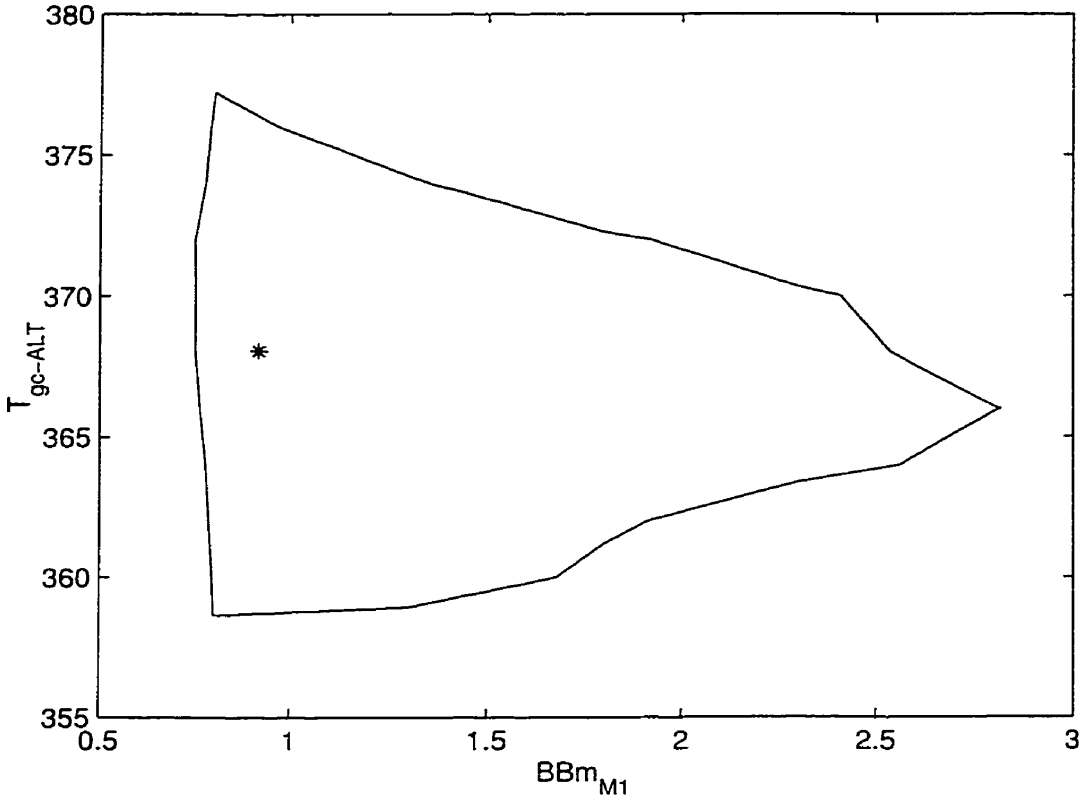


Figure 6.28: 95 percent confidence region for the parameters BBm_{M1} and $Tg_{M_1M_2}$

