

University of Alberta

Design of Experiments for Large Scale Catalytic Systems

by

Siddhartha Kumar

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

In

Process Control

Department of Chemical and Materials Engineering

©Siddhartha Kumar

Fall 2011

Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

To Mummy, Papa, Anjani and my lovely sisters

Abstract

Parameter estimation for mathematical models is performed based on the data collected by experiments using system identification techniques. However, since performing experiments can be time consuming as well as expensive, experiments must be designed prior to performing, so that the data collected will be optimal for parameter estimation. This thesis aims at performing experimental design while addressing three different design problems: (1) *non-identifiability* for large scale catalytic systems, (2) *uncertainty* in parametric values being used for design, and (3) parameter estimation for a specific subset of reactions. Hierarchical clustering, stochastic optimization and computational singular perturbation are the methodologies used in this study. Catalytic systems under investigation are ammonia decomposition and preferential oxidation for hydrogen production for fuel cells.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Vinay Prasad for all the efforts he has put in me, for all the time he has invested in me, for all the guidance he has given me not only through my studies but also in improving me as a person. He has prepared me to face the real world with all my might. I would also like to thank him for all the wonderful opportunities he has provided me with right from the very first conference till the completion of my thesis. This thesis is a representation of what I have learnt academically, but there is a huge part of the knowledge I have gained under Dr. Prasad's guidance which cannot be documented. I would like to thank Dr. Semagina, who has also been a constant source of inspiration for me. Sometimes when I got caught in research with no solution to the rescue, it is the company of my friends that helped me maintain my calm. I would like to thank my friends: Rishi, Aditya, Nitesh, Dipen, Neha, Rajesh, Vikram, Ajay and Vishnu for the wonderful time I have had in Edmonton. I would also like to thank some of my seniors: Vivek, Yashasvi, Sameedh, Johnny, Arjun and Anuj for their guidance throughout my stay in Edmonton. A special thanks to Aditya and Anuj for their help during my research. Last but not the least, I need to thank my wonderful family and my sweet girlfriend Anjani who has always been very supportive and loved me so much.

Table of Contents

1. Introduction	1
1.1 Design of Experiments	1
1.2 The D-Optimal Design	3
1.3 Kinetic Systems	6
1.3.1 Ammonia Decomposition	6
1.3.2 Preferential Oxidation	9
1.3.3 Reactor Modeling	11
1.4 Thesis Overview	12
1.4.1 Thesis Outline	12
1.4.2 Thesis Contributions	15
2. Parameter Estimation for Large-Scale Kinetic Systems	17
2.1 Introduction	17
2.1.1 Identifiability	18
2.1.2 The D-Optimal Design	19
2.2 The Methodology	20
2.2.1 Principal Component Analysis	22
2.2.2 Hierarchical Clustering	23
2.2.3 D-Optimal design for PROX chemistry on Pt catalyst	24
2.2.4 D-Optimal design for PROX chemistry on Rh catalyst	29
2.3 Conclusions	31
3. Stochastic Nonlinear Optimization for Optimal Design of Experiments	32
3.1 Introduction	32
3.2 Microkinetic and Reactor Model under Uncertainty	34

3.2.1	Basic Kinetic Models.....	34
3.2.2	Stochastic Description and Sampling	34
3.3	Stochastic Optimization using Sample Average Approximation and Constrained Particle Swarm Optimization	36
3.3.1	Sample Average Approximation	39
3.3.2	Function Evaluation.....	40
3.3.3	Deterministic Optimization	40
3.3.4	Particle Swarm Optimization.....	40
3.3.4	Optimization Parameters	43
3.3.5	Constrained PSO.....	46
3.4	Computational Results	49
3.5	Conclusions	53
4.	Computational Singular Perturbation for Design of Experiments	55
4.1	Introduction	55
4.2	Computational Singular Perturbation Theory	56
4.2.1	Mathematical Representation of the Kinetic System.....	57
4.2.2	Basis Vectors	58
4.2.3	The Classification of Fast and Slow Modes	60
4.2.4	The CSP Refinement Procedure	61
4.2.5	Radical Pointers and Participation Indices	62
4.3	The Reaction System and CSP Analysis	63
4.3.1	Basic Kinetic Models.....	63
4.4	Design of Experiments using CSP Analysis	69
4.4.1	The Reaction Metric DOE	69
4.4.2	The Search Method.....	70

4.4.3	Application to Ammonia decomposition model.....	70
4.5	Conclusions	72
5.	Concluding Remarks and Future Work.....	73
5.1	Concluding Remarks	73
5.2	Future Work.....	75
5.2.1	Stochastic Sampling	75
5.2.2	Input Sequence design for Design of Experiments	75
6.	Bibliography.....	76
7.	Appendix	80

List of Tables

Table 1-1 Rate constants for different types of reactions	7
Table 1-2 Elementary reactions representing the decomposition of ammonia on a Ru surface	8
Table 1-3 Preferential Oxidation (PROX) reactions;.....	10
Table 2-1 Ranges of Operating Conditions.....	21
Table 3-1 Laboratory scale reactor and catalyst specifications	34
Table 3-2 Bounds of decision variables for the optimization problem.....	41
Table 3-3 PSO Parameter values used in the Optimization Algorithm	44
Table 3-4 Comparison of objective function values for different values of β	46
Table 3-5 Comparison of approaches to handle constrained PSO.....	49
Table 3-6 Stochastic and Deterministic Optimal Solutions	50

List of Figures

Figure 1.1 A schematic for parameter estimation using DOE	13
Figure 2.1 Reaction index of kinetic parameters (pre-exponentials) clustered together for Pt. Reaction indices are the same as in Table 1-3. Figure is generated for the first cluster.	26
Figure 2.2 Reaction index of kinetic parameters (pre-exponentials) clustered together for Pt. Reaction indices are the same as in Table 1-3. Figure is generated for the second cluster.....	26
Figure 2.3 D-optimal index for all the sample points and their frequency of occurrence in the cluster for Pt. The variance is large and there are many sample points which score high on the D-optimal index.....	28
Figure 2.4 Distribution of D optimal points for Rh catalyst	29
Figure 2.5 Reaction index of kinetic parameters (pre-exponentials) clustered together for Rh.....	30
Figure 3.1 Uniform distribution of the order of pre-exponentials; a single random value is selected from each of the four regions	36
Figure 3.2 Change in Objective Function per iteration for PSO.....	45
Figure 3.3 Comparison of objective function values for different values of β	47
Figure 3.4 A plot showing the variation of objective function with respect to temperature and pressure, for the deterministic case.....	51
Figure 3.5 Contour plot of the objective function against temperature and pressure. The figure does not reveal any specific optimal region. ..	51
Figure 3.6 A plot showing the variation of objective function with respect to temperature and pressure, for the stochastic case	52
Figure 3.7 Top view of Figure 3.6. The figure reveals optimal region (marked on the plot) where the objective function reaches a maxima.....	52
Figure 4.1 Time variation of the number of exhausted modes for the ammonia decomposition model. After 0.1ms, all the modes get exhausted and none contribute to the dynamics	65

Figure 4.2 Time variation of amplitude of each mode for the ammonia decomposition catalytic system. Modes 8 and 9 appear to be an active mode.....	65
Figure 4.3 Time scale of each mode for the ammonia decomposition catalytic system combined with a PFR model. The thin black line indicates the current time sample. Curves lying above this line are the active modes for the corresponding time sample	67
Figure 4.4 Participation Index of each reaction in Mode 9, which is an active mode for most of the times. Broken curves suggest that CSP data fails to give information at some time samples. Reactions 5 and 6 have negligible participation in this mode, however they might be participating in some other active mode	68
Figure 4.5 Radical Pointer for each species of mode 9, which is an active mode for most of the times. We can see that species 5, 6, 7 and 8 have negligible probability of being the radical in this mode, however they can be radicals in some other active mode	68
Figure 4.6 Variation in the value of reaction metric with respect to temperature (K) and NH ₃ mole fraction. The plot reaches a peak at 830K, but is not affected much by the NH ₃ mole fraction	70
Figure 4.7 Variation in the value of reaction metric shown with respect to NH ₃ and H ₂ mole fraction. The plot shows no specific shape and no direct relationship can be deduced.....	71

List of Symbols

S	sensitivity matrix
y	output
θ	parameter set
FIM	fisher information matrix
e	measurement noise
k	rate constant
R	universal gas constant ($\text{kcal.mol}^{-1}.\text{K}^{-1}$)
s	sticking coefficient (dimensionless)
T	temperature (K)
M	molecular mass (g.mol^{-1})
T_0	reference temperature (300K)
E_a	activation energy (kcal.mol^{-1})
A	pre-exponential (s^{-1})
σ	site density (mol.cm^{-2})
n	reaction order (dimensionless)
F_j	molar flow rate (mol.s^{-1})
r_j	rate of reaction ($\text{mol.cm}^{-2}.\text{s}^{-1}$)
L	length of reactor (cm)
P	pressure (atm)
V	volume of reactor (cm^3)
v	volumetric flow rate ($\text{cm}^3.\text{s}^{-1}$)
τ	residence time (s)
a_v	Ratio of catalyst surface area to volume (cm^{-1})
x_i	mole fraction (dimensionless)
ϕ_{ik}	angle between sensitivity vectors (radians)
s_i	sensitivity vector
u	inlet conditions

D	diameter (cm)
ΔH_i	change in enthalpy (kcal.mol ⁻¹)
ΔS_i	change in entropy (kcal.mol ⁻¹ .K ⁻¹)
X	set of decision variables for optimization
ε	random variable
E	expectation operator
LB	lower bound on decision variables
UB	upper bound on decision variables
J	logarithm of the determinant of FIM
β	weight imparted to variance of J
$\hat{f}(X)$	estimated value of $f(X)$
N_{iter}	number of iterations i to be performed
N_{pt}	number of particles
N_d	number of search dimensions
N_p	number of parameters to be estimated
C_1	weight parameter for PSO
C_2	weight parameter for PSO
W_{int}	weight parameter for PSO
x^{min}	lower limit of decision variables
x^{max}	upper limit of decision variables
v_j^{max}	maximum particle velocity
g	reaction rate vector
S_r	stoichiometric column vector for the r^{th} elementary reaction
F^r	reaction rate of the r^{th} elementary reaction
b^i	orthonormal row basis vector for CSP
a_k	orthonormal column basis vector for CSP
f^i	amplitude of i^{th} mode for CSP
J	jacobian of g with respect to y
$\lambda(i)$	eigenvalues of J
$\tau(m)$	time scale of the m^{th} exhausted mode

$a_i^0(M)$	refined set of column basis vectors
$b_0^m(M)$	refined set of row basis vectors
$Q(M)$	fast subspace projection matrix
$P_m(r)$	participation index of r^{th} reaction in m^{th} mode
RM	reaction metric

List of Abbreviations

DOE	Design of Experiments
OVAT	One Variable At a Time
RSM	Response Surface Methodology
FIM	Fisher Information Matrix
RDS	Rate Determining Step
MARI	Most Abundant Reaction Intermediate
UBIQEP	Unity Bond Index Quadratic Exponential Potential
PROX	Preferential Oxidation
PFR	Plug Flow Reactor
CSP	Computational Singular Perturbation
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
VSS	Value of Stochastic Solution
ODE	Ordinary Differential Equation
CSTR	Continuous Stirred-Tank Reactor
TVAT	Two Variables At a Time

1. Introduction

1.1 Design of Experiments

Experiments are normally carried out to collect data based on which a particular hypothesis can be tested. However, as experiments are often expensive and time-consuming, their choice must be optimized to reduce the number necessary to accomplish a user defined objective. The Design of Experiments (DOE) is a method regarded as accurate for testing a hypothesis. It is a tool to develop an experimentation strategy that maximizes information gained about a system using a minimum of resources. A sound experimental design follows established scientific protocols and generates good statistical data. The basic theory of DOE was developed some 50 years ago (Lucas *et al.*, 1959). The objective of experimental design is to establish a causal relationship between variables by manipulating an independent variable to assess the effect upon dependent variables. The dependent variable, in the context of DOE, is called the response, and the independent variables are called factors. The aim is to select combinations of factors that will provide the most information on the causal relationship. In general, the design of an experiment is influenced by (1) the objectives of the experiment, (2) the extent to which sequential experimentation will be performed, if at all, (3) the number of factors under investigation, (4) the possible presence of identifiable and non-identifiable factors, (5) the model for the response variable (Ryan, 2007). The objective for each experiment should be clearly delineated, as these objectives will dictate the construction of the designs.

Many classical design methods have been used for experimental design. The first type of design is the one factor design which is a one-variable-at-a-time

(OVAT) approach, i.e. only one factor is under investigation. Another method for DOE is factorial design wherein multiple factors are investigated simultaneously. Different kinds of factorial designs are commonly used, such as, general full factorial design, full factorial design with two levels, fractional factorial designs with two levels and factorial designs with more than two levels (Ryan, 2007). Factorial designs with two levels is also known as 2^k factorial design, and are the simplest possible designs, requiring 2^k experiments, where k is the number of variables under study. In these designs, each variable has two levels, -1 and $+1$, and the variables can be either quantitative (e.g., temperature, pressure, concentration) or qualitative (e.g., type of catalyst, type of apparatus, sequence of operations). For understanding, let us take an example of designing experiments for a 3 variable system. In this case, factorial design implies $2^3 = 8$ experiments would have to be performed. The mathematical model is therefore the following:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3$$

... (1.1)

As a consequence, with just eight experiments, it is possible to estimate a constant term, the three linear terms, the three two-term interactions and the three-term interaction. Contrary to what happens in the one-variable-at-a-time (OVAT) approach, the factorial design is suitable for estimating the interactions between variables. Upon detailed study of factorial designs, it can be inferred that when more than one response is studied, all of them should be considered at the same time, thus finding a combination better than given by the OVAT approach. In the case of a large number of factors, the experimental cost required is very high, since all possible combinations of factor values must be taken into consideration. In this case, fractional factorial designs can be used to reduce the size of the problem. Factorial methods are not suited to the situation where there are some constraints on the outputs (or internal states of the experiment such as temperatures, pressures, etc.). They are also not well suited to handle dynamic experiments, where both “factors” and “responses” may not be single values (say, a constant temperature or a conversion) but complex time profiles of the same

variables. However, in view of their simplicity, these methods are still widely used in the DOE.

Another method used for DOE is response surface methodology (RSM) which is a collection of tools for fitting a surface to a set of data, and determining optimum factor levels (Sarabia *et al.*, 2009). The shape of the surface is determined by the model that is fit to the data, as well as the response values, and hence the term “response surface”. Typically, a full second-order model is fit in trying to determine the optimum combination of factor levels. Before that is done, generally a first-order model is fit and then the method of steepest ascent or descent is used to try to zero in on the operating region, and then a design is used to accommodate a second-order model for the purpose of characterizing the region.