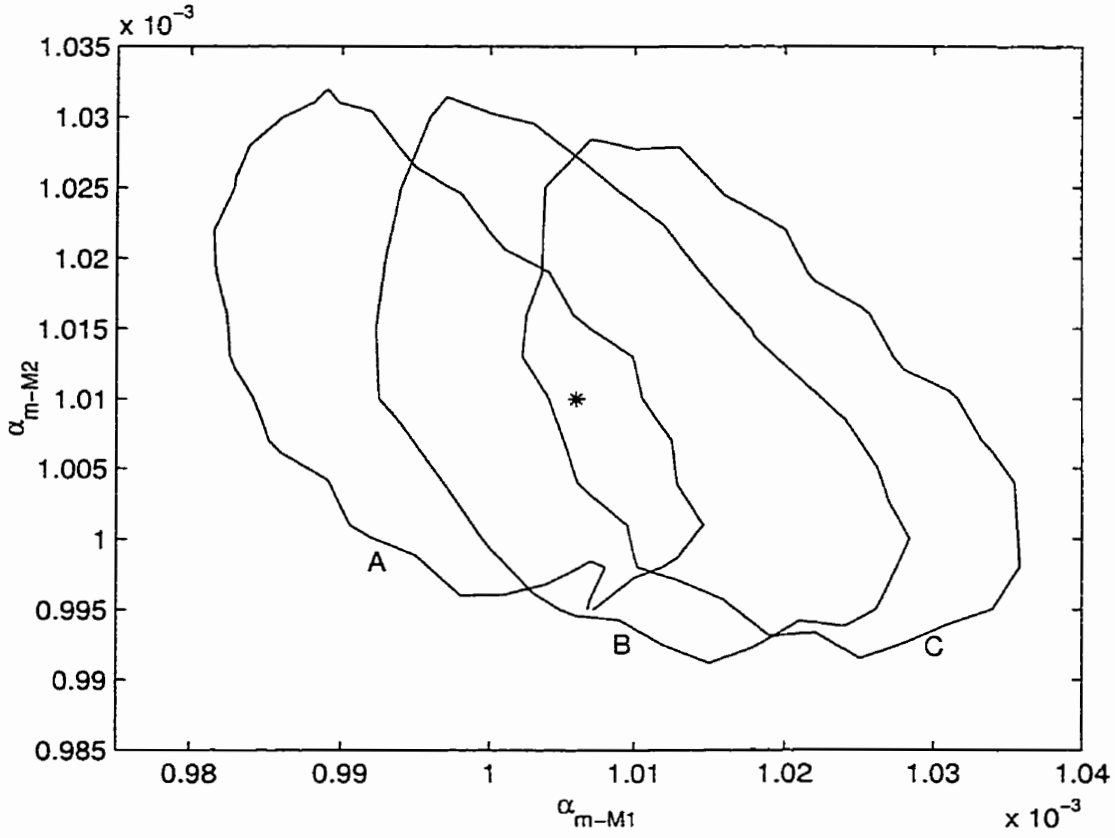


Figure 6.29: 95 percent confidence region for the parameters $\alpha_{m\ M1}$ and $\alpha_{m\ M1}$ at different values of $Tg_{M_1M_2}$, where $Tg_{M_1M_2}$ equals 360.0(A), 368.1(B) and 375.0(C)

Chapter 7

Concluding Remarks

7.1 Conclusions

In the previous chapters a protocol for the estimation of parameters in large dynamic models has been presented. Each of the four parameter estimation steps were discussed and three case studies were shown to illustrate the parameter estimation methodology. The major points of the parameter estimation protocol will be summarised in the paragraphs that follow.

In performing a sensitivity analysis it was shown how local sensitivity can be much more useful than a global measure due to the additional information that it provides. A limitation in using local sensitivities is the large volume of data that may be present. It was shown that by using gradient plots and normalizing the gradient values the analysis of local sensitivity measures is simplified. An added benefit of using the gradient plots is that they can provide guidance in the design of an optimal experiment.

The experiment design process is a difficult optimization problem that was shown to be plagued with local optima. It was found that if the initial guess in the optimization is based on the information provided by the gradient plots a very good experiment will be found. This approach was found to perform much better that multiple starting points

or using a robust optimization algorithm.

When designing experiments for nonlinear models the parameter values are required. Since these values are unknown and the values used will rarely be equal to the true values (in practice never), the design may be very poor for other values of the parameters. Therefore, how sensitive the design is to a change in the parameter values needs to be determined. A method was presented on how to perform this analysis so that the robustness of the design can be determined.

In estimating the parameters a significant problem was the number of local optima that were found. It was also noted that the number of local optima increased with an increase in the model complexity. Guidelines were presented to deal with this problem and the simulated annealing algorithm was found to work well.

Confidence regions are a measure of the uncertainty in the parameter estimates. Therefore they should be generated to obtain a measure of the quality of the parameter estimates. When more that two parameters are estimated, the confidence regions obtained are usually the conditional joint confidence region of two parameters. It was shown that these conditional confidence regions can underestimate the true marginal confidence regions of the parameter estimates.

Box and Draper (1965) showed that improved parameter estimates can be obtained if multiple responses are used. They showed this improvement by the significant decrease in the joint confidence regions that occurred with each extra response that was added. The author has found that the joint confidence regions generated when multiple responses are used and the sample size is small can be much smaller than the true joint confidence regions. A sample size is considered small if it is equal to $2p$, where p is the number of parameters. Therefore using the joint confidence regions in these circumstances will indicate that a much greater level of certainty in the parameters exists than is true. This does not imply that the parameter estimates are not improved with the use of extra responses. Through Monte Carlo studies it was found that use of multiple responses does

improve the parameter estimates, it is only the confidence regions that are an incorrect representation of the parameter uncertainty.

Since the true shape joint confidence regions for two parameters can be numerically difficult to obtain. The author has developed an algorithm that is both robust and efficient in obtaining the joint confidence regions.

The problem of parameter estimation in large models has been investigated and the above observations have been made. In performing the investigation a number of issues arose that were not addressed. These are outlined in the next section and can act as a starting point for further work in this area.

7.2 Future Work

In the course of this research a number of issues arose that were not addressed. These issues represent a starting point for an extension of the research presented in this thesis.

The D-optimality criterion for the design of experiments is a measure proportional to the volume of the linear joint confidence. It should be investigated if this is appropriate, not in the sense that it accurately represents the uncertainty in the parameters of a nonlinear model (as it is known that in most cases it does not), but whether it is proportional to it (therefore the best experiment will have the smallest linear joint confidence region and the smallest true joint confidence region). If this is true then the D-optimality criterion is appropriate even though it uses the linear joint confidence region. If it is not, then perhaps some measure can be developed to determine the difference between an experiment designed using linear versus true joint confidence regions.

In the design of experiments the physical cost of obtaining a given response was assumed to be equal for all responses. This in general in not true and the experiment design criterion should take this into account. A method was also presented to determine how sensitive the experiment design is to a perturbation of the parameter values. Therefore there is the potential to extend the experiment design criterion to a mixture

of criteria that would include the cost of a response as well as the experiment sensitivity to perturbations in the parameter values.

In the parameter estimation process the problems of local optima, conditional confidence regions, and the failure of the confidence region theory when multiple responses and a small sample size are used, were noted. A relatively new development, the Gibbs Sampler (Casella and George, 1992; Garcia-Cortes and Sorensen, 1996), is a promising approach to solving many of these issues. Although at the present time it requires further development in its implementation (so that it is more computationally efficient), to be applicable to the types of problems proposed in this thesis in particular on how to efficiently sample the conditional posterior distributions.

It was shown that the joint confidence regions can be incorrect when using multiple responses and the sample size is less than $2p$ where p is the number of parameters. Based on this observation the following questions present themselves.

- Is the model prediction confidence region also incorrect if multiple responses are used and if so, to what degree?
- How is the experiment design process affected, since it is based on the linearized joint confidence region volume of the parameter estimates?
- How is the model discrimination process affected, since most model discrimination methods are a function of the model prediction confidence region or the parameter estimates joint confidence region?

Appendix A

Nomenclature

$A111, A112, A212$	cumulative triad fraction of monomer-1-centered triads
$A222, A221, A121$	cumulative triad fraction of monomer-2-centered triads
A_{ijk}	instantaneous fraction of monomer-j-centred triads with sequence ijk
BBm	rate of decrease of k_p with free volume of monomer (L/mol min per free volume unit)
C	normalizing constant
E_1, E_2	activation energies in the ABC model
f_1, f_2	mole fractions of monomer 1 and 2 in the co-monomer feed
$F1$	copolymer composition, mole fraction of monomer one
f	initiator efficiency
f(x, θ)	a nonlinear model
$f_{act-eng}$	activation energy in the Arrhenius expression for f, initiator efficiency
$f_{pre-exp}$	pre-exponential factor in the Arrhenius expression for f, initiator efficiency
I	an initiator molecule
k_1, k_2	reaction rate constants in the ABC model
k_d	initiator decomposition rate constant (s^{-1})
k_{fm}	overall rate constant for transfer to monomer (L/(mol s))
$k_{fm\ \ act-eng}$	activation energy in the Arrhenius expression for k_{fm}

$k_{fm\ \ pre-exp}$	pre-exponential factor in the Arrhenius expression for k_{fm}
k_p	overall propagation rate constant (L/(mol s))
$k_{p\ act-eng}$	activation energy in the Arrhenius expression for the rate of propagation
$k_{p\ pre-exp}$	pre-exponential factor in the Arrhenius expression for the rate of propagation
k_t	overall termination rate constant (L/(mol s))
M_i	monomer i
M_n	number average molecular weight
M_w	weight average molecular weight
n	number of trials
N	number of sample points
N_I	moles of initiator
N_p	moles of monomer bound in the copolymer
p	number of parameters
P_n	polymer molecule containing n monomer units
Q_0, Q_1, Q_2	first three moments of the polymer molecular weight distribution
$r_{11}, r_{21}, r_{12}, r_{22}$	monomer reactivity ratios in the penultimate model
r_1, r_2	reactivity ratios
R	ideal gas constant
$[R\bullet]$	concentration of radicals (mol/L)
$R_{i,\,j}$	a radical of chain length i, ending in monomer j
$R_{in}\bullet$	a primary radical formed by the decomposition of the initiator
$R_n\bullet$	a free radical chain containing n monomer units
s_1, s_2	radical reactivity ratios in the penultimate model
s^2	estimate of the variance
$S(\hat{\beta})$	residual sum of squares at $\hat{\beta}$
$S(\hat{\theta})$	residual sum of squares at $\hat{\theta}$
T	temperature
Tg_{M_i}	glass transition temperature of monomer i (K)
$Tg_{M_iM_j}$	glass transition temperature of the alternating copolymer of monomer i and monomer j (K)

Tgp glass transition temperature of the polymer (K)

v_{ij} sum of the product of the residuals of responses i and j

$\mathcal{V}$ matrix of the elements v_{ij}

V matrix of gradient values of the response with respect to the parameters evaluated at the trial points in the experiment design; and volume (L) in the polymerization model in section 6.2

V_f total free volume

$V_{f_{crit}}$ critical free volume for propagation

V_{fom} constant of free volume with temperature for monomer (free volume units/K)

V_{fop} constant of free volume with temperature for polymer (free volume units/K)

V_{F_m} free volume of monomer

V_m volume of monomer

V_p volume of polymer

V_T volume of reacting mixture

W diagonal matrix of weights used in the MWLS criterion

W_{M_i} weight fraction of monomer i

X matrix of regressor variables, or the design matrix

x independent variable in a model

Y vector of responses

y_i model response

Z matrix of residuals

Greek Letters

α significance level

α_m variation of free volume with temperature for monomer (free volume units/K)

β a parameter in a linear model, and a parameter in the calculation of the moments of the molecular weight distribution in the polymerization model section 6.2

γ	the proportion of free radical chains terminating by disproportionation
τ	a parameter in the calculation of the moments of the molecular weight distribution
θ_i	parameter in a nonlinear model
Σ	covariance matrix
σ^2	variance
σ_{ij}	ij-th element of the response covariance matrix
σ^{ij}	ij-th element of the inverse of the response covariance matrix
σ_{resp}	standard deviation of the response measurement error
ϵ	measurement error
Ω	hessian of the objective function with respect to the parameters
$\phi_i\bullet$	mole fraction of radical type i

Superscripts

Superscripts indicate the condition of the variable.