1.2 The D-Optimal Design

As opposed to these “black-box” statistical experiment design methods, another form of optimal design has been developed. These methods take advantage of the knowledge of the structure of the system under consideration, where the system is represented by a mathematical model. These model-based experiment design techniques can be applied to any system, including linear, non-linear, steady state or transient processes. Their goal is typically to assist in the rapid development, refinement and statistical validation of process models. These model-based experiment designs use the model equations (including any constraints) and current parameters explicitly to estimate parameter values and conditions for the next experiment. It applies an optimization framework, which requires the evaluation of an appropriate objective function. Generally good designs are constructed such that they meet a certain optimality criterion, and the collection of such criteria is referred to as alphabetic optimality. Many different criteria have been suggested and labeled by single capital letters A, E, D, I and so on, which explains the term alphabetic optimality. Most of these designs start with

evaluating the influence of a perturbation in the estimated parameter vector over the predicted output. For this purpose, the sensitivity matrix (\mathbf{S}) is calculated as

$$\mathbf{S} = \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^T} \right) \quad (1.2)$$

where \mathbf{y} is the output we are interested in, and $\boldsymbol{\theta}$ is the parameter set. The measured output is a function of the parameters, and may be affected by measurement noise. In practice, the measurement noise is often assumed to be normally distributed with zero mean and a covariance given by the matrix $\boldsymbol{\Sigma}$. As a result, the measurements are also normally distributed even for a nonlinear system. The Fisher information matrix, \mathbf{FIM} , is defined as:

$$\mathbf{FIM}(\boldsymbol{\theta}) = \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^T} \right)^T \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^T} \right) \quad (1.3)$$

If the measurement noise is uncorrelated and has a constant variance with time, the covariance matrix becomes $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$ and the Fisher information matrix (FIM) becomes the transpose of the sensitivity matrix multiplied by itself. Without loss of generality, it can be assumed that $\sigma^2 = 1$, which gives:

$$\mathbf{FIM}(\boldsymbol{\theta}) = \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^T} \right)^T \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^T} \right) \quad (1.4)$$

The FIM is thus always a positive semi-definite symmetric matrix. Its inverse is also the lower bound on the variance of any statistic, which explains why the quantity in equation 1.4 is called the ‘‘information’’ matrix. The larger the value of the FIM is, the smaller the variance becomes, and therefore, there is more certainty about the location of the unknown parameter value.

The most common approach to optimal experiment design is the D-optimal experiment design, wherein the determinant of the Fisher information matrix is maximized. Other alphabetic optimality criteria include the average-variance criterion, the smallest-eigenvalue criterion, and the trace criterion

(Pukelsheim, 1993). The D-optimal criterion for experimental design has been used for most the different methods proposed in this work.

Consider the example of a system with one input, two parameters, and two output variables:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} u^2\theta_1 + \theta_2 \\ \theta_1^2 + u\theta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (1.5)$$

where $\mathbf{y} = [y_1 \ y_2]^T$ are output variables, $\boldsymbol{\theta} = [\theta_1 \ \theta_2]^T$ are parameters, u is an input variable determining the experimental condition, and $\mathbf{e} = [e_1 \ e_2]^T$ represent noise with a Gaussian distribution with

$$E(\mathbf{e}) = \mathbf{0}, \quad \text{Var}(\mathbf{e}) = \mathbf{I}$$

The nominal values of the parameters are θ_{10} and θ_{20} . The sensitivity matrix (\mathbf{S}) and \mathbf{FIM} are calculated as,

$$\mathbf{S} = \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^T} = \begin{bmatrix} u^2 & 1 \\ 2\theta_{10} & u \end{bmatrix} \quad (1.6)$$

$$\mathbf{FIM} = \mathbf{S}^T \mathbf{S} = \begin{bmatrix} u^2 & 2\theta_{10} \\ 1 & u \end{bmatrix} \begin{bmatrix} u^2 & 1 \\ 2\theta_{10} & u \end{bmatrix} = \begin{bmatrix} u^4 + 4\theta_{10}^2 & u^2 + 2u\theta_{10} \\ u^2 + 2u\theta_{10} & 1 + u^2 \end{bmatrix} \quad (1.7)$$

$$\varphi_D^* = \max_u \log \det(\mathbf{FIM}) \quad (1.8)$$

$$\det(\mathbf{FIM}) = 4\theta_{10}^2 + u^6 - 4u^3\theta_{10} \quad (1.9)$$

Taking the derivative in order to find the u for which the $\det(\mathbf{FIM})$ can be maximized, yields:

$$u^3 = 2\theta_{10} \quad (1.10)$$

Hence, the experiments performed at input conditions given by equation (1.10) are D-optimal and the data generated will be good for estimating the values of parameters. However it is important to note here that the value of input conditions

is directly related to the nominal value of the parameter. Hence, uncertainty in the nominal values will affect the experimental design directly.

1.3 Kinetic Systems

1.3.1 Ammonia Decomposition

The overall reaction for the decomposition of ammonia into hydrogen and nitrogen is:



This reaction has also been studied on a variety of catalysts. The reaction proceeds with the adsorption of ammonia gas onto the catalyst surface, where hydrogen atoms are abstracted one at a time. Atomic nitrogen and hydrogen combine to produce N_2 and H_2 , respectively. Each of the elementary steps in the reaction network was assigned a modified Arrhenius equation, as shown in Table 1-1. Each reaction used an activation energy that was dependent both on temperature and the surface coverage of various species (Table 1-2). We describe this using microkinetic models, which consider a detailed reaction mechanism consisting of all relevant elementary reactions, and the species concentrations are solved rigorously using numerical methods, without any assumptions made about a rate determining step (RDS), the most abundant reaction intermediate (MARI), or partial equilibrium conditions. Table 1-1 shows the expressions for the rate constants of elementary reactions in the modified Arrhenius form.

Table 1-1 Rate constants for different types of reactions

Reaction Type	Rate Constant
Adsorption	$k = \frac{s}{\sigma^n} \sqrt{\frac{RT}{2\pi M}} \left(\frac{T}{T_0}\right)^\beta e^{-\frac{E_a(\theta,T)}{RT}}$
Desorption	$k = \frac{A}{\sigma^{n-1}} \left(\frac{T}{T_0}\right)^\beta e^{-\frac{E_a(\theta,T)}{RT}}$
Surface Reaction	$k = \frac{A}{\sigma^{n-1}} \left(\frac{T}{T_0}\right)^\beta e^{-\frac{E_a(\theta,T)}{RT}}$

There are two major sets of parameters in microkinetic models – the activation energies and the pre-exponential factors. The activation energy for each reaction is specified using the unity bond index - quadratic exponential potential (UBIQEP) method (Prasad *et al.*, 2008), and is not estimated using the experimental design, whereas the pre-exponentials are considered as parameters for estimation. The system we use to demonstrate our approach is the catalytic decomposition of ammonia to produce hydrogen. It is a well-studied model system often used for demonstrations of proof-of-concept in catalytic studies (Prasad *et al.*, 2010). The model was developed considering the irreversible elementary reactions shown in Table 1-2, along with the nominal values for the sticking coefficients and pre-exponentials to be used for the D-optimal design, and the coverage and temperature dependencies of the activation energies.

Table 1-2 Elementary reactions representing the decomposition of ammonia on a Ru surface

No.	Reaction	Nominal Values of Sticking Coefficient [unitless] or Pre-exponential factor [s ⁻¹]	Activation Energy [$\frac{kcal}{mol}$]
1	$H_2 + 2 * \xrightarrow{k_1} 2H *$	8.7×10^{-1}	0.0
2	$2H * \xrightarrow{k_2} H_2 + 2 *$	1.1×10^{11}	$19.6 - 7.0\theta_H$
3	$N_2 + 2 * \xrightarrow{k_3} 2N *$	2.0×10^{-1}	$7.1 - 26.3\theta_N$
4	$2N * \xrightarrow{k_4} N_2 + 2 *$	1.7×10^{12}	$51.0 - 43.8\theta_N$
5	$NH * + * \xrightarrow{k_5} N * + H *$	1.9×10^{12}	$5.3 + 15.5\theta_N + \theta_H$
6	$N * + H * \xrightarrow{k_6} NH *$	7.6×10^9	$37.6 - 19.5\theta_N - 2.5\theta_H$
7	$NH_2 * + * \xrightarrow{k_7} NH * + H *$	2.0×10^{12}	$20.1 + 1.2\theta_H$
8	$NH * + H * \xrightarrow{k_8} NH_2 * + *$	1.4×10^{10}	$15.9 - 2.3\theta_H$
9	$NH_3 * + * \xrightarrow{k_9} NH_2 * + H *$	2.0×10^{12}	$18.7 + 1.3\theta_H$
10	$NH_2 * + H * \xrightarrow{k_{10}} NH_3 * + *$	3.4×10^9	$11.6 - 2.2\theta_H$
11	$NH_3 + * \xrightarrow{k_{11}} NH_3 *$	1.5×10^{-4}	0.0
12	$NH_3 * \xrightarrow{k_{12}} NH_3 + *$	8.1×10^{11}	17.7

θ_H and θ_N are the fractional coverages of H* and N* respectively. * indicates a vacant site or, in conjunction with a chemical species, an adsorbate

1.3.2 Preferential Oxidation

The catalytic kinetic system described in this section is the preferential oxidation of CO on platinum and rhodium catalysts. This process is of importance in the production of hydrogen for use in fuel cells through the reforming of fuels such as methane, methanol or propane. The carbon monoxide produced during reforming acts as a poison for proton exchange membrane fuel cells, and must be removed or converted before the hydrogen rich stream is fed to the fuel cell. Preferential oxidation of CO (with hydrogen not being converted to water) converts it to carbon dioxide, while maintaining process efficiency in terms of hydrogen production. A detailed kinetic model for this system may be found in (Mhadeshwar *et al.*, 2004, 2005), and is the starting point for our investigation. While the model is able to describe available data qualitatively and, to some extent, quantitatively, the aim is to develop optimal experimental designs to develop parameter estimates that improve the quantitative predictive capabilities of the model. Table 1-3 shows the 23 reversible elementary reactions conducted on the catalyst, and the pre-exponentials for these reactions are the parameters to be estimated. For each reversible reaction, the pre-exponential factors for the forward and backward reactions are constrained to be in specific ratios based on thermodynamic consistency (Mhadeshwar *et al.*, 2003). The activation energy for each reaction is specified using the UBIQEP method (Prasad *et al.*, 2008), and is not estimated using the experimental design. The model consists of the kinetic expressions for each of the elementary reactions, which are used to form the rate equations for all gas phase and surface species, and these are integrated along the length of a plug flow microreactor, which is representative of the typical system for kinetic experimental studies. There are 12 gas phase species and 9 surface species, resulting in 21 state variables.

Table 1-3 Preferential Oxidation (PROX) reactions;
 * indicates vacant site or adsorbed species, as applicable

Index	Reaction
1	$H_2 + 2 * \leftrightarrow 2H *$
2	$O_2 + 2 * \leftrightarrow 2O *$
3	$OH * + * \leftrightarrow O * + H *$
4	$H_2O * + * \leftrightarrow H * + OH *$
5	$H_2O * + O * \leftrightarrow 2OH *$
6	$OH + * \leftrightarrow OH *$
7	$H_2O + * \leftrightarrow H_2O *$
8	$H + * \leftrightarrow H *$
9	$O + * \leftrightarrow O *$
10	$CO + * \leftrightarrow CO *$
11	$CO_2 + * \leftrightarrow CO_2 *$
12	$CO_2 * + * \leftrightarrow CO * + O *$
13	$CO_2 * + H * \leftrightarrow CO * + OH *$
14	$COOH + * \leftrightarrow COOH *$
15	$COOH * + * \leftrightarrow CO * + OH *$
16	$COOH * + * \leftrightarrow CO_2 * + H *$
17	$CO * + H_2O * + * \leftrightarrow COOH * + H *$
18	$CO_2 * + OH * \leftrightarrow COOH * + O *$
19	$CO_2 * + H_2O * \leftrightarrow COOH * + OH *$
20	$HCOO + 2 * \leftrightarrow HCOO **$
21	$CO_2 * + H * \leftrightarrow HCOO **$
22	$CO_2 * + OH * \leftrightarrow HCOO ** + O *$
23	$CO_2 * + H_2O * \leftrightarrow +HCOO ** + OH *$

1.3.3 Reactor Modeling

Fluid going through a PFR is modeled as flowing through the reactor as a series of many “plugs”, each with uniform composition, traveling in the axial direction of the reactor, with each plug having a different composition from the ones before and after it. The key assumption is that as a plug flows through a PFR, the fluid is perfectly mixed in the radial direction but not in the axial direction. An ideal plug flow reactor has a fixed residence time, i.e. any fluid that enters the reactor a time t will exit the reactor at time $t + \tau$, where τ is the residence time of the reactor. A plug flow reactor with the tube packed with catalyst is known as packed bed reactor (PBR). The catalyst surface area is modeled as a continuous distribution of catalyst sites throughout the reactor. The catalyst was characterized through an effective catalyst surface area to volume ratio and site density (sites per surface area). The inlet gas compositions, temperature and pressure were treated as variables to be solved for, by the experimental design. Argon was used as the inert gas. The mole balance for flow systems is given by:

$$F_{j_0} - F_j + \int_0^V r_j dV = \frac{dN_j}{dt} \quad (1.12)$$

where, F_{j_0} is the inlet molar flow rate of species j , F_j is the molar flow rate out of species j , V is the volume of the reactor, r_j is the total reaction rate for the species j and N_j is the number of moles of species j . For steady state models, the derivative on the right hand side of the above equation is set to zero. The PFR steady state equation is given by:

$$\frac{dF_j}{dL} = A \cdot r_j \quad (1.13)$$

where, dL is the differential element along the length of the PFR and A is the area of cross section of the PFR. The individual rate of the species is given by:

$$r_j = \sum_{i=1}^N r_{ij} \quad (1.14)$$

where, N is the number of reactions. The concentration of the species for gaseous phase is given by:

$$C_j = C_{T_0} \frac{F_j}{F_T} \frac{P}{P_0} \frac{T_0}{T} \quad (1.15)$$

where C_{T_0} is the initial total concentration, F_T is the total molar flow rate, P_0 is the initial pressure, T_0 is the initial temperature. The concentration of the gaseous species is related to the molar flow rate by:

$$C_j = \frac{F_j}{v} \quad (1.16)$$

Since the system we are dealing with is a catalytic system, the surface species also have to be taken into account. The rate of reaction for surface reactions is calculated in the units $mol.cm^{-2}.s^{-1}$ and the rate constants are given by the table 1.1. The rate of change of surface concentration of surface species in the units $mol.cm^{-2}$.

1.4 Thesis Overview

1.4.1 Thesis Outline

In the field of chemical reaction engineering, experimental data is collected in order to estimate parameters of a kinetic model which are pre-exponential factors and activation energies. The objective is to estimate these parameters as precisely as possible with the minimum experimental effort. Also, since the quality of experimental data used for the estimation of the parameters of a model influences directly the accuracy of the estimated parameters, it is necessary to carefully design the experiments so that the chosen experimental conditions bring the most meaningful information. In this study, the model structure is already known and the objective of performing experiments is the evaluation of model parameters. This is the so-called parameter estimation problem and many solutions have been proposed. Hunter *et al.* (1965) came up with one of the first solutions by developing optimal strategies for parameter estimation. This study is based on a similar idea where the task was to perform experiments sequentially such that the volume of the confidence region formed by

the estimated parameters in the parameter space is minimized. The parameters of interest are the pre-exponentials for the individual reactions only. Firstly, the model structure for such systems is established. The aim is to find the parameter values (pre-exponentials) in order to completely model the system. DOE is performed with the complete mathematical model which uses the nominal values of the parameters. This step gives the factor values (experimental conditions) at which an experiment should be performed and data should be collected. With the collected data, a new and improved set of parameter values are estimated by fitting onto the model. The whole procedure is sequentially repeated in order to get accurate estimates for the parameters. Figure 1.1 below explains this procedure in the form of a flowchart.

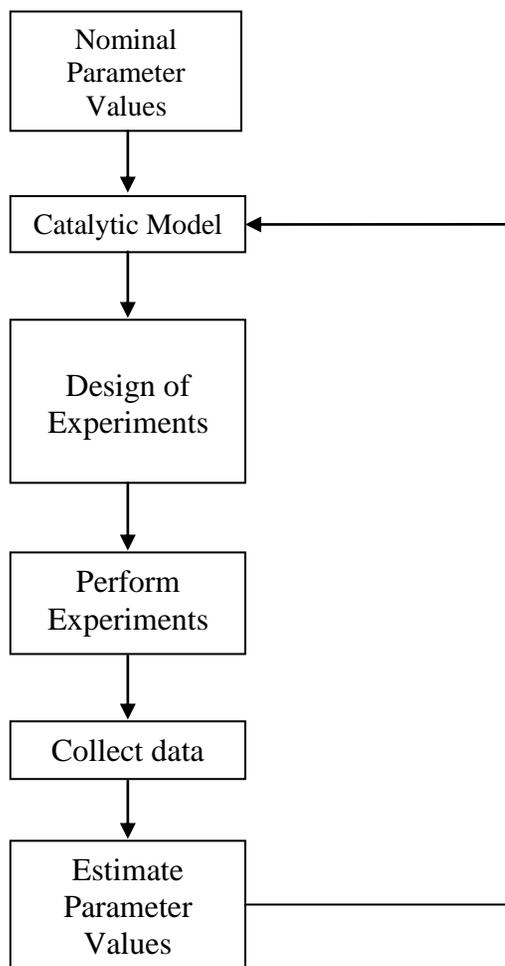


Figure 1.1 A schematic for parameter estimation using DOE

Unfortunately, since kinetic models are nonlinear in nature (due to the presence of exponential terms), the design of experiments can be a difficult task due to the presence of multiple suboptimal solutions (non-convexity), or multiple equivalent solutions (a result of non-identifiability of the full set of parameters). Non-convexity may be surmounted by the use of global optimization methods (Balsacanto *et al.*, 2007). Non-identifiability also presents formidable complications. Two types of identifiability problems are found: *structural* identifiability is related to the model structure independent of experimental data, and *practical* identifiability, which also takes into account the amount and quality of measured data that was used for parameter calibration (Walter *et al.*, 1997). Such identifiability issues are addressed for the ammonia decomposition kinetic system in Chapter 2.

There are other issues to consider while performing experimental design using the D-optimal criterion. It is well known that this design metric generally depends on the value of the parameters to be estimated. Since this value is of course unknown, the usual practice is to select an experiment that is D-optimal for some reasonable nominal value of the parameters. If the actual value of the parameters to be estimated happens to differ too much from this nominal value, the designed experiment may prove very far from optimal. In chapter 3, we present a methodology which allows uncertainty in the model parameters to be taken into account during the design of the experiment. The parameters to be estimated are assumed to belong to a distribution with known statistics, and we look for an experiment maximizing the expectation of the determinant of the Fisher information matrix over the distribution using a suitable optimization technique.

Sometimes, in chemical reaction engineering, interest might lie in only a few species or reactions, or more specifically in particular reaction chemistries. In chapter 4, we present a methodology in order to design experiments influenced by such an objective. The challenge here is to outline a procedure in order to choose

a design criterion that maximizes a metric related to that particular reaction. We use computational singular perturbation (CSP) theory in order to extract kinetically useful information (Lam *et al.*, 1994). This method is computationally very expensive; hence in order not to increase the computational cost further, the D-optimal design criterion was not used. We therefore used a grid based experimental design with the reaction metric being evaluated over a grid of input conditions.

The outline of the thesis in brief is as follows: Chapter 2 describes a method for addressing the identifiability issues by performing DOE based on hierarchical clustering for reducing the number of parameters to be identified; Chapter 3 outlines a stochastic optimization technique which considers the uncertainty in nominal parameter values to maximize the value of the D-optimal metric averaged over a number of samples, Chapter 4 describes another method which uses CSP in order to perform DOE for the estimation of desired reaction parameters, and Chapter 5 concludes the thesis as well as outlines the scope for future work.

1.4.2 Thesis Contributions

Modifications in the conventional techniques for optimal experimental design have been presented in this thesis. The main thesis contributions are as follows:

- 1) Chapter 2 presents a comparison of different techniques for dimension reduction such as principal component analysis, hierarchical clustering and singular value decomposition in order to perform D-optimal experimental design for two different catalysts for preferential oxidation. This part of the thesis has been published in the International Journal of Advanced Mechatronic Systems (Subramanian *et al.*, 2011).
- 2) In chapter 3, uncertainties in nominal parameter values have been taken into account while performing DOE using a stochastic optimization approach. This work has been presented as a poster at the AIChE Annual Meeting 2010 in Salt lake City, USA.

- 3) A comparison of two different approaches to constrained non-linear particle swarm optimization has also been demonstrated in chapter 3.
- 4) In chapter 4, the mathematical tool, computational singular perturbation, has been applied to a transient ammonia decomposition system modeled in a PFR approximated by a series of CSTRs. It gives important dynamic kinetic information about the system. This work has been presented as an oral presentation at the CSChE 2010 conference in Saskatoon, Canada.
- 5) A novel metric has been formulated using the computational singular perturbation analysis for the experimental design of kinetic systems which is not sensitivity based as other conventional methods for optimal experimental design. This work has been accepted for an oral presentation at the AIChE Annual Meeting 2011 in Minneapolis, USA.

2. Parameter Estimation for Large-Scale Kinetic Systems¹

2.1 Introduction

Many techniques in process systems engineering have been developed for small-scale, lumped parameter systems, and are not easy to extend to large scale systems. For example, methods for the design of experiments, such as full factorial designs with two levels, fractional factorial designs with two levels, factorial designs with more than two levels, response surface designs, and optimal designs are easily applicable to systems with less than five factors (Ryan, 2007). Also, since kinetic models are basically non-linear in nature (due to the presence of exponential terms), the design of experiments for these systems can be a daunting task due to the presence of multiple suboptimal solutions (non-convexity), or multiple equivalent solutions (a result of non-identifiability). A function is said to be convex on an interval if the graph of the function lies below the line segment joining any two points of the graph. Hence, a non-convex function on an open set has more than one minimum. Non-convexity issues can be solved by the use of global optimization methods (Balsacanto *et al.*, 2007), but non-identifiability is difficult to deal with. Two types of identifiability problems

¹ A version of this chapter has been published. Subramanian, K., Kumar, S., Patwardhan, S. C. and Prasad, V. (2011), Extensions to experiment design and identification algorithms for large-scale and stochastic processes. *International Journal of Advanced Mechatronic Systems*, 2011 – Vol. 3, No.1 pp.3 - 13.

are found: *structural* identifiability which depends on the identifiability of all the parameters given perfect knowledge of the outputs and the model structure, and *practical* identifiability problems which are related to the amount and quality of measured data from the experiments being performed (Walter *et al.*, 1990). In the case of large-scale catalytic kinetic systems, most of the experiments performed are steady-state experiments, and many of the parameters are not identifiable from the measured variables that are available. Consequently, a method must be devised to select a subset of parameters that can be estimated, either by eliminating or grouping the other parameters.

2.1.1 Identifiability

While developing process models, the first step is to determine an appropriate model structure. Once the structure is decided, the next step is to estimate the model parameters. Since, estimation of model parameters is time consuming, it is desirable to first check whether the parameters are structurally identifiable or not. Structural identifiability analysis is always performed a priori and it is possible to distinguish between *global*, *local*, and *un-identifiability*. In contrast, during quantitative estimation of parameters, a posterior analysis is performed, based on collected experimental information, where only *local* identifiability results are obtained and this is known as practical identifiability analysis. In microkinetic reaction models, the structure of the model is pre-determined by the rate expression of each of the species evaluated in accordance with the underlying reaction mechanism. Hence, structural identifiability is not a problem for our case. However, practical identifiability analysis has to be performed. Local identifiability is tested by computing the rank of the sensitivity matrix (Chu & Hahn, 2009).

A parameter θ_0 is said to be locally identifiable if there exists an open neighbourhood of θ_0 containing no other θ which produces the identical output \mathbf{y} . If θ_0 is a parameter point and the sensitivity matrix $\mathbf{S}(\theta) = \partial \mathbf{y} / \partial \theta^T$ has constant rank in a neighborhood of θ_0 , then θ_0 is locally identifiable if and only if $\mathbf{S}(\theta_0)$ has the full column rank. Numerically, identifiability can be determined from the

parameter covariance matrix. Parameters are considered not to be numerically identifiable if the entries in the covariance matrix are large. However, since this matrix can be actually computed only after the parameters have been estimated, hence, the Fisher information matrix is used because its inverse gives a lower bound for the covariance matrix (Ljung *et al.*, 1999). One of the most common model based experimental design methods, the D-optimal design method, involves the evaluation of the Fisher information matrix.

2.1.2 The D-Optimal Design

A detailed description of this method has already been given in Chapter 1. In this chapter, the aim is to develop a modified D-optimal method for this problem in identification and estimation for large-scale kinetic systems. This is important because if all the parameters cannot be identified, and still all the parameters are estimated, the solution might not be unique and stable. Hence, a method should be formulated in order to select just a subset of parameters to be estimated rather than the full set which is not identifiable.

In catalytic kinetic systems, the sensitivity of kinetic parameters is negligible in regions of thermodynamic equilibrium in the operating space, leading to a reduction in the number of identifiable parameters in these regions. The lack of transient data also leads to a smaller number of parameters being identifiable from the steady state data. Consequently, it is important to develop techniques for reduction of the number of parameters to be identified. In this work, the use of hierarchical clustering and principal component analysis is explored for developing a reduced set of parameters for identification, and attempt to rationalize the results of the parameter selection using kinetic arguments. The techniques are demonstrated on the system for the preferential oxidation of carbon monoxide on platinum and rhodium catalysts (Mhadeshwar *et al.*, 2005).

2.2 The Methodology

The well-known D-optimal method for the design of experiments is employed. The Fisher information matrix (FIM) is generated from the parametric sensitivity matrix, S (Prasad *et al.*, 2008), and the D-optimal design seeks the input conditions that maximize the determinant of the FIM, or its logarithm (Walter & Pronzato, 1990). The input conditions to be specified are the variables temperature, pressure, residence time, catalyst surface area per unit reactor volume, and the inlet feed composition (mole fractions of hydrogen, oxygen, carbon monoxide, carbon dioxide, water and nitrogen). The 23 parameters are the pre-exponential factors for the elementary reactions, which include the pathways for hydrogen oxidation, CO oxidation and coupling of hydrogen and carbon monoxide chemistries. Initial estimates of these quantities are obtained from (Mhadeshwar *et al.*, 2004, 2005) for Pt and Rh catalyst, respectively. The outputs whose sensitivities are computed with respect to the parameters to generate the sensitivity matrix and the FIM are CO conversion, selectivity, mole fractions of hydrogen, oxygen, carbon monoxide, carbon dioxide, water, and fractional coverage of the surface species, making 15 outputs in all. Note that not all of these outputs are independent of each other, and the conversion and the selectivity are the most important quantities of interest to practitioners.

To begin with, the system is simulated at many different randomly chosen operating conditions that sample the feasible space uniformly (approximately 22,000 points are chosen for the case of Pt, and 36,000 points for Rh; see (Prasad *et al.*, 2008) for a description of the sampling algorithm). Table 2.2 shows the operating conditions chosen for the analysis. The scaling used for each variable in the table, is the scale (linear or logarithmic) in which the corresponding variable will be dealt. For example, by saying that the scale of P is logarithmic, it is meant that the random sampling is done in the logarithmic scale i.e. $\log_{10} P$.

Table 2-1 Ranges of Operating Conditions

Decision Variable	Symbol	Lower Bound (LB)	Upper Bound (UB)	Scaling
Temperature	$T(K)$	325	700	Linear
Pressure	$P(atm)$	0.1	5	Logarithmic
Residence Time	$\tau(s)$	0.05	5	Logarithmic
Ratio of catalyst surface area to volume	$a_v(1/cm)$	150	15000	Logarithmic
Mole fraction of (H_2)	x_1	0.65	0.85	Linear
Mole fraction of (O_2)	x_2	0	0.15	Linear
Mole fraction of (CO)	x_3	0	0.1	Linear
Mole fraction of (CO_2)	x_4	0	0.3	Linear
Mole fraction of (H_2O)	x_5	0	0.2	Linear
Mole fraction of (N_2)	x_6	0	0.5	Linear

2.2.1 Principal Component Analysis

PCA is a much-used multivariate data analysis method for explorative data analysis, outlier detection, rank (dimensionality) reduction, graphical clustering, classification, and regression (Esbensen *et al.* 2009). It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. The main advantage of PCA is that once these patterns are found in the data, the data can be compressed by reducing the number of dimensions, without much loss of information. A brief description of the method is given below.

1. The mean is subtracted from each of the data dimensions. So, all the x values have \bar{x} (the mean of the x values of all the data points) subtracted, and all the y values have \bar{y} subtracted from them. This produces a data set whose mean is zero.
2. The covariance matrix is then calculated. Covariance is always measured between 2 variables. Hence, for an n -dimensional data set, $\frac{n!}{(n-2)! \times 2}$ different covariance values are calculated and placed in a matrix. This matrix is symmetric about the main diagonal.

$$cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n - 1)} \quad (2.1)$$

For example, from a 3 dimensional data set (dimensions x, y, z), $cov(x, y)$, $cov(x, z)$ and $cov(y, z)$ can be calculated. Then, the covariance matrix has 3 rows and 3 columns, and will be:

$$C = \begin{bmatrix} cov(x, x) & cov(x, y) & cov(x, z) \\ cov(y, x) & cov(y, y) & cov(y, z) \\ cov(z, x) & cov(z, y) & cov(z, z) \end{bmatrix} \quad (2.2)$$

It is to be noted that since $cov(a, b) = cov(b, a)$, the matrix is symmetric about the main diagonal.

3. Eigenvectors and eigenvalues for this matrix are calculated. By this process of taking eigenvectors of the covariance matrix, lines have been

extracted that characterise the data. The rest of the steps involve transforming the data so that it is expressed in terms of these lines.

4. Once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalue, highest to lowest. The eigenvector with the highest eigenvalue is the *principal component* of the data set. The components with lesser significance (small eigenvalues) are ignored so that the final data set will have fewer dimensions than the original. A new matrix is constructed by taking the eigenvectors that want to be kept from the list of eigenvectors, and forming a matrix with these eigenvectors in the columns.
5. The final step in PCA is deriving the new data set which is given by:

$$Final\ Data = \mathbf{A} \times \mathbf{B} \quad (2.4)$$

where \mathbf{A} is the matrix with chosen significant eigenvectors in the columns *transposed* so that the eigenvectors are now in the rows, with the most significant eigenvector at the top, and \mathbf{B} is the mean-adjusted data *transposed*, i.e. the data items are in each column, with each row holding a separate dimension.

PCA transforms our data so that it is expressed in terms of the patterns between them, where the patterns are the lines that most closely describe the relationships between the data. This is helpful because our data is now classified as combinations of contributions from each of those lines.

2.2.2 Hierarchical Clustering

Another popular method for extracting significant data and dimensionality reduction is *hierarchical clustering*. This method starts with a set of distinct points, each of which is considered a separate cluster. The two clusters that are closest according to some metric are agglomerated. This is repeated until all of the points belong to one hierarchically constructed cluster. The final hierarchical cluster structure is called a dendrogram, which is simply a tree that shows which clusters were agglomerated at each step. In this work, hierarchical clustering is

used to identify correlated kinetic reactions that can be eliminated (by choosing to estimate only one of the parameters in the cluster, and keeping the others fixed at their nominal values). The clustering method used is similar to that proposed by Chu & Hahn (2009). In this case, parameters are clustered based on a dissimilarity measure between their sensitivity vectors. The similarity/dissimilarity measure is based on the angle between the sensitivity vectors. This approach does not require the sensitivity vectors to be parallel; however a low (or zero) value of the angle between two parameters means that the parameters are pairwise indistinguishable in terms of their effect on the output with a certain numerical precision. The similarity measure of the effect of two parameters on the output is defined by;

$$\cos \phi_{ik} = \frac{|\mathbf{s}_i^T \mathbf{s}_k|}{\|\mathbf{s}_i\|_2 \|\mathbf{s}_k\|_2} \quad (2.5)$$

where $\phi_{ik} \in [0, \pi/2]$ is the angle between the sensitivity vectors \mathbf{s}_i and \mathbf{s}_k . The value of the similarity ranges from 0 to 1, where a value of unity means that the two vectors are parallel to one another, and the two parameters cannot be distinguished based on sensitivity. A value of 0 refers to the sensitivity vectors of the parameters being orthogonal, that is, the parameters have a distinct effect on the outputs. Using this measure, parameters are hierarchically clustered using the ‘clusterdata’ function in MATLAB. Details about this function are given in the Appendix.