$\hat{}$	an estimate based on sample data
$\bar{}$	an initial estimate based on prior knowledge
*	a true value
M_i	regarding monomer i

Appendix B

The following is a listing of the simulation options used with the Watpoly model for the examples in Chapter 2 and 3.

Initial temperature	65 C
Heat transfer parameter	1.00 Cal/K min
Temperature Case	Isothermal
Solution end time	960 min.
Numerical solution spacing	5 min
Induction time	0 min
Conversion limit	0.999
Tolerance parameter	4
Diffusion controlled propagation	[x]
Diffusion controlled termination	[x]
Residual termination model	[RNG]
SSH for radicals	[x]
Polymer Tg	Johnston

Appendix C

The following is a copy of the Matlab source code for the algorithm to obtain the joint confidence region contour. The algorithm is coded as a function where the contour level, parameter range, grid resolution, and name of the function that will return the criterion value are passed. The function can either return the data collected or plot the joint confidence region contour.

```
% ========================================================================
%
% contour_bf function
%
% ========================================================================
% algorithm to find the confidence contour of a set of parameter
% estimates, by finding the contour and then following it by fliiping
% a box of function evaluations in the appropriate dirrection
%
% []=contour_bf(x_range,y_range,point,contour_level,func_name)
%
% x_range = [x_min, x_step, x_max]  range and step size of first parameter
% y_range = [y_min, y_step, y_max]  range and step size of second parameter
% point = point estimate values of the two parameters
% contour_level = value of contour
% func_name = function that will return the criterion value where the two
%             parameter values are passed to it as a vector
%             [criterion]='func_name'([parameter 1, parameter 2])
%
%
% NOTE: this function requires the update_cmat m-file
%
% 00/01/14

% the location of the square as it moves along is defined by the
% top left hand corner (the square index point).  The other points
% in the square are always relative to the index point and in the
% orientation as shown below
%       1 2   (1 is the index point)
%       4 3



function []=contour_bf(x_range,y_range,point,contour_level,func_name)

%func_name='abc_crit_wolf';

%contour_level = 0.0001;
```

```
% parameter point estimate
%point(1)=.4 ;
%point(2)=.3 ;

% setup alorithm values, i.e. ranges, and step sizes
x_min = x_range(1) ;
x_max = x_range(3) ;
y_min = y_range(1) ;
y_max = y_range(3) ;

x_step = x_range(2) ;
y_step = y_range(2) ;


x_val = x_min:x_step:x_max;
y_val = y_min:y_step:y_max;

x_num = length(x_val);
y_num = length(y_val);

crit_mat = zeros(y_num,x_num);

x_mat = ones(y_num,1)*x_val;
y_mat = (ones(x_num,1)*y_val)';


% find location of square with the point estimate
box_x=ceil((point(1)-x_min)/x_step);
box_y=ceil((point(2)-y_min)/y_step);

% assign dummy values to starting box index
box_start=[0,0];

% get criterion value at box corners 1 and 2
crit_mat=update_cmat(1,box_x,box_y,x_val,y_val,crit_mat,func_name);
crit_mat=update_cmat(2,box_x,box_y,x_val,y_val,crit_mat,func_name);

% loop to move the box in the x-direction until we find the contour or plot edge
% =>assuming that the current crit_mat value is less than the contour_level
while (crit_mat(box_y,box_x+1) < contour_level) & (box_x < x_num-1)
box_x=box_x+1;
   crit_mat=update_cmat(2,box_x,box_y,x_val,y_val,crit_mat,func_name);
end
box_index = 2;

% get points 3 and 4 in the box on the contour or edge
crit_mat=update_cmat(3,box_x,box_y,x_val,y_val,crit_mat,func_name);
crit_mat=update_cmat(4,box_x,box_y,x_val,y_val,crit_mat,func_name);

% test if we are on a contour or edge
if (crit_mat(box_y,box_x) < contour_level) & ...
      (crit_mat(box_y,box_x+1) < contour_level) & ...
      (crit_mat(box_y+1,box_x+1) < contour_level) & ...
      (crit_mat(box_y+1,box_x) < contour_level)
   % on an edge
   edge = 1;
   movedir = 'down ';

elseif (crit_mat(box_y,box_x) > contour_level) & ...
      (crit_mat(box_y,box_x+1) > contour_level) & ...
      (crit_mat(box_y+1,box_x+1) > contour_level) & ...
```

```
      (crit_mat(box_y+1,box_x) > contour_level)
   % error! contour smaller than box step resolution
   disp('Error!  the contour is smaller than the resolution of the')
   disp('        chosen axis step size')
   return

else
   % on a contour
   edge = 0;
end

% main loop to move the box along the contour on the criterion surface
while 1
   box_move = 0;

   if edge == 0

   switch box_index
   case 1
      if crit_mat(box_y,box_x)>contour_level & crit_mat(box_y,box_x+1)<contour_level
         if box_y > 1
            % move box up and get points 1 and 2
            box_y=box_y-1;
            crit_mat=update_cmat(1,box_x,box_y,x_val,y_val,crit_mat,func_name);
            crit_mat=update_cmat(2,box_x,box_y,x_val,y_val,crit_mat,func_name);
            box_move = 1; % the box has moved
         else
            edge = 1;
            movedir = 'right';
         end
      else
         box_index=2;
      end

   case 2
      % move box right and get points 2 and 3
      if crit_mat(box_y,box_x+1)>contour_level & crit_mat(box_y+1,box_x+1)<contour_level
         if box_x < x_num-1
            % move box right and get points 2 and 3
            box_x=box_x+1;
            crit_mat=update_cmat(2,box_x,box_y,x_val,y_val,crit_mat,func_name);
            crit_mat=update_cmat(3,box_x,box_y,x_val,y_val,crit_mat,func_name);
            box_move = 1; % the box has moved
         else
            edge = 1;
            movedir = 'down ';
         end
      else
         box_index=3;
      end

   case 3
      if crit_mat(box_y+1,box_x+1)>contour_level & crit_mat(box_y+1,box_x)<contour_level
         if box_y < y_num-1
            % move box down and get points 3 and 4
            box_y=box_y+1;
            crit_mat=update_cmat(3,box_x,box_y,x_val,y_val,crit_mat,func_name);
            crit_mat=update_cmat(4,box_x,box_y,x_val,y_val,crit_mat,func_name);
            box_move = 1; % the box has moved
         else
            edge = 1;
            movedir = 'left ';
```

```
      end
   else
      box_index=4;
   end

case 4
   if crit_mat(box_y+1,box_x)>contour_level & crit_mat(box_y,box_x)<contour_level
      if box_x > 1
         % move box left and get points 1 and 4
         box_x=box_x-1;
         crit_mat=update_cmat(1,box_x,box_y,x_val,y_val,crit_mat,func_name);
         crit_mat=update_cmat(4,box_x,box_y,x_val,y_val,crit_mat,func_name);;
         box_move = 1; % the box has moved
      else
         edge = 1;
         movedir = 'up    ';
      end
   else
      box_index=1;
   end
end % end of switch box_index

else % move along the edge until we find the contour again
   switch movedir
      case 'right'
         if box_x < x_num-1
            box_x = box_x +1;
            crit_mat=update_cmat(2,box_x,box_y,x_val,y_val,crit_mat,func_name);
            box_move = 1; % the box has moved
            if crit_mat(1,box_x+1) > contour_level
               crit_mat=update_cmat(3,box_x,box_y,x_val,y_val,crit_mat,func_name);
               crit_mat=update_cmat(4,box_x,box_y,x_val,y_val,crit_mat,func_name);
               edge = 0;
               box_index = 2;
            end
         else
            movedir = 'down ';
         end

      case 'down '
         if box_y < y_num-1
            box_y = box_y +1;
            crit_mat=update_cmat(3,box_x,box_y,x_val,y_val,crit_mat,func_name);
            box_move = 1; % the box has moved
            if crit_mat(box_y+1,box_x+1) > contour_level
               crit_mat=update_cmat(1,box_x,box_y,x_val,y_val,crit_mat,func_name);
               crit_mat=update_cmat(4,box_x,box_y,x_val,y_val,crit_mat,func_name);
               edge = 0;
               box_index = 3;
            end
         else
            movedir = 'left ';
         end

      case 'left '
         if box_x > 1
            box_x = box_x - 1;
            crit_mat=update_cmat(4,box_x,box_y,x_val,y_val,crit_mat,func_name);
            box_move = 1; % the box has moved
            if crit_mat(box_y+1,box_x) > contour_level
               crit_mat=update_cmat(1,box_x,box_y,x_val,y_val,crit_mat,func_name);
               crit_mat=update_cmat(2,box_x,box_y,x_val,y_val,crit_mat,func_name);
```

```
                edge = 0;
                box_index = 4;
            end
         else
            movedir = 'up   ';
         end

      case 'up   '
         if box_y > 1
            box_y = box_y - 1;
            crit_mat=update_cmat(1,box_x,box_y,x_val,y_val,crit_mat,func_name);
            box_move = 1; % the box has moved
            if crit_mat(box_y,box_x) > contour_level
               crit_mat=update_cmat(2,box_x,box_y,x_val,y_val,crit_mat,func_name);
               crit_mat=update_cmat(3,box_x,box_y,x_val,y_val,crit_mat,func_name);
               edge = 0;
               box_index = 1;
            end
         else
            movedir = 'right';
         end

   end   %end of switch

end      %end of if edge == 0

if box_move == 1 &[box_x,box_y] == box_start
   % we are at the starting point
   break
elseif box_start == [0,0];
   % save starting point so we know when to stop
   box_start=[box_x,box_y];
end

%crit_mat


end   % end of main loop (i.e. while 1)

% get rid of zeros in the contour matrix
for i=1:y_num
   for j=1:x_num
      if crit_mat(i,j) == 0
         crit_mat(i,j) = NaN;
      end
   end
end

contour(x_mat,y_mat,crit_mat,[contour_level, contour_level])
hold on
plot(point(1),point(2),'*')
plot3(x_mat,y_mat,crit_mat,'r.')
hold off
```

```
% ================================================================
%
% updata_cmat function
%
% ================================================================
%
% function called by the contour_bf algorithm to get a point
% in the moving square
%
%

function [crit_mat]=update_cmat(box_point,box_x,box_y,x_val,y_val,crit_mat,func_name)

loud = 0;

switch box_point
case 1
   if crit_mat(box_y,box_x) == 0
      para_val=[x_val(box_x),y_val(box_y)];
    crit_mat(box_y,box_x) = eval([func_name,'(para_val)']);

      if loud == 1
      plot(box_x,box_y*(-1),'o')
      end
   end

case 2
   if crit_mat(box_y,box_x+1) == 0
      para_val=[x_val(box_x+1),y_val(box_y)];
    crit_mat(box_y,box_x+1) = eval([func_name,'(para_val)']);

      if loud == 1
      plot(box_x+1,box_y*(-1),'o')
      end
   end

case 3
   if crit_mat(box_y+1,box_x+1) == 0
      para_val=[x_val(box_x+1),y_val(box_y+1)];
    crit_mat(box_y+1,box_x+1) = eval([func_name,'(para_val)']);

      if loud == 1
      plot(box_x+1,(box_y+1)*(-1),'o')
      end
   end

case 4
   if crit_mat(box_y+1,box_x) == 0
      para_val=[x_val(box_x),y_val(box_y+1)];
    crit_mat(box_y+1,box_x) = eval([func_name,'(para_val)']);

      if loud == 1
      plot(box_x,(box_y+1)*(-1),'o')
      end
   end

end  % end of switch

return
```