2.2.3 D-Optimal design for PROX chemistry on Pt catalyst

First principal component analysis (PCA) is used as a linear measure of reducing the number of parameter directions. Due to the fact that some of the operating conditions are equilibrium limited, the rank of the FIM varies in the space; however, most conditions lead to an FIM rank of 8, meaning that at most 8 parameters can be identified at these conditions. This places an upper limit on the parameter identification. In our work, PCA is used for reducing the dimension of large data sets and extracting their structural features. In this present work, of the outputs considered, the magnitude of the sensitivity vectors with respect to each output is computed, and PCA is used to find the most important output direction

for sensitivity (using all points). The surface concentration of oxygen is the most important output. PCA is then performed on the sensitivity vector of the surface concentration of oxygen with respect to each kinetic parameter. This indicates that 4 output directions are uncorrelated in a linear sense. However, the important parameter directions are combinations of many different parameters, and do not provide a clear indication of the important chemistry that is coupled and can be reduced.

Consequently, the use of hierarchical clustering is explored. Since the FIM is most often of rank 8, parameters are clustered into 8 clusters. Of these, only the first cluster has multiple (16) parameters, while the other clusters are just single parameters. This means that a large portion of the chemistry is correlated and operates in an indistinguishable manner, and this chemistry can be reduced in terms of parameters to be estimated. Figure 2.1 gives the reaction indices of kinetic parameters, and their frequency of occurrence in the first cluster. This indicates that the seven reactions 2, 3, 4, 12, 15, 16 and 18 have independent parameters, and the other parameters are indistinguishable. Figure 2.2 gives similar information for the 2nd cluster wherein the frequency of occurrence of the 12th reaction was the highest whereas, the frequency of occurrence of other reactions was negligible, thus suggesting that no other reaction is correlated to reaction 12. Similar was the case for the other 6 clusters as well, and no correlation could be established between any of the reactions 2, 3, 4, 12, 15, 16 and 18. Hence, it was concluded that these seven reactions formed singleton clusters, whereas the other remaining reactions were correlated and belonged to the same cluster. These reactions correspond to oxygen adsorption on the catalyst surface, dissociation of adsorbed water to produce adsorbed OH and O, further dissociation of hydroxyl on the surface, dissociation of carbon dioxide on the surface to produce carbon monoxide or COOH (depending on the presence of adsorbed H), further dissociation of COOH to form carbon monoxide on the surface, and the combination of adsorbed CO₂ and OH to form COOH. This indicates that the coupling between the hydrogen and carbon monoxide chemistries is primarily

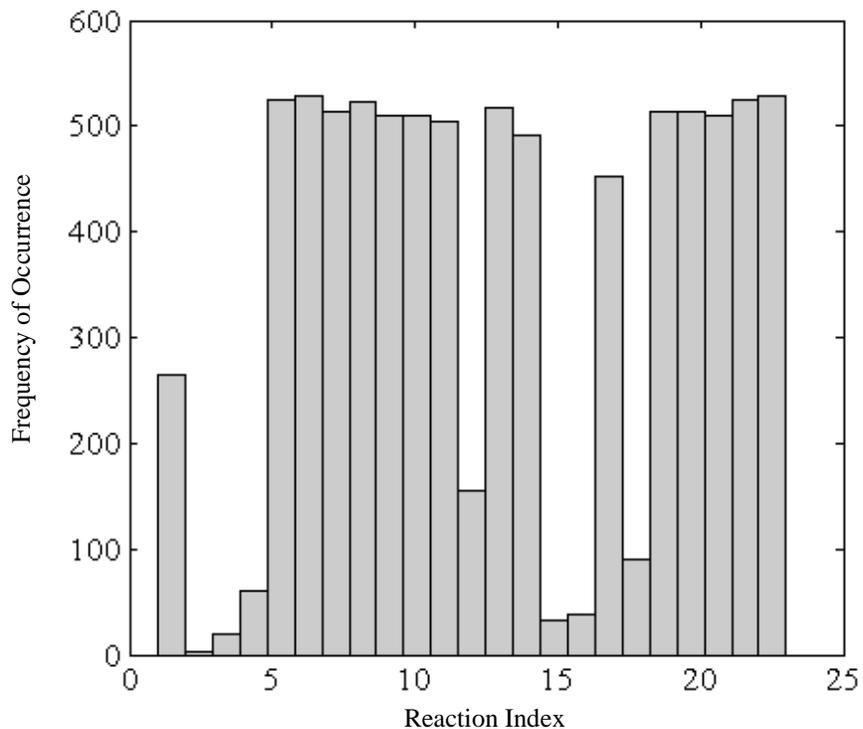


Figure 2.1 Reaction index of kinetic parameters (pre-exponentials) clustered together for Pt. Reaction indices are the same as in Table 1-3. Figure is generated for the first cluster.

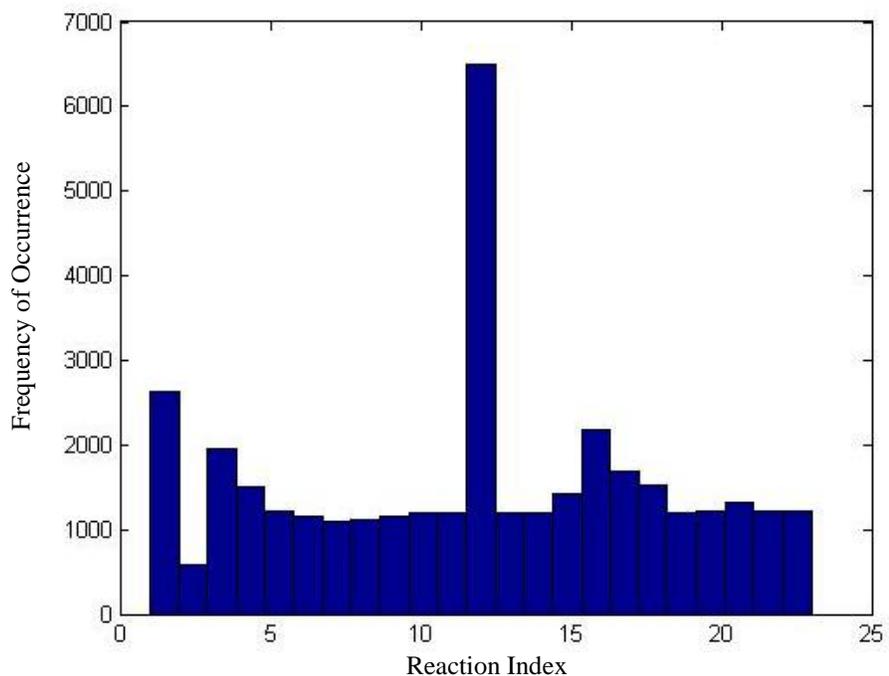


Figure 2.2 Reaction index of kinetic parameters (pre-exponentials) clustered together for Pt. Reaction indices are the same as in Table 1-3. Figure is generated for the second cluster.

important in kinetic investigations for parameter identification, which is in accord with fundamental catalytic studies on this system (Maestri *et al.*, 2008).

Figure 2.3 shows the value of D-optimal index for all the sample points and their frequency of occurrence. PCA and clustering analysis is repeated on a subset of the data, focusing on the operating conditions that score high on the D-optimal metric (i.e., at the right side of the distribution in Figure 2.3). This is to verify that the important chemistry based on all operating conditions is the same as that at the D-optimal conditions. Both the PCA and the clustering analysis produced very similar results in this step, indicating that the same chemistry is sensitive at D-optimal conditions. The D-optimal conditions are found to be: 622 K (temperature), 0.7 atm (pressure), 0.39 s (residence time), 9047 cm⁻¹ (catalyst surface area per unit reactor volume), 0.74 (hydrogen inlet mole fraction), 0.09 (oxygen inlet mole fraction), 0.04 (carbon monoxide inlet fraction), 0.05 (carbon dioxide inlet fraction), 0.03 (inlet water fraction), 0.05 (inlet nitrogen fraction).

Since it is very difficult to ‘dial in’ the exact input conditions of the D-optimal point experimentally, hierarchical clustering is used in an attempt to obtain a region in the operating space where there is a high probability of finding points that score high on the D-optimal metric. This is done by clustering the points in the data subset that correspond to high D-optimal metric scores (Prasad *et al.* 2008), i.e., high values of the determinant of the FIM.

However, it is observed that there is no significant pattern in the region around the D optimal point, and most regions have approximately the same probability of containing a point that scores high on the D optimal metric. Since the two outputs, conversion and selectivity are key outputs, and because measurements of other outputs such as surface coverages are very difficult to obtain in practice, the study described above was also conducted considering the 23 parameters, but only conversion and selectivity as the outputs.

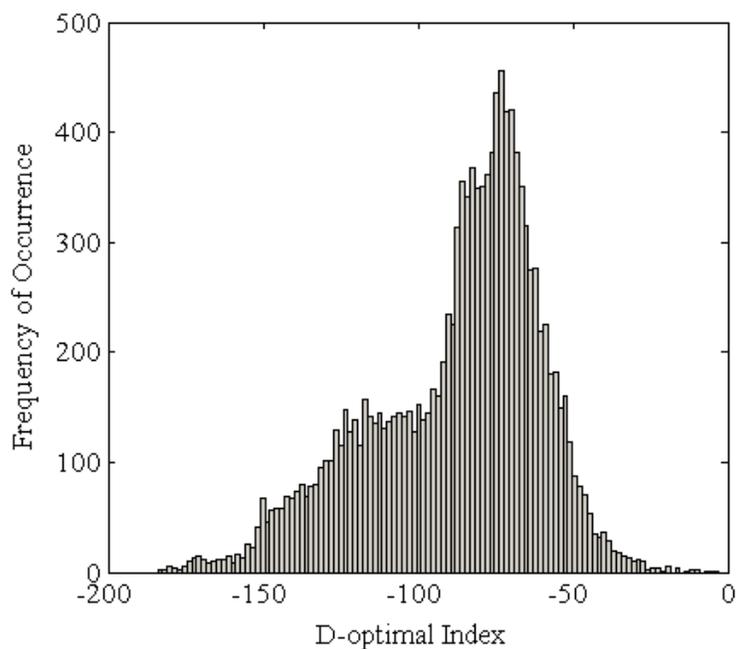


Figure 2.3 D-optimal index for all the sample points and their frequency of occurrence in the cluster for Pt. The variance is large and there are many sample points which score high on the D-optimal index.

The PCA and clustering analysis both provide similar indications about the important chemistry as the case with 15 outputs, and the same conclusion is reached on the probability of regions containing points that score high on the D optimal metric. However, the D optimal point in this case corresponds to: 602 K (temperature), 4.5 atm (pressure), 2.1 s (residence time), 568 cm⁻¹ (catalyst surface area per unit reactor volume), 0.74 (hydrogen inlet mole fraction), 0.05 (oxygen inlet mole fraction), 0.05 (carbon monoxide inlet fraction), 0.02 (carbon dioxide inlet fraction), 0.03 (inlet water fraction), 0.11 (inlet nitrogen fraction). In the case that measurements of other outputs are difficult to make, this D-optimal point may be used for the design. Note that no mass transfer constraints have been taken into account for this study.

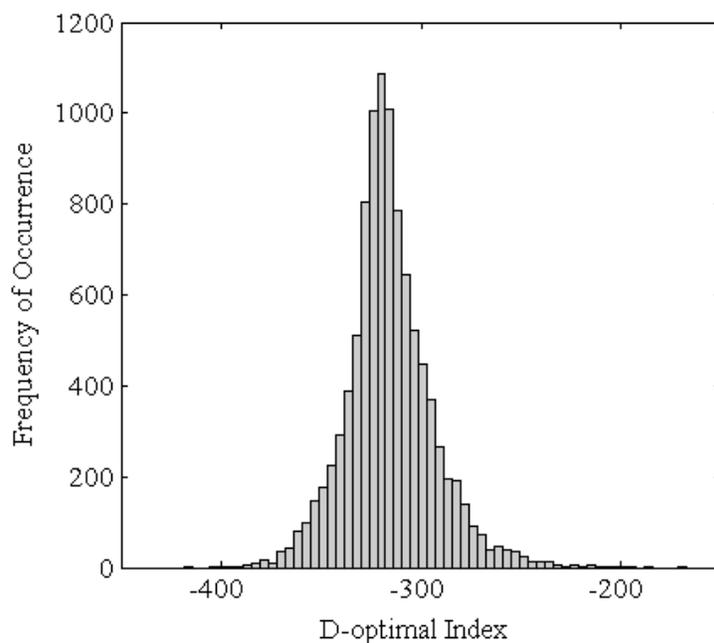


Figure 2.4 Distribution of D optimal points for Rh catalyst

2.2.4 D-Optimal design for PROX chemistry on Rh catalyst

Since the method of experiment design and the kinetic mechanism are the same for Rh catalyst as for the Pt catalyst, the results on the Rh system are not described exhaustively; instead, the discussion focuses on the main points that distinguish this system from the Pt system. For the case with 15 outputs, the principal component analysis provides the same insight about the most significant output, which is the surface concentration of oxygen. Hierarchical clustering, however, indicates that 11 parameters are grouped together and have indistinguishable chemistry; these relate to the reactions with indices 5-9, 13-14, and 19-23. While slightly fewer parameters are clustered together than for the case with Pt, the main features of the independent chemistry remain the same, i.e., the coupling between the carbon monoxide and hydrogen chemistries is important in parameter identification. The D-optimal point in this case is: 512 K (temperature), 2.4 atm (pressure), 1.4 s (residence time), 10937 cm^{-1} (catalyst surface area per unit reactor volume), 0.67 (hydrogen inlet mole fraction), 0.13 (oxygen inlet mole fraction), 0.05 (carbon monoxide inlet fraction), 0.01 (carbon

dioxide inlet fraction), 0.02 (inlet water fraction), 0.12 (inlet nitrogen fraction). Once again, no region in the input space presents significantly higher probabilities of finding points that scored high on the D-optimal metric.