Appendix D

All of the gradient plots obtained in the case studies are available as a package of encapsulated postscript files. Within this package there is a `readme.txt` text file that contains a listing and desctiption of all of the gradient plots contained within it.

A copy of the package can be obtained by contacting either Prof. T.A. Duever or Prof. A. Penlidis at the following address:

Department of Chemical Engineering
University of Waterloo
Waterloo, Ontario
Canada
N2L 3G1

Bibliography

[1] Andres T.H. (1997), "Sampling methods and sensitivity analysis for large parameter sets", *J. Statist. Comput. Simul.*, 57, 77-110.

[2] Atkinson A.C. (1996), "The usefulness of optimum experimental designs", *J. R. Statist. Soc. B*, 58(1), 59-76.

[3] Atkinson A.C. and Donev A.N. (1992), *Optimum Experimental Designs*, Oxford: Clarendon Press.

[4] Atkinson A.C. and Hunter W.G. (1968), "The design of experiments for parameter estimation", *Technometrics*, 10(2), 271-289.

[5] Bard Y. (1974), *Nonlinear parameter estimation*, New York: Academic Press.

[6] Bartus D.B. (1987), "Parameter Estimation in Dynamic Systems", Ph.D. thesis, The University of Michigan.

[7] Bates D.M. and Watts D.G. (1980), "Relative curvature measures of nonlinearity", *J. R. Statist. Soc. B*, 42(1), 1-25.

[8] Bates D.M. and Watts D.G. (1988), *Nonlinear Regression Analysis and its Applications*, Toronto: John Wiley & Sons.

[9] Beal S.L. and Sheiner L.B. (1988), "Heteroscedastic nonlinear regression", *Technometrics*, 30(3), 327-338.

[10] Beale E.M.L. (1960), "Confidence regions in non-linear estimation", *J. R. Statist. Soc. B*, 22, 41-88.

[11] Biegler L.T., Damiano J.J., and Blau G.E. (1986), "Nonlinear parameter estimation: A case study comparison", *AIChE Journal*, 32(1), 29-43.

[12] Bilardello P., Joulia X., LeLann J.M., Delmas H., and Koehret B. (1993), "A general strategy for parameter estimation in differential-algebraic systems", *Computers & Chemical Engineering*, 17(5/6), 517-525.

[13] Box G.E.P. and Cox D.R. (1964), "An analysis of transformations", *J. R. Statist. Soc. B*, 26(2), 211-252.

[14] Box G.E.P. and Draper N.R. (1965), "The Bayesian estimation of common parameters from several responses", *Biometrika*, 52(3/4), 355-365.

[15] Box G.E.P., Hunter W.G., MacGregor J.F., and Erjavec J. (1973), "Some problems associated with the analysis of multiresponse data", *Technometrics*, 15(1), 33-51.