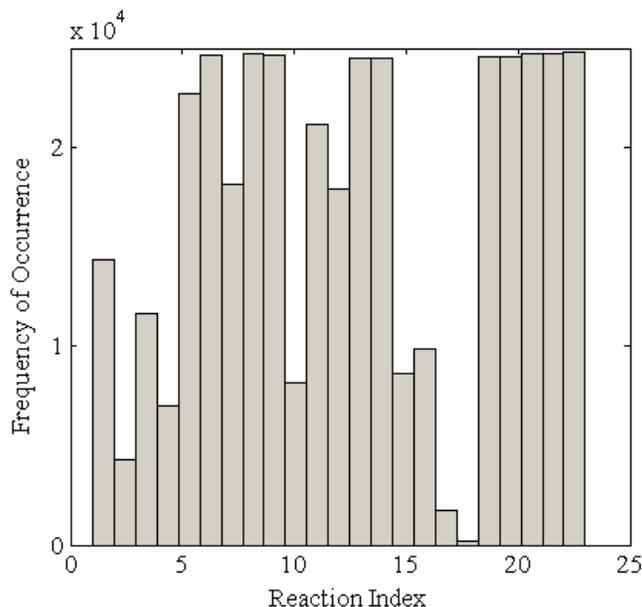


Figure 2.5 Reaction index of kinetic parameters (pre-exponentials) clustered together for Rh

Finally, the analysis for Rh catalyst is repeated with conversion and selectivity being considered as the only two outputs. In this case, the hierarchical clustering provides the interesting result that a much smaller subset of the chemistry is grouped together and is indistinguishable. The reactions that are clustered together have indices 6, and 19-23. This means that many more reactions are independent and kinetically significant for parameter identification than in all the other cases considered so far. This implies that the adsorption behavior of the carbon monoxide and hydrogen families is also important, in addition to the coupling chemistry. Clustering in the input space again does not reveal any regions of high probability, and the D-optimal point is very close to that found for the case with two outputs in Pt. This could be a coincidence, but it may indicate that the optimal point for both systems is essentially the same, if only conversion and selectivity are considered as outputs. Moreover, a comparison of Figures 2-2 and 2-3 suggests that although the distribution of D-optimal points for Rh catalyst shows a smaller variance than with platinum, there

are many number of sample points (input conditions) for Pt where the D-optimal index is much larger than that for Rh. Hence, it can be concluded that the experiments to be performed for the case with platinum as the catalyst will give better data (more sensitive) to estimate the parameters (pre-exponentials) for the PROX catalytic system.

2.3 Conclusions

Most of the issues with process systems engineering of large-scale and complex systems relate to the large number of parameters associated with them. This study attempts to show that modifications are required to traditional methods of experiment design and parameter estimation. Principal component analysis and hierarchical clustering are the methods that are used for this study. The work leads to the following conclusions:

- 1) In particular, it has been shown that hierarchical parameter clustering is very useful and effective in obtaining a reduced set of identifiable parameters for the preferential oxidation of carbon monoxide on platinum and rhodium catalysts.
- 2) The clustering results demonstrate a correlation between the carbon monoxide and hydrogen chemistries for the PROX system. Principal component analysis (PCA) when performed over the full set of data gives the most significant output to be the surface concentration of oxygen.
- 3) A comparison between the D-optimal indices for both the catalysts suggests that experiments performed for the case with platinum as catalyst is more sensitive to parameter estimation of the preferential oxidation system as compared to the case with rhodium as catalyst.

3. Stochastic Nonlinear Optimization for Optimal Design of Experiments

3.1 Introduction

The objective of experimental design is to find experimental conditions which lead to estimating parameters with minimum experimental effort. The most common approach to experiment design is D-optimal design, where the determinant of the Fisher information matrix is maximized. It has been shown with the help of an example in chapter 1 that this design metric depends on the nominal value of the parameters to be estimated. Since the values of the parameters are unknown, a reasonable nominal value of these parameters is chosen and then a D-optimal experimental design is performed. It is noteworthy here that this designed experiment may not yield accurate results if the actual value of the parameters differs too much from the nominal value used for experimental design. A different strategy for optimal experimental design is based on optimization. Optimization can be broadly classified into two categories: (1) *deterministic*, where the objective function is deterministic as well as the method, and (2) *stochastic*, where either the objective function is stochastic (involves randomness or uncertainty), or the method. This is important as these parameter uncertainties can lead to the deterministically identified conditions not being the best ones. Hence, it is very necessary to develop methods in order to minimize the

effect of these uncertainties during experimental design. The optimization is performed to maximize the D-optimal metric for an optimal combination of inlet conditions, i.e.

$$\max_u \det(\mathbf{FIM}(\boldsymbol{\theta})) \quad (3.1)$$

where $\mathbf{FIM}(\boldsymbol{\theta})$ is the Fisher information matrix evaluated over the parameter set $(\boldsymbol{\theta})$, and u is the set of inlet conditions. Deterministic optimization in our problem is done by assuming that the nominal value of parameters is near to the actual value and hence the uncertainty is not taken into account while evaluating the objective function. Although, this chapter focuses on an experimental design methodology based on stochastic optimization by incorporating the uncertainty in the parameter (pre-exponentials) nominal values, it will also be compared with the deterministic results.

This present work therefore uses stochastic optimization techniques to maximize the value of the D-optimal statistical metric averaged over a number of samples. This technique takes the advantage of the fact that probability distributions governing the uncertain parameters are known or can be assumed reasonably. The solution from the stochastic optimization is more robust and provides more realistic predictions than its deterministic counterpart. A gradient-free population-based search method, referred to as particle swarm optimization (PSO), is used in the optimization step in order to avoid getting trapped in local optima. For kinetic systems, multiscale kinetic models have been used quite in the past for DOE, as they are more detailed and fewer assumptions are incorporated as compared to global kinetic models (Prasad *et al.*, 2009).

3.2 Microkinetic and Reactor Model under Uncertainty

3.2.1 Basic Kinetic Models

The kinetic system used for this section of the thesis is the ammonia decomposition model explained thoroughly in section 1.3.1. The reactor model used is that of a steady state plug flow reactor (PFR) explained in section 1.3.3. Table 3-1 shows details of the system such as site density and reactor specifications used for the analysis of this part of the thesis.

Table 3-1 Laboratory scale reactor and catalyst specifications

Term	Symbol	Value
Length	L	1.7 [cm]
Diameter	D	0.41 [cm]

3.2.2 Stochastic Description and Sampling

In chemical reaction engineering, the objective of experimental design is often parameter estimation, or more precisely, estimation of pre-exponentials and activation energies. In this study, the activation energies have been previously calculated from UBIQEP are accurate and do not need to be estimated. The only parameters to be estimated are the sticking coefficients and the pre-exponentials. While employing the D-optimal design method, a reasonable estimate of the nominal values of these parameters should be known beforehand. The need for stochastic modeling arises because the actual value of the parameters might be different from this nominal value, and hence the D-optimal design and parameters estimated by the actual experiment data may end up being far from optimal. The solution for similar problems proposed in literature is to solve a stochastic optimization problem which will take into account the uncertainty in the nominal values of the parameters (Lee *et al.*, 2010).

Microkinetic analysis is a very promising tool for modeling surface reactions; however, in mechanism development and parameter estimation for these models, an issue often overlooked is thermodynamic consistency at both enthalpic and entropic levels (Mhadeshwar *et al.* 2003). Hence, both these inconsistencies distort the underlying equilibrium constant, which affects the prediction of equilibrium states. In our work thermodynamic consistency of the entire mechanism is incorporated through constraints. In the mechanism outlined in Table 1.2, there are 12 irreversible reactions which are essentially 6 reversible reactions. In general, for the i^{th} reversible reaction in the mechanism, the following equations form the basis of the enthalpic and entropic constraints:

$$E_i^f - E_i^b = \Delta H_i \quad i = 1, \dots, 6 \quad (3.2)$$

$$\frac{A_i^f}{A_i^b} = e^{\Delta S_i/R} \quad i = 1, \dots, 6 \quad (3.3)$$

where f and b stand for the forward and backward steps, A is the pre-exponential factor, E is the activation energy, R is the universal gas constant and ΔH and ΔS are the enthalpy change and entropy change of the reversible reaction respectively. Since the parameters to be estimated in this case are only the pre-exponentials (sticking coefficients in adsorption and desorption reactions), the constraint of equation (3.3) will be active while designing the experiment. The UBIQEP method ensures that the values of activation energies used in the microkinetic model are in accordance with equation (3.2). After taking into consideration the thermodynamic constraints, there are 6 parameters whose uncertainty needs to be considered while solving the problem. The nominal values of these parameters are taken from Table 1.2. In order to model the uncertainty in the parameters, a distribution for each of the parameter is assumed and then a sampling procedure is applied. The uncertainties associated with parameters are modeled as exogenous random inputs following uniform distribution added to the corresponding nominal values.

$$\log_{10} A_i = \log_{10} A_{i_0} + \varepsilon_{A_i} \quad (3.4)$$

where ε_{A_i} follows $\mathcal{U}(-1,1)$ and A_i is the pre-exponential of the i^{th} reaction and the A_{i+1} pre-exponential is calculated using equation (3.3). The next step is to pick a certain number of random values from this distribution. This number is chosen to be four, so that the entire random parameter space is easily spanned. Figure 3.1 shows the four regions spanning the random parameter space, each covering $1/4^{\text{th}}$ of the distribution. One random value is chosen from each region, which gives a total of four values in ε_{A_i} . A similar procedure is repeated for all the 6 pre-exponentials to be modeled stochastically. Then all possible combinations of these 6 vectors containing four values each, is generated using the ‘combvec’ function in MATLAB. Details about this function are given in the Appendix.

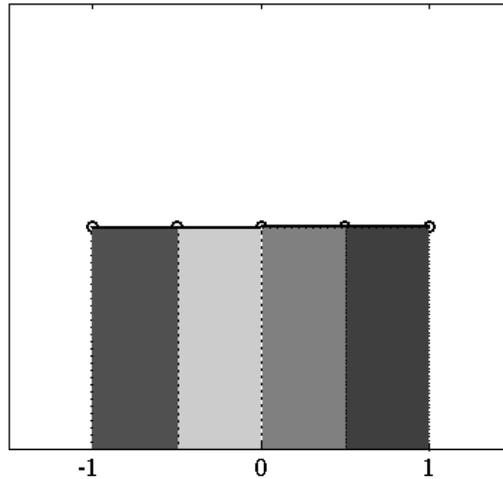


Figure 3.1 Uniform distribution of the order of pre-exponentials; a single random value is selected from each of the four regions

3.3 Stochastic Optimization using Sample Average Approximation and Constrained Particle Swarm Optimization

In order to perform DOE for parameter estimation of the ammonia decomposition model on ruthenium catalyst, the stochastic optimization problem should be formulated in a way as to estimate the inlet conditions while maximizing the modified stochastic D-optimal statistical metric. The decision variables for the optimization problem are temperature (T), pressure (P), ratio of catalyst surface

area to the volume (a_v), residence time (τ) and inlet mole fractions of hydrogen (x_1), nitrogen (x_2), ammonia (x_3) and argon (x_4). The stochastic optimization problem can be formulated as follows:

$$X = [T \quad P \quad a_v \quad \tau \quad x_1 \quad x_2 \quad x_3 \quad x_4] \quad (3.5)$$

$$J = \log(\det(FIM)) \quad (3.6)$$

$$\max_X \quad \mathbb{E}_\varepsilon[J] - \beta(\text{var}(J)) \quad (3.7)$$

$$\text{subject to} \quad LB \leq X \leq UB \quad (3.8)$$

$$A_{eq}X = B_{eq}$$

where FIM is the Fisher information matrix, \mathbb{E}_ε is the expectation operator, (UB) and (LB) are upper bounds and lower bounds on all the decision variables respectively, A_{eq} and B_{eq} are matrices required to formulate the constraint equation for the set of decision variables. Since the optimization problem is stochastic, the expectation of the logarithm of the determinant of the FIM is used in the objective function. The logarithm of the FIM is used for scaling because of the large range of values of determinant of the FIM. The $\text{var}(J)$ used in equation 3.7 is given by:

$$\text{var}(J) = \mathbb{E} \left[(J - \mathbb{E}(J))^2 \right] \quad (3.9)$$

$$= \mathbb{E} \left[J^2 + (\mathbb{E}(J))^2 - 2J \cdot \mathbb{E}(J) \right]$$

$$= \mathbb{E}(J^2) + \mathbb{E} \left[(\mathbb{E}(J))^2 \right] - 2\mathbb{E}(J) \cdot \mathbb{E}[\mathbb{E}(J)]$$

$$= \mathbb{E}(J^2) + (\mathbb{E}(J))^2 - 2(\mathbb{E}(J))^2$$

$$= \mathbb{E}(J^2) - (\mathbb{E}(J))^2 \quad (3.10)$$

The equality constraint of equation (3.8) is that the sum of mole fractions of the gaseous species should be equal to one. Table 3.2 shows the upper bounds

(*UB*) and lower bounds (*LB*) on all the decision variables. The scaling used for each variable in the table is the scale (linear or logarithmic) in which the corresponding variable will be used. For example, when it is said that the scale of P is logarithmic, it is meant that the values that the optimizer treats as decision variables will be in the logarithmic scale i.e. $\log_{10} P$, and when these variables are being used in the model, they will be converted to the normal scale. In equation (3.7), β is a weight applied to the variance of the expectation operator (Kall *et al.* 1994). As described in the previous section, the unknown parameter vector, ε , is modelled as a random vector with known probability distributions. This optimization problem is challenging due to the following reasons:

- The objective function is highly nonlinear.
- Sensitivities have to be calculated in order to find the Fisher information matrix (FIM). This is done by perturbing each parameter one at a time, and then numerically integrating the individual rate expressions for each species for the given values of decision variables and uncertain parameters.
- The expectation operator is also difficult to evaluate because the components of the random vector ε are independent of each other.
- The equality constraint of the sum of mole fractions is difficult to incorporate because it results in a constrained nonlinear optimization problem.

The ratio of catalyst surface area to volume is one of the decision variables for the optimization problem. Due to the fact that catalyst synthesis cannot be done precisely enough to accurately obtain the desired value, its uncertainty must be taken into account. The nominal value for the ratio of catalyst surface area to volume will be the value decided by the optimizer while treating it as a decision variable. The uncertainties associated with calculation of the random variables in this case are modeled as exogenous random inputs following a Gaussian distribution added to the nominal value (value of the decision variable).

$$a_v = a_{v_0} + \varepsilon_{a_v} \quad (3.11)$$

where ε_{a_v} follows $\mathcal{N}(0, (0.15a_{v_0})^2)$, a_v is the ratio of catalyst surface area to the volume and a_{v_0} is the nominal value of the ratio of catalyst surface area to the volume suggested by the optimizer as a decision variable.

The six parameters (pre-exponentials) and one decision variable result in 7 random variables. The number of possible combinations to be considered is $4^7 = 16384$. This is the number required for one objective function evaluation. Hence, reducing the number of possible combinations is necessary to decrease the computational time considerably. The surface reactions in the ammonia decomposition mechanism are of prime interest. Hence, the reactions 3, 4, 5, 6, 7, 8 from Table 3.2 are the reactions which are considered to be affected by randomness in their reaction parameters. This reduces the number of random variables are reduced to four (three reaction pre-exponentials and one decision variable). In order to further reduce the computational time, instead of sampling from four areas of the distribution, sampling is done only from two areas (one to the left and the other to the right of the mean). This further reduces our problem and now the number of possible combinations reduces to just $2^4 = 16$.

3.3.1 Sample Average Approximation

In order to evaluate the expectation operator in equation (3.7), the method that has been used is the sample average approximation (SAA). Sample average approximation (SAA) is a technique that evaluates the expected objective value ($f(X) = \mathbb{E}[J(X, \varepsilon)]$) by generating samples based on the sampling technique explained in the previous subsection and calculating the mean of the objective function values over the 16 combinations (samples) generated:

$$\hat{f}(X) = \frac{1}{16} \sum_{k=1}^{16} J(X, \varepsilon_k) \quad (3.12)$$

where $\hat{f}(X)$ is denoting the estimate of $f(X)$, $J(X, \varepsilon_k)$ is the $\log(\det(FIM))$ and X is the set of decision variables.

3.3.2 Function Evaluation

The sample-based approaches are easy to implement even for a system such as ours, with complex constraints and numerical integration in the function evaluation for the optimization. The function evaluation for our optimization problem starts with evaluating the sensitivity matrix (\mathbf{S}) for every sample during the sample average approximation. In order to do this, each element of the parameter set ($\boldsymbol{\theta}$) is perturbed one at a time, and then the model is solved in order to evaluate the change in the outputs with respect to the perturbed parameter.

3.3.3 Deterministic Optimization

For comparison, we perform deterministic optimization. Here, the uncertainty is not accounted for, and the optimization is formulated as:

$$X = [T \quad P \quad a_v \quad \tau \quad x_1 \quad x_2 \quad x_3 \quad x_4] \quad (3.13)$$

$$J = \log(\det(FIM)) \quad (3.14)$$

$$\max_X J \quad (3.15)$$

$$\begin{aligned} & \text{subject to} \\ & LB \leq X \leq UB \\ & A_{eq}X = B_{eq} \end{aligned} \quad (3.16)$$

3.3.4 Particle Swarm Optimization

The objective function is not in an explicit form with respect to the decision and uncertain parameter vectors; therefore it is difficult to apply gradient-based deterministic optimization methods. In the present work, a global population-based sampling optimization method known as particle swarm optimization (PSO) has been used. This optimization technique is motivated by the social behavior of collection of animals.

Table 3-2 Bounds of decision variables for the optimization problem

Decision Variable	Symbol	Lower Bound (LB)	Upper Bound (UB)	Scaling
Temperature	$T(K)$	500	1000	Linear
Pressure	$P(atm)$	0.1	10	Logarithmic
Residence Time	$\tau(s)$	0.05	5.0	Logarithmic
Ratio of catalyst surface area to volume	$a_v(1/cm)$	150	15000	Logarithmic
Mole fraction of (NH_3)	x_1	0	1	Linear
Mole fraction of (H_2)	x_2	0	1	Linear
Mole fraction of (N_2)	x_3	0	1	Linear
Mole fraction of Ar	x_4	0	1	Linear

It starts with randomly generated swarms of particles (initial guesses for the optimal states) and remembers the best solution found (Jia *et al.*, 2011). The particles move around the solution space with velocities determined by the algorithm, and they move toward the global optimal solution over many iterations. The velocity for each particle is calculated based on the difference between the current particle position and the best particle position globally as well as locally in the swarm. Hence, if the particle is far from the current global best solution, then the particle velocity will be higher and in the direction of the global best solution, and if the particle is near to the global best solution, then the particle velocity will be small and still in the direction of the global best solution. The advantages of PSO are that it does not need to evaluate derivatives of the objective function and constraints, and the number of parameters to adjust is relatively small. The steps involved in the PSO algorithm (Schwaab *et al.*, 2008) are as follows:

1. Initialize the search parameters.
 - a. N_{iter} = number of iterations i to be performed
 - b. N_{pt} = number of particles
 - c. N_d = number of search dimensions
 - d. N_p = number of parameters to be estimated
 - e. C_1, C_2, W_{int} = weight parameters
 - f. x^{min} and x^{max} = lower and upper limits of the variables to be optimized
2. Calculate the maximum particle velocities for the N_p parameters.

$$v_j^{max} = \frac{x_j^{max} - x_j^{min}}{2} \quad (3.17)$$

3. Calculate the initial particle positions and velocities:

$$x_{p,d,j}^i = x_j^{min} + r(x_j^{max} - x_j^{min}) \quad (3.18)$$

$$v_{p,d,j}^i = v_j^{max}(2r - 1) \quad (3.19)$$

where p , d , j , and r denote the particle, the search direction, the parameter index, and a random number in the range $[0,1]$, respectively.

4. Evaluate the objective function for each particle which moves in N_d search dimensions with all N_p parameters.
5. Update $x_{p,j}^{ind}$ and N_{pt} (vectors with dimension of N_p) that contain the best position found by each particle in the swarm.
6. Update x^{glo} (a vector of dimension N_p) that contains the best position found by the whole particle swarm.
7. When the maximum number of iterations is reached or a convergence criterion is met, the PSO search is terminated. Otherwise, update the particle velocities and position using

$$v_{p,d,j}^{i+1} = W_{int} v_{p,d,j}^i + C_1 r_1 (x_{p,j}^{ind} - x_{p,d,j}^i) + C_2 r_2 (x_j^{glo} - x_{p,d,j}^i) \quad (3.20)$$

$$x_{p,d,j}^i = x_{p,d,j}^i + v_{p,d,j}^{i+1} \quad (3.21)$$

8. If the absolute particle velocity is higher than the maximum velocity, then

$$v_{p,d,j}^i = v_j^{max} \text{sign}(v_{p,d,j}^{i+1}) \quad (3.22)$$

9. If the particle position is out of the inequality constraints (upper and lower bounds), then the particle is placed at $x_{p,j}^{ind}$.
10. Increase the iteration counter by 1 and return to Step 4.

3.3.4 Optimization Parameters

The PSO algorithm requires tuning of the algorithm parameters. The choice of the parameters x^{min} and x^{max} was based on Table 3.2. The number of parameters to be estimated, N_p , is the size of vector x , which is equal to 8 (equation 3.5). Other parameters such as C_1 , C_2 , W_{int} , N_{pt} , and N_{iter} were selected based on tuning of trial values for better performance. Apart from these parameters required for the PSO,

the parameter β , which is the variance weight in the objective function also had to be decided. The weight β is necessary for the reformulation of the objective function. A detailed explanation of how N_{iter} and β were decided is given below. Table 3.3 shows the parameter values used in the algorithm.

Table 3-3 PSO Parameter values used in the Optimization Algorithm

Parameter	Value
N_{iter}	80
N_{pt}	20
N_d	1
N_p	8
C_1	$2(0.99)^i$
C_2	$2(0.982)^i$
W_{int}	$(0.99)^i(2r - 1)$

r denotes a random variable in the range $[0, 1]$.

- N_{iter} is the number of iterations to be performed. Every optimization problem to be solved by numerical methods needs a stopping criterion for termination. Many different stopping criteria have been used in the past (Zielinski *et al.* 2007), and can be divided into three major groups: (1) improvement-based criteria, which terminate an optimization run only if relatively very small improvements in the objective function are observed, (2) movement-based criteria, where rather than the improvement, small changes in positions of the particle results in the termination of the optimization run, and (3) distribution-based criteria, which considers the diversity in the population; if the diversity is low, the individuals are close to each other, so it is assumed that convergence has been obtained.

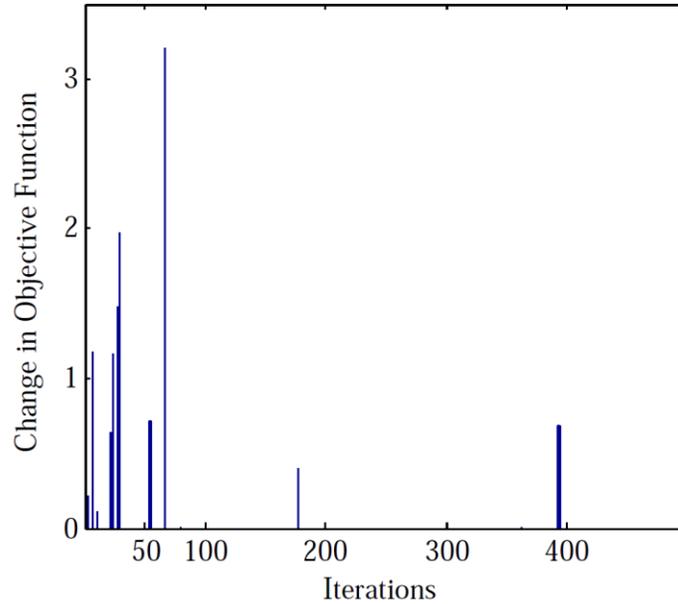


Figure 3.2 Change in Objective Function per iteration for PSO

In this case, the improvement of the objective function has been monitored for a single case. Figure 3.2 shows that at the 66th iteration, the change in the value of objective function was the most significant one, and after that, the improvement in the objective function was at the cost high computational time. An upper limit was put on the number of iterations being performed, which was 80. Although the global optimum may not be attained, this offers a compromise between computational load and accuracy of the solution.

- β is the weight given to the variance of J which is the $\log(\det(FIM))$. In equation (3.7), the expectation operator is being evaluated using SAA, but unless the value of β is known, the value of objective function cannot be evaluated. Once again, parallel runs for the optimization problem were done with a wide range of values of β (from 0 to 100). Table 3.4 shows a comparison of the values of the objective function obtained for six different values of β .

Table 3-4 Comparison of objective function values for different values of β

Type of Optimization	β	Value of Objective Function
Stochastic	0	3.41
	0.01	-29.19
	0.1	-27.88
	1	-28.74
	10	-27.69
	100	-28.82
Deterministic	N/A	-25.56

Figure 3.3 on the next page shows a visual comparison of the same results. It can be seen that the value of the objective function is the maximum in the stochastic case for $\beta = 0$. Note that $\beta = 0$ is the case where the weight imparted to the variance of J is the minimum. The difference in the objective function values for the values of $\beta = 0$ and $\beta = 0.01$ is very high. The main reason for this difference is that the variance of J is very high, and a weight even as small as $\beta = 0.01$ is not able to nullify the effect of this variance on the value of the objective function. Hence, the high variance of the D-optimal metric was neglected and $\beta = 0$ was considered as the best case..

3.3.5 Constrained PSO

One of the major issues while using PSO is handling the constraints. The upper and lower bound constraints can be incorporated into the particle generation step. However, other constraints such as the linear equality constraints are difficult to handle because constraints other than simple bounds cannot be incorporated directly.

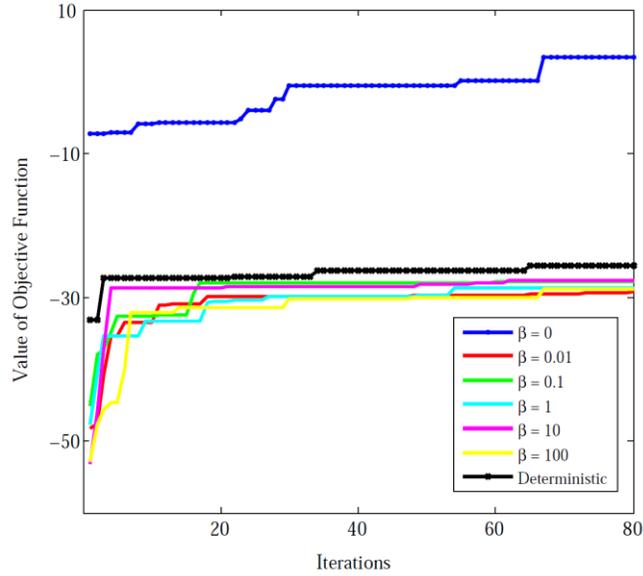


Figure 3.3 Comparison of objective function values for different values of β

In this case, the equality constraint was that the sum of mole fractions should be equal to one. There are two categories of constraint-handling methods in PSO: (1) penalty function methods, which penalize infeasible solutions by applying a penalty based on the magnitude of the constraint violation, and (2) projection methods, where an infeasible solution is projected on to the feasible space. Two approaches based on projection have been tested either manipulating the positions of the particles (*Approach 1*), or by manipulating their velocities (*Approach 2*) as explained below.

Approach 1: Normalizing mole fractions after every iteration so that their sum adds up to one.

$$x_i = \frac{x_i}{\sum x_i} \quad (3.23)$$

Approach 2: In unconstrained PSO, the particles are updated using velocities based on:

$$x(j+1) = x(j) + v(j+1) \quad (3.24)$$

For $x(j)$, the sum of the mole fractions is unity. In the second approach a constraint is applied to the velocities (in the directions of the mole fractions) such that their sum equals zero. This ensures that the sum of $x(j + 1)$ in the directions of mole fractions is unity. This is explained in the equations below:

$$\text{for } j = 1; \quad \sum x_i = 1 \quad (3.25)$$

$$\text{for } j \neq 1; \quad \sum x_i(j + 1) = \sum x_i(j) + \sum v_i(j + 1) \quad (3.26)$$

But the constraint that to be followed all the times is:

$$\sum x_i(j) = 1 \quad (3.27)$$

$$\sum x_i(j + 1) = 1 \quad (3.28)$$

Using equations (3.24), (3.25) and (3.26),

$$\sum v_i(j + 1) = 0 \quad (3.29)$$

The order of velocities is: mole fractions of hydrogen (v_1), nitrogen (v_2), ammonia (v_3) and argon (v_4). In order to keep the sum of the velocities equal to zero after each iteration, the fourth one was calculated using the first three. The mole fractions were then kept within their individual maximum and minimum limits after each update using the projection method. Once the four velocities are calculated followed by calculation of new particle positions, it is found that the mole fraction of hydrogen (x_1) is not within its lower or upper bounds, then the following algorithm is used for calculating new velocities such that all the particles are within their lower and upper bounds.

1. New set of velocities are computed. Their directions remain the same (i.e. towards the best global solution) but the magnitudes are re-calculated by:

$$v_{i_{new}} = r_i v_i \quad i = 1, \dots, 4 \quad (3.28)$$

where r_i is a random variable. When r_i is in the range $[0,1]$, the method is known as contraction because the new velocity will be less than the original value, and when r_i is in the range $[1,2]$, the method is known as expansion because the new velocity will be greater than the original value.

2. The particle's new position is calculated using:

$$x_{i_{new}} = x_i + v_{i_{new}} \quad i = 1, \dots, 4 \quad (3.29)$$

3. The steps 1 and 2 are repeated until $LB \leq x_{i_{new}} \leq UB$, i.e if the particle's new position is within the known lower and upper bounds.

Table 3-5 Comparison of approaches to handle constrained PSO

	Conversion		D-Optimal Value (Objective Function)	
	<i>Approach 1</i>	<i>Approach 2</i>	<i>Approach 1</i>	<i>Approach 2</i>
Run 1	9.45	9.7	-26.5	-26.2
Run 2	7.68	9.9	-26.67	-26.34
Run 3	60.23	62.53	-27.16	-25.56
Run 4	11.77	57.94	-26.93	-25.93
Run 5	9.1	56.86	-26.87	-26.35

3.4 Computational Results

Equation (3.5) gives the decision variables considered for the optimization problem. In this present work the deterministic and stochastic solutions have been compared. It is better to use the solution obtained by stochastic optimization where the uncertainty in such variables has already been accounted for, and hence even if the catalyst is not synthesised to the exact precision, the experiment would still be D-optimal. Table 3.8 gives the best stochastic and deterministic optimal solutions for the set of decision variables. The value $\beta = 0$ is considered for the

stochastic solution and the third run with *Approach 2* from table 3-5 is considered for the deterministic solution.

Table 3-6 Stochastic and Deterministic Optimal Solutions

Result	Stochastic	Deterministic
$T(K)$	812.58	699.52
$P(atm)$	5.71	0.13
$\tau(s)$	0.07	0.063
$a_v(1/cm)$	556.81	964
$x_1(NH_3)$	0.715	0.144
$x_2(H_2)$	0.016	0.001
$x_3(N_2)$	0.084	0.006
$x_4(Ar)$	0.185	0.849

The benefit of the stochastic solution can be quantified using the concept of “value of stochastic solution” (VSS):

$$\begin{aligned}
 VSS &= \max_X \mathbb{E}_\varepsilon[J(X, \varepsilon)] - \mathbb{E}_\varepsilon[J(X^{d*}, \varepsilon)] \\
 &= 3.41 - (-25.56) = 28.97
 \end{aligned}
 \tag{3.26}$$

A VSS value of 28.97 implies that the average performance based on the stochastic solution is an improvement by about 113% compared to that of the deterministic solution. Figure (3.4), shows the value of the objective function in the vicinity of the optimal solution obtained with respect to different decision variables for the deterministic case. This plot and other subsequent plots have been generated by evaluating the objective function for each case (stochastic and deterministic) at samples of decision variables generated in the neighborhood of the respective optimal solutions. Figure (3.5), which is the top view (contour plot) of figure (3.4), displays multiple red vertical lines which suggests that there is no specific set of coordinates or a region where the objective function (D-optimal)

reaches a maximum. The plot shows very little variation along the pressure axis as compared to temperature.