[16] Box GEP and Jenkins GM. (1976), *Time series analysis: Forecasting and Control*, 2nd-ed, Holden-Day, San Francisco.

[17] Box G.E.P. and Lucas H.L. (1959), "Design of experiments in nonlinear situations", *Biometrika*, 46, 77-90.

[18] Box G.E.P. and Tiao G.C. (1973), *Bayesian inference in statistical analysis*, Toronto: Addison-Wesley.

[19] Caracotsios M. and Stewart W.E. (1985), "Sensitivity analysis of initial value problems with mixed ODE's and algebraic equations", *Computers & Chemical Engineering*, 9(4), 359-365.

[20] Casella G. and George E.I. (1992), "Explaining the Gibbs sampler", *The American Statistician*, 46(3), 167-174.

[21] Chappell M.J. and Godfrey K.R. (1992), "Structural identifiability of the parameters of a nonlinear batch reactor model", *Mathematical Biosciences*, 108, 241-251.

[22] Chappell M.J., Godfrey K.R., and Vajda S. (1990), "Global identifiability of the parameters of nonlinear systems with specified inputs: A comparison of methods", *Mathematical Biosciences*, 102, 41-73.

[23] Chaudhuri P. and Mykland P.A. (1993), "Nonlinear experiments: Optimal design and inference based on likelihood", *JASA, Theory and Methods*, 88(422), 538-546.

[24] Clarke G.P.Y. (1987), "Marginal curvatures and their usefulness in the analysis of nonlinear regression models", *JASA, Theory and Methods*, 82(399), 844-850.

[25] Cook R.D. and Goldberg M.L. (1986), "Curvatures for parameter subsets in nonlinear regression", *Annals of Statistics*, 14(4), 1399-1418.

[26] Cook R.D. and Witmer J.A. (1985), "A note on parameter-effects curvature", *JASA, Theory and Methods*, 80(392), 872-878.

[27] D'Agnillo L., Soares J.B.P., and Penlidis A. (1998), "A hierarchical data analysis of a replicate experiment in polyethylene synthesis with high-temperature gel permeation chromatography", *Polymer Reaction Engineering*, 7(2), 259-281.

[28] DeWit M.S. (1997), "Identification of the important parameters in thermal building simulation models", *J. Statist. Comput. Simul.*, 57, 305-320.

[29] Dhib R. and Oxby P. (1995), *A robust algorithm for determining contours*, internal report, University of Waterloo, Waterloo.

[30] Draper N.R. and Guttman I. (1995), "Confidence intervals versus regions", *The Statistician*, 44(3), 399-403.

[31] Draper N.R. and Hunter W.G. (1966), "Design of experiments for parameter estimation in multiresponse situations", *Biometrika*, 53(3/4), 525-533.

[32] Draper N.R. and Pukelsheim F. (1996), "An overview of design of experiments", *Statistical Papers*, 37, 1-32.

[33] Dube M.A. and Penlidis A. (1996), "Hierarchical data analysis of a replicate experiment in emulsion terpolymerization", *AIChE Journal*, 42(7), 1985-1994.

[34] Duever T.A. and Penlidis A. (1998), "Issues in nonlinear parameter estimation and multivariate model discrimination: Applications in polymer reaction engineering", *Appl. Math. and Comp. Sci.*, 8(4), 815-840

[35] Fletcher R. (1987), *Practical Methods of Optimization*, 2nd ed. Toronto: John Wiley & Sons.

[36] Ford I., Titterington D.M., and Kitsos C.P. (1989), "Recent advances in nonlinear experimental design", *Technometrics*, 31(1), 49-60.

[37] Gao J. (1999), "Advances in mathematical modelling of multicomponent free-radical polymerizations in bulk, solution and emulsion", PhD thesis, University of Waterloo.

[38] Gao J. and Penlidis A. (1996), "A comprehensive simulator/database package for reviewing free-radical homopolymerizations", *JMS Rev. Macromol. Chem. Phys.*, C36(2), 199-404.

[39] Gao J. and Penlidis A. (1998), "A comprehensive simulator/database package for reviewing free-radical copolymerizations", *JMS Rev. Macromol. Chem. Phys.*, C38(4), 651-780.

[40] Garcia-Cortes L.A. and Sorensen D. (1996), "On a multivariate implementation of the Gibbs sampler", *Genet. Sel. Evol.*, 28, 121-126.

[41] Gonin R. and Money A.H. (1985), "Nonlinear L_1-norm estimation: Part I: On the choice of the exponent, p, where the errors are additive", *Comm. in Statist. Part A, Theory and Methods*, 14, 827-840.

[42] Guay M. and McLean D.D. (1995), "Optimization and sensitivity analysis for multiresponse parameter estimation in systems of ordinary differential equations", *Computers & Chemical Engineering*, 19(12), 1271-1285.

[43] Hamilton DC and Watts DG. (1985), "A quadratic design criterion for precise estimation in nonlinear regression models", *Technometrics*, 27(3), 241-250.