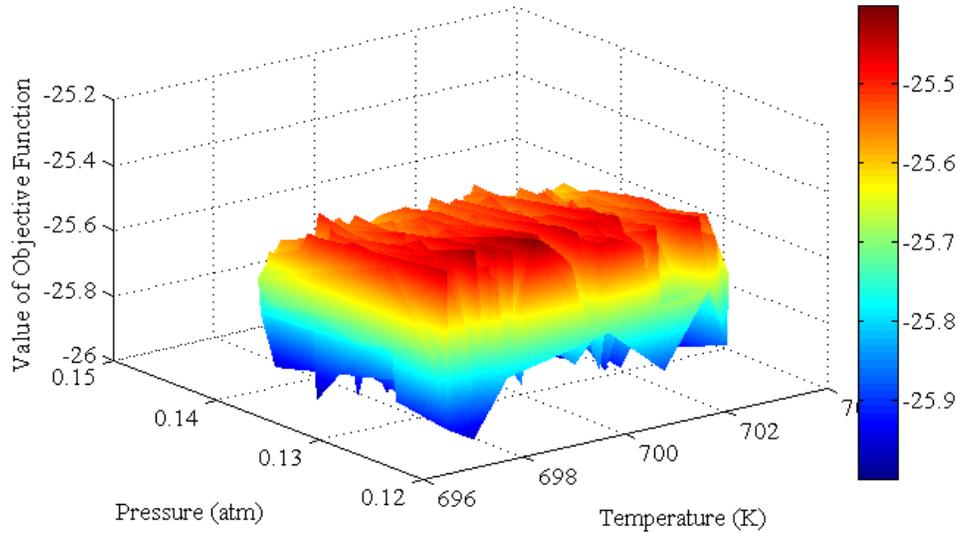


Figure 3.4 A plot showing the variation of objective function with respect to temperature and pressure, for the deterministic case

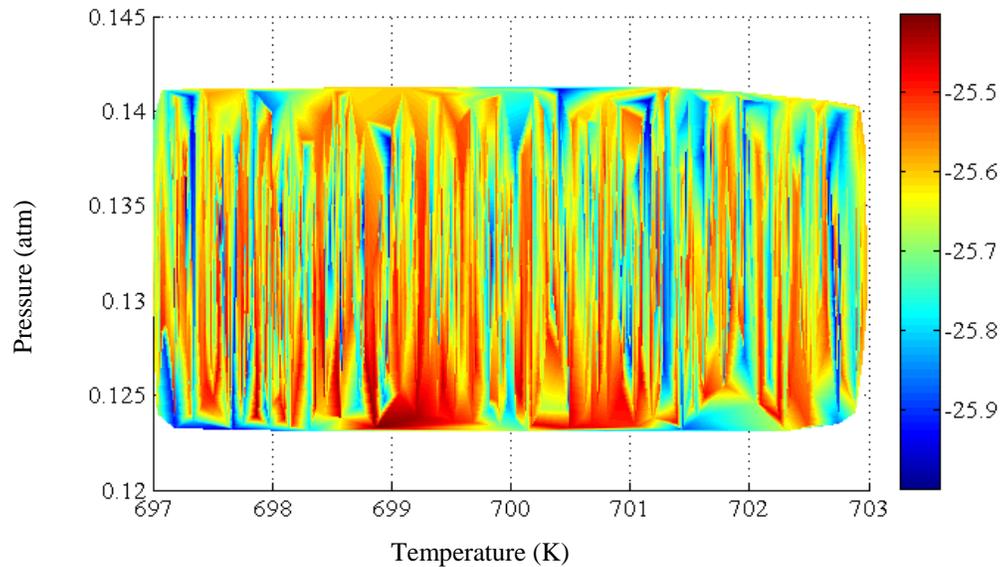


Figure 3.5 Contour plot of the objective function against temperature and pressure. The figure does not reveal any specific optimal region.

Figure (3.6) shows a similar plot for the stochastic case. In contrast to the deterministic case, it can be seen that the plot has less points where the objective

function reaches a local maximum. Figure (3.7), which is a contour plot of the figure (3.6), displays an optimal region around the point (810.3 K, 5.2 atm) in which the experiment can be performed.

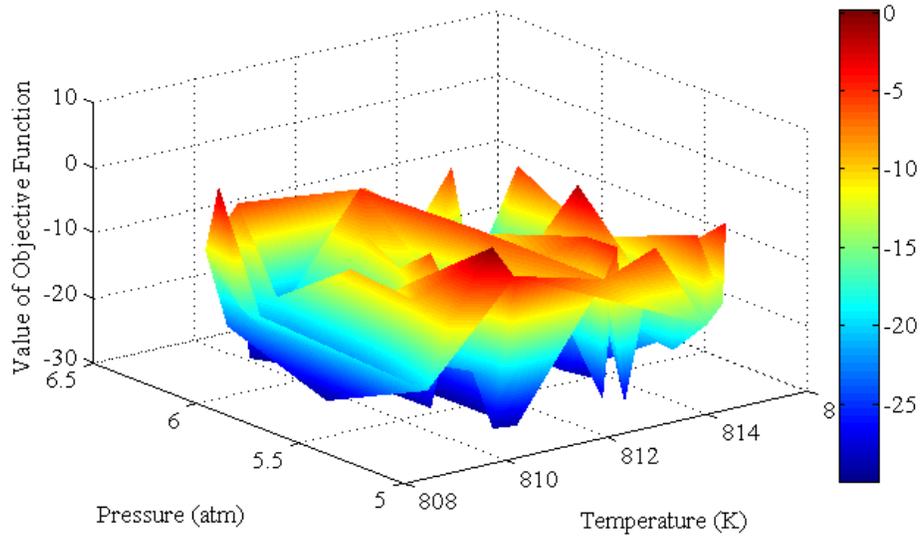


Figure 3.6 A plot showing the variation of objective function with respect to temperature and pressure, for the stochastic case

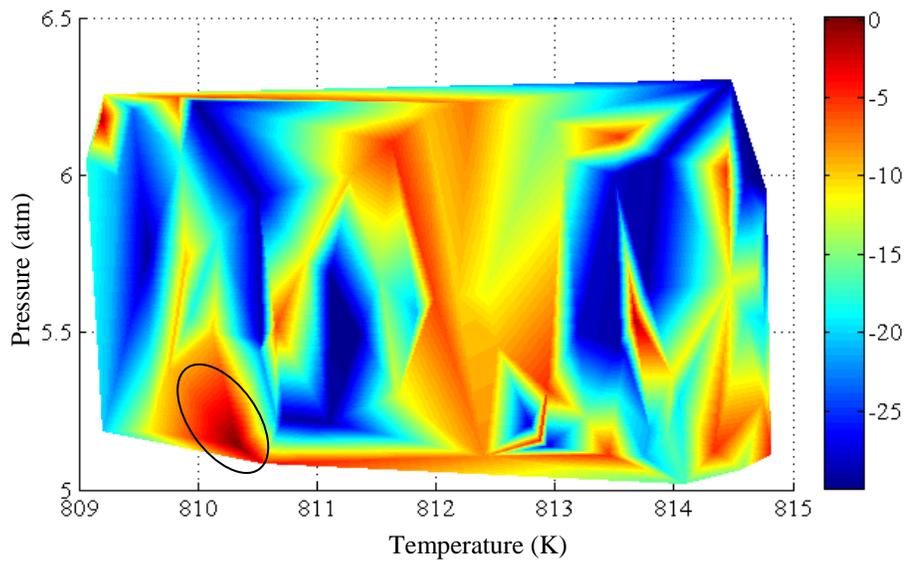


Figure 3.7 Top view of Figure 3.6. The figure reveals optimal region (marked on the plot) where the objective function reaches a maxima

Plots for the deterministic optimization case, exhibits the presence of many valleys and peaks within very short ranges of the decision variables. This kind of sensitivity towards variables for the deterministic case can be attributed to the non-linear structure of the problem. The stochastic optimization which identifies regions, provides a more robust solution.

It can be seen that in comparison to its deterministic counterpart, in the stochastic case, small changes in the values of the variables have a negligible effect in the value of the objective function, and hence the experiment can be performed at values even in the vicinity of the optimal operating point, if not at the exact point, which can be hence referred to as an optimal operating region.

Table 3.6 presents a comparison of the operating conditions for the experiment between the stochastic and deterministic cases. It can be seen that although the operating conditions are much more favorable for the reaction system, still the stochastic results are considered to be better. The reason is that the problem under consideration is the design of experiments followed by parameter estimation rather than the ease of performing experiments. The data collected after performing the experiment will be good for estimating the pre-exponentials, which is the main objective of this study.

3.5 Conclusions

In this study, a comparison between deterministic and stochastic optimization techniques also suggests that stochastic results are much better as compared to the deterministic results. This chapter leads to the following conclusions:

- 1) The optimization has a VSS value of 28.97 and hence, the stochastic optimization gives better results. Since it is not always possible to dial in the exact values of variables solved by the optimizer while performing the experiment, uncertainty should be taken into account.
- 2) Nonlinear constrained particle swarm optimization was incorporated and different approaches for the same have been tried. Projection methods have been used rather than the penalty function methods.

- 3) The variance of the objective function is very high and hence even a small weight of 0.01 could not minimize the effect of considering the variance.
- 4) Stochastic optimization gives optimal operating regions rather than points given by deterministic optimization.

4. Computational Singular Perturbation for Design of Experiments

4.1 Introduction

In chemical reaction engineering researchers may sometimes be interested only in a few species or reactions or more specifically in particular reaction chemistries. Designing experiments for such an objective needs a model-based design, where firstly a method should define important reaction chemistries dynamically, and secondly, an appropriate design method should select input conditions in order to perform experiments to estimate parameters for those chemistries. Model reduction techniques help in getting the most vital information at any point of time. A model reduction technique is used in this work which can also focus on particular reactions and species. This method is based on computational singular perturbation (CSP) analysis of the system at every instant. CSP is a mathematical technique which has been used in the past to identify the important reactions and species by breaking the system into active, exhausted and dormant modes (Lam *et al.*, 1994). It has mainly been used in the investigation of combustion models and chemistries (Brad *et al.*, 2007, Lee *et al.*, 2007, Lu *et al.*, 2001 and 2008, Bykov *et*

al., 2008, Massias *et al.*, 1999, Valorani *et al.*, 2003 and 2006, Gou *et al.*, 2010, Polifke *et al.*, 1998).

An experiment design metric is required for dynamic experiments, where both inputs and outputs may not be single values (say, a constant temperature, a conversion) but complex time profiles of the same variables. Other methods for DOE, such as response surface methodology and alphabetical optimal designs, can become computationally very expensive while calculating the sensitivities at times. Hence, the method we use is a search over a grid of input conditions in order to find the best input condition in the grid which can fulfill our objective of focusing on a particular reaction.

4.2 Computational Singular Perturbation Theory

The chemical kinetics of any system can be represented in two ways: (1) using a global model, where there is one rate expression that represents the whole reaction mechanism taking into account the rate determining step (RDS) and the most abundant reaction intermediate (MARI), and (2) using a microkinetic model, where all the individual rate expressions for the intermediate reactions are solved together without any assumption of the RDS or the MARI. Since, the global model has many underlying assumptions related to it, it is preferable to use microkinetic model to represent a kinetic system. For catalytic systems, the number of intermediate reactions may sometimes increase up to the order of hundreds (Bernaerts *et al.*, 2000), making the microkinetic model more complex. Although our interest may lie in only a few species and reactions, the detailed reaction model usually involves a much larger number of species.

There are some species which are important intermediates in the whole reaction mechanism, but their concentrations are considerably low. These are referred to as *radicals* in the context of computational singular perturbation (CSP). Among the elementary reactions in the system, some may be fast and some may be slow. CSP decomposes this kinetic model into three types of modes at every instant: (1) *fast* or *exhausted* modes, which have negligible contribution to the dynamics at that instant, (2) *slow* or *active* modes, which have considerable

contribution to the dynamics, and (3) *dormant* modes, which are neither exhausted nor active and represent conservation laws (Lam *et al.*, 1994). Although the aim of the CSP method is modal decomposition of the reaction system, it can also be used as a model reduction tool, where a simplified model is obtained by neglecting contributions from the fast modes and considering contributions from the slow and dormant modes. It retains the essential features of the full reaction system.

Conventional methods used to obtain simplified models apply steady-state approximation to the appropriate radicals, partial-equilibrium approximation to the fast reactions, and ignore the very slow reactions. These steps, however, are not easy to apply for large scale catalytic systems. Moreover, the results obtained are expected to be valid only in some limited domain of initial and operating conditions in some limited interval of time. The CSP method, in contrast, can generate time-resolved simplified kinetic models for large reaction systems by a rigorous systematic mathematical algorithm (Zagaris *et al.*, 2004). CSP not only performs the modal decomposition and can be used as a model reduction tool, but a careful inspection of the numerical CSP data highlights important chemistries and species at each time scale. In the next subsection, the CSP method is discussed in brief, followed by design of experiments using CSP. A detailed description of the complete CSP method can be found in Lam *et al.*, (1993).

4.2.1 Mathematical Representation of the Kinetic System

A reaction system consisting of R elementary chemical reactions and N species (concentrations of the chemical species) is considered. The total number of unknowns is N which is represented by a N-dimensional column vector,

$$y = [y^1, y^2, y^3 \dots, y^N]^T \quad (4.1)$$

Any chemical reaction model is represented mathematically in the form of rate expressions of individual reactions given by ordinary differential equations (ODEs) for the rate of change of concentrations of species. The governing equations for the species vector y , can hence be written as follows

$$\frac{dy}{dt} = g(y) \quad (4.2)$$

where the column vector g is the reaction rate for all the species and consists of contributions from each of the R elementary reactions:

$$g = [g^1, g^2, g^3, \dots, g^N]^T = \sum_{r=1}^R S_r F^r \quad (4.3)$$

where S_r and F^r are the stoichiometric (column) vector and the reaction rate of the r^{th} elementary reaction, respectively. Each additive term of equation (4.3) corresponds to the overall rate of the formation or decay of the corresponding species in vector form.