[44] Hemker P.W. and Kok J. (1993), *A project on parameter identification in reaction kinetics*, . Stichting Mathematisch Centrum, Amsterdam; NM-R9301

[45] Jeffreys H. (1961), *Theory of probability*, 3rd-ed, Clarendon, Oxford.

[46] Kiefer J. (1974), "General equivalence theory for optimum designs (approximate theory)", *Annals of Statistics*, 2(5), 849-879.

[47] Kiefer J. and Wolfowitz J. (1960), "The equivalence of two extremum problems", *Can. J. Mathematics*, 12, 363-366.

[48] Klaus R.A. (1981), "A computer based methodology for regression and experimental design with nonlinear algebraic and ordinary differential equation multiresponse models", Ph.D. thesis, Swiss Federal Institute of Technology, Zurich, Switzerland.

[49] Leis J.R. and Kramer M.A. (1985), "Sensitivity analysis of systems of differential and algebraic equations", *Computers & Chemical Engineering*, 9(1), 93-96.

[50] Leis J.R. and Kramer M.A. (1988), "The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations", *ACM Transactions on Mathematical Software*, 14(1), 45-60.

[51] Ljung L. and Glad T. (1994), "On global identifiability for arbitrary model parametrizations", *Automatica*, 30(2), 265-276.

[52] Markussen C.R. and DiStefano III J.J. (1982), "Evaluation of four methods for computing parameter sensitivities in dynamic sytem models", *Mathematical Biosciences*, 61, 135-148.

[53] McManus N. (1999), personal communication, University of Waterloo, Waterloo

[54] Moros R., Kalies H., Rex H.G., and Schaffarczyk S. (1996), "A gentic algorithm for generating initial parameter estimations for kinetic models of catalytic processes", *Computers & Chemical Engineering*, 20(10), 1257-1270.

[55] Oberhofer W. (1982), "The consistency of nonlinear regression minimizing the L_1-norm", *The Annals of Statistics*, 10, 316-319.

[56] O'Brien T.E. (1992), "A note on quadratic designs for nonlinear regression models", *Biometrika*, 79(4), 847-849.

[57] Oxby P.W. (1997), "Multivariate weighted least squares as a preferable criterion for multiresponse parameter estimation", Ph.D. thesis, University of Waterloo.

[58] Pierce TH, Cukier RI, and Dye JL. (1981), "Application of nonlinear sensitivity analysis to enzyme mechanisms", *Mathematical Biosciences*, 56, 175-208.

[59] Press W.H., Flannery B.P., Teukolsky S.A., and Vetterling W.T. (1989), *Numerical recipes, the art of scientific computing, Fortran version*, Cambridge: Cambridge University Press.

[60] Pronzato L. and Walter E. (1988), "Robust experiment design via maximin optimization", *Mathematical Biosciences*, 89, 161-176

[61] Rabitz H., Kramer M., and Dacol D. (1983), "Sensitivity analysis in chemical kinetics", *Annual Rev. Phys. Chem.*, 34, 419-461.

[62] Rahni N., Ramdani N., Candau Y. and Dalicieux (1997), "Application of group screening to dynamic building energy simulation models", *J. Statist. Comput. Simul.*, 57, 285-304.

[63] Seber G.A.F. and Wild C.J. (1989), *Nonlinear Regression*, Toronto: John Wiley & Sons.

[64] Shaw B.M. (1994), "Nonlinear experimental design for precise parameter estimation", Masters thesis, University of Waterloo.

[65] Stewart W.E., Caracotsios M., and Sorensen J.P. (1992), "Parameter estimation from multiresponse data", *AIChE Journal*, 38(5), 641-650.

[66] Sulieman H. (1998), "Parametric sensitivity analysis in nonlinear regression", Ph.D. thesis, Queen's University, Kingston, Ontario.

[67] Tidwell P.W. and Mortimer G.A. (1965), "An improved method of calculating copolymer reactivity ratios", *Journal of Polymer Science, Part A: Polymer Chemistry*, 3, 369-387.

[68] Vajda S., Rabitz H., Walter E., and Lecourtier Y. (1989), "Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models", *Chem. Eng. Comm.*, 83, 191-219.

[69] Vila J.P. (1990), "Exact experimental designs via stochastic optimization for non-linear regression models", *Compstat*,

[70] Wald A. (1943), "On the efficient design of statistical investigations", *Annals of Mathematical Statistics*, 14, 134-140

[71] Watts D.G. (1994), "Estimating parameters in nonlinear rate equations", *Can. J. Chem. Eng.*, 72, 701-710.

[72] Welsh A.H., Carrol R.J., and Ruppert D. (1994), "Fitting heteroscedastic regression models", *JASA, Theory and Methods*, 89(425), 100-116.

[73] Yao L. and Sethares W.A. (1994), "Nonlinear parameter estimation via the genetic algorithm", *IEEE trans. signal processing*, 42(4), 927-935.

[74] Ziegel E.R. and Gorman J.W. (1980), "Kinetic modelling with multiresponse data", *Technometrics*, 22(2), 139-151.