4.2.2 Basis Vectors

The basic idea of CSP is to project the R terms in equation (4.3) into N linearly independent modes, and group these N modes into a fast group and a slow group. In general, the amplitude of fast modes decays rapidly with time. The simplified kinetics model is obtained when the contribution of the fast group to the dynamics becomes small enough to be neglected. The CSP procedure described below was developed by Lam and Goussis (1993). The process of projecting and grouping of the terms is done by the use of a special set of basis vectors. Since g is an N -dimensional vector, it can be expressed in terms of any set of N linearly independent basis vectors ($a_1(t), a_2(t), \dots, a_N(t)$). According to the CSP algorithm, a corresponding set of (row) vectors ($b^1(t), b^2(t), \dots, b^N(t)$) also has to be calculated, satisfying the orthonormal relations (Lam *et al.*, 1993):

$$b^i \cdot a_k = \delta_k^i \quad i, k = 1, 2, \dots, N \quad (4.4)$$

where δ_k^i is the $[N \times N]$ identity matrix. For the linear case, these basis vectors are the left and right eigenvectors of the Jacobian of g with respect to y . Hence, once we chose our linearly independent basis (column) vectors ($a_1(t), a_2(t), \dots, a_N(t)$) and the set of (row) vectors ($b^1(t), b^2(t), \dots, b^N(t)$), equation (4.3) can now be alternatively expressed in terms of these basis vectors:

$$g = \sum_{i=1}^N a_i f^i \quad (4.5)$$

where f^i is the amplitude of g in the direction of a_i . Each of the N additive terms in equation (4.5) is interpreted as representing a *reaction mode*. Therefore, a_i and f^i are the effective stoichiometric vector and the effective reaction rate of the i^{th} *mode*, respectively. The dot product of b^k with g gives us the amplitude of each mode:

$$f^k = b^k \cdot g \quad (4.6)$$

Once the set of basis vectors (either the a_i 's or the b^k 's) is chosen, f^i 's are readily computed from equations (4.4) and (4.6), respectively (Valorani *et al.* 2006). Since, it is unlikely that the magnitude will be the same for each of the basis vectors a_i at every instant; hence, without loss of generality, all the a_i vectors are scaled such that their orders of magnitude are constant with time. Consequently, the order of magnitude of the contribution of the i^{th} mode in equation (4.5) is dependent only on the order of magnitude of f^i . The time evolution of the f^i 's is given by:

$$\frac{df^i}{dt} = \sum_{k=1}^N \Lambda_k^i f^k \quad (4.7)$$

where

$$\Lambda_k^i = \phi_k^i + \frac{db^i}{dt} \cdot a_k = \phi_k^i - b^i \cdot \frac{da_k}{dt} \quad i, k = 1, 2, \dots, N, \quad (4.8)$$

$$\phi_k^i = b^i \cdot J \cdot a_k \quad i, k = 1, 2, \dots, N, \quad (4.9)$$

and J is the Jacobian of g with respect to y . The time evolution of f^i 's is controlled solely by Λ_k^i which, according to equation (4.8), is completely determined by $a_i(t), b^i(t)$ (and their time derivatives) in addition to J . If Λ_k^i was diagonal, the modes would be completely uncoupled from each other, and if the system was linear, J would be a constant matrix. In this case the basis vectors

would be constant through time, and would be the right and left eigenvectors of \mathbf{J} . The resulting Λ_k^i would be the diagonal matrix (or the Jordan form) of the eigenvalues, $\lambda(i)$'s. Consequently, the amplitude of each of the uncoupled modes f^i would evolve with its own characteristic time scale. If $\lambda(i)$ was real and negative, the amplitude f^i of the i^{th} mode would decay exponentially toward zero and become eventually exhausted for some time $t \gg |1/\lambda(i)|$ (Marzouk *et al.* 2007). The algorithm for the linear case is: whenever the amplitude of the currently fastest mode falls below a user-specified threshold, the term representing that mode is dropped from equation (4.5), thus generating a simplified equation to be integrated (Massias *et al.* 1999).

4.2.3 The Classification of Fast and Slow Modes

Lam & Goussis have extended their CSP algorithm to non-linear problems as well. Unlike the linear case, the basis vectors and \mathbf{J} are not constant. However, the left and right eigenvectors of \mathbf{J} , β^i and α_i , are always defined, and can be computed at any time. The reciprocal of the absolute value of $\lambda(i)$ is denoted by $\tau(i)$ and is called as the current time scale of the reaction modes. The CSP algorithm needs a precise classification of fast and slow modes at every time sample. If the current time sample is t_s , the group of M modes whose time scales are shorter than t_s , ($|\tau(m)| < t_s$, with $m = 1, 2, \dots, M$) are considered fast modes, and all others are considered slow modes. The time derivatives of $a_i(t)$, $b^i(t)$ is calculated numerically using the values of the left and right eigenvectors of \mathbf{J} at every time sample. Λ_k^i is then calculated using equations (4.8) and (4.9).

The main effect of non-linearity is that if eigenvectors of \mathbf{J} were used directly as the trial set of basis vectors, the resulting Λ_k^i would not be diagonal. The non-zero off-diagonal elements cause mixing of the modes, and as a consequence, the fast modes do not decay and become small as in the linear case. Although, these eigenvectors and eigenvalues of \mathbf{J} , cannot be used directly as the set of basis vectors, they are used as a guide for choosing a set of trial basis vectors. Lam & Goussis have devised “refinement” algorithms which can

generate from any reasonable trial set of basis vectors an improved set which has less mode mixing than before.

4.2.4 The CSP Refinement Procedure

The refinement algorithm described here follows Goussis *et al.* (2006). The indices m and n are used to refer to the fast subspace ($m = 1, 2, \dots, M$) and I and J are used to refer to the slow subspace ($I, J = M + 1, \dots, N$). Indices i and k refer to the whole N -dimensional space ($i, k = 1, 2, \dots, N$). The right-hand side of equation (4.7) is divided into a fast and a slow group. The fast and the slow equations written separately are:

$$\frac{df^m}{dt} = \sum_{n=1}^M \Lambda_n^m f^n + \sum_{J=M+1}^N \Lambda_J^m f^J \quad m = 1, 2, \dots, M, \quad (4.10)$$

$$\frac{df^I}{dt} = \sum_{n=1}^M \Lambda_n^I f^n + \sum_{J=M+1}^N \Lambda_J^I f^J \quad I = M + 1, \dots, N. \quad (4.11)$$

The refinement procedure of the trial basis vectors is described below, where the set of basis vectors $a_i^0(M)$ and $b_0^i(M)$ is an ‘‘improvement’’ over the original trial set, a_i and b^i :

$$b_0^m(M) \equiv b^m + \sum_{J=M+1}^N p_J^m(M) b^J \quad m = 1, 2, \dots, M, \quad (4.12)$$

$$a_J^0(M) \equiv a_J - \sum_{n=1}^M a_n p_J^n(M) \quad J = M + 1, \dots, N, \quad (4.13)$$

$$b_0^I(M) \equiv b^I - \sum_{n=1}^M q_n^I(M) b_0^n(M) \quad I = M + 1, \dots, N, \quad (4.14)$$

$$a_m^0(M) \equiv a_m + \sum_{J=M+1}^N a_J^0(M) q_m^J(M) \quad m = 1, 2, \dots, M, \quad (4.15)$$

where $p_j^m(M)$ and $q_m^l(M)$ are given by:

$$p_j^m(M) = \sum_{n=1}^M \tau_n^m(M) \Lambda_j^n \quad m = 1, 2, \dots, M, \text{ and } j = M + 1, \dots, N. \quad (4.16)$$

$$q_m^l(M) = \sum_{n=1}^M \Lambda_n^l \tau_m^n(M) \quad m = 1, 2, \dots, M, \text{ and } l = M + 1, \dots, N. \quad (4.17)$$

where $\tau_n^m(M)$ is the inverse of $\omega_n^m(M)$ which is the $M \times M$ principal submatrix of Λ_k^i . The use of refined $b_0^m(M)$ and $a_j^0(M)$ obtained with $q_m^l(M)$ purifies the slow modes. The amplitude of the modes can now be calculated using these refined basis vectors:

$$f_0^i = b_0^i(M) \cdot g \quad i = 1, 2, \dots, N. \quad (4.18)$$

Since for linear cases, the amplitude of the fast modes decay exponentially, hence the use of refined basis vectors for the non-linear case enables the fast f_0^m 's to exponentially approach to smaller values for $t \gg \tau(M)$. More the number of refinements, lesser is the mode mixing. However, since increase in the number of refinements results in a considerable increase in the computational time, for most of the kinetic systems, one or two refinements is sufficient (Goussis *et al.*, 2006).

4.2.5 Radical Pointers and Participation Indices

Following Lam & Goussis (1994), a $N \times N$ matrix $Q(M)$ can be formed as follows:

$$Q(M) = \sum_{m=1}^M a_m b^m \quad (4.19)$$

This is called as the fast subspace projection matrix. When evaluated with refined basis vectors, it is denoted by $Q_0^0(M)$. It is then decomposed into its M components:

$$Q_0^0(M) = \sum_{m=1}^M Q_{m,0}^0 \quad (4.20)$$

where

$$Q_{m,0}^0 = a_m^0 b_0^m, \quad m = 1, 2, \dots, M. \quad (4.21)$$

$Q_{m,0}^0$ is called the m^{th} fast mode projection matrix. The *radical pointer* of the m -th mode, $Q_m(i)$, is given by:

$$Q_m(i) = \text{the } i\text{-th diagonal element of } Q_{m,0}^0, \\ i = 1, 2, \dots, N. \quad (4.22)$$

The quantity $Q_m(i)$ is dimensionless, and its sum over all N components is unity.

The *participation index* of the r -th reaction in the m -th mode, $P_m(r)$, is given by:

$$P_m(r) = (s_r)^{-1} \cdot Q_{m,0}^0 \cdot s_r, \quad r = 1, 2, \dots, R, \quad (4.23)$$

It is a dimensionless number of the order of unity. Pointers computed from “guessed” trial basis vectors without any refinement at all produces erroneous results. Hence, only pointers computed from refined basis vectors can provide reliable information.

4.3 The Reaction System and CSP Analysis

4.3.1 Basic Kinetic Models

The CSP analysis was performed for the ammonia decomposition model explained in chapter 1, on both the batch and flow reactor models. However, the PFR was modeled as 10 CSTRs in series. Argon was used as the inert gas. A typical plug flow reactor could be a tube packed with catalyst, and is better known as packed bed reactor (PBR). Table 4.1 shows the reactor specifications.

Table 4-1 Lab scale plug flow reactor specifications

Term	Symbol	Value
Length	L	1 [cm]
Diameter	D	0.3175 [cm]

4.3.2 The CSP Analysis

The ammonia decomposition model, as detailed above, has 10 species and 12 reactions. Hence, the vector y for our case is:

$$y = [C_{H_2}, \theta, \theta_H, C_{N_2}, \theta_N, \theta_{NH}, \theta_{NH_2}, \theta_{NH_3}, C_{NH_3}, C_{Ar}]^T \quad (4.24)$$

The CSP tool is most often used by the people investigating combustion models and chemistries in order to get rough estimates for the important reaction and species at desired time scales, and hence, most of their cases use batch reactor for modeling and consequent analysis. Since the second part of our work incorporates the design of experiments which is essentially is estimating inlet conditions at every time sample, hence the PFR model is preferred over the batch reactor model for the analysis.

As mentioned earlier, the reactor used was a PFR modeled as 10 CSTRs in series. The first step to CSP analysis was to identify the number of exhausted modes at each time instant. Based on the time scales for each mode, the number of exhausted modes is evaluated at each instant as explained in section (4.2.3). Figure 4.1 shows how the number of exhausted modes changes with time. It can be seen that after $t \approx 10^{-4}s$, all the modes have been exhausted. This suggests that the system is dynamic only for times $t < 10^{-4}s$. We also see that for this time interval, the number of exhausted modes experiences a shift from 5 to 6. It is important to note that when the exhausted modes are 5, the number of active modes is also 5, since there are no dormant modes in our system.

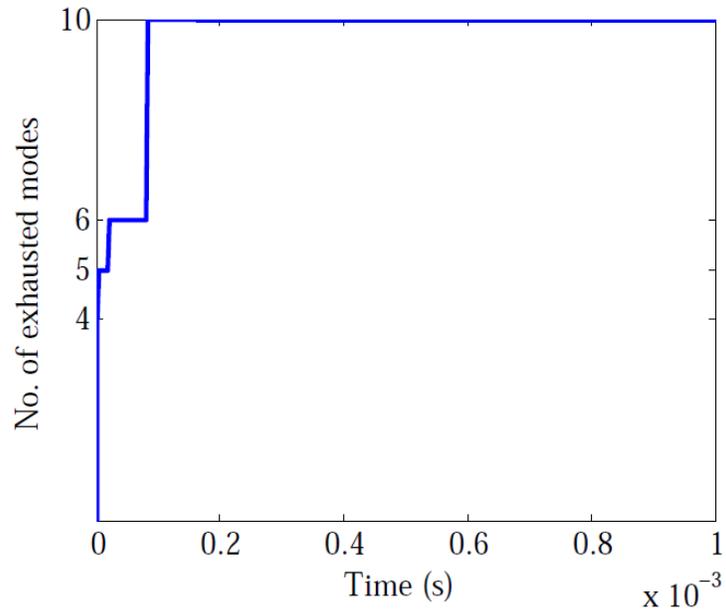


Figure 4.1 Time variation of the number of exhausted modes for the ammonia decomposition model. After 0.1ms, all the modes get exhausted and none contribute to the dynamics

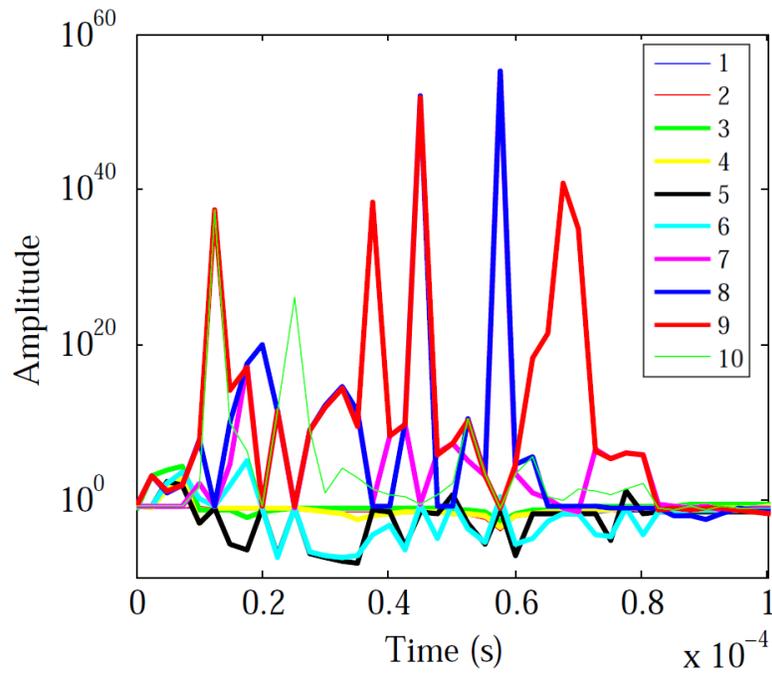


Figure 4.2 Time variation of amplitude of each mode for the ammonia decomposition catalytic system. Modes 8 and 9 appear to be an active mode.

When the number of exhausted modes is 6, the number of active modes is 4. However, the active modes are decided only by the magnitude of their amplitudes, which can be different at every instant. Figure 4.2 shows how the amplitude of each mode changes with time. The plot has been generated only for $t < 10^{-4}s$, and hence gives us an idea of the transient state of the system. We can see that at every instant, different combinations of modes are active. However, figure 4.2 suggests that mode 8 and mode 9 are active modes most of the times for $t < 10^{-4}s$. These conclusions are strengthened by Figure 4.3, which shows the time scales of different modes. It can be seen clearly in this figure that, for $t < 2 \times 10^{-5}s$, when the number of exhausted modes are 5, it is modes 1, 2, 3, 4 and 5 that are exhausted, while the rest are active and are participating in the dynamics of the system. We can see that mode 6 becomes exhausted for $t > 2 \times 10^{-5}s$. Each reaction is participating in each mode, but the amount of participation is decided by the participation indices of a particular reaction in a particular mode. In order to find the important reactions contributing to the dynamics, we will have to take a closer look at the participation indices of the active modes. Figure 4.4 shows the participation indices of all the 12 reactions in mode 9, which is an active mode for $t < 10^{-4}s$. We cannot qualitatively say which reaction is important at which time instant; hence, a detailed quantitative analysis at every time sample is required. We can however conclude that reactions 5 and 6 (NH* formation) are not at all important in this particular mode. However, these reactions can be important and significantly contributing in some other active mode.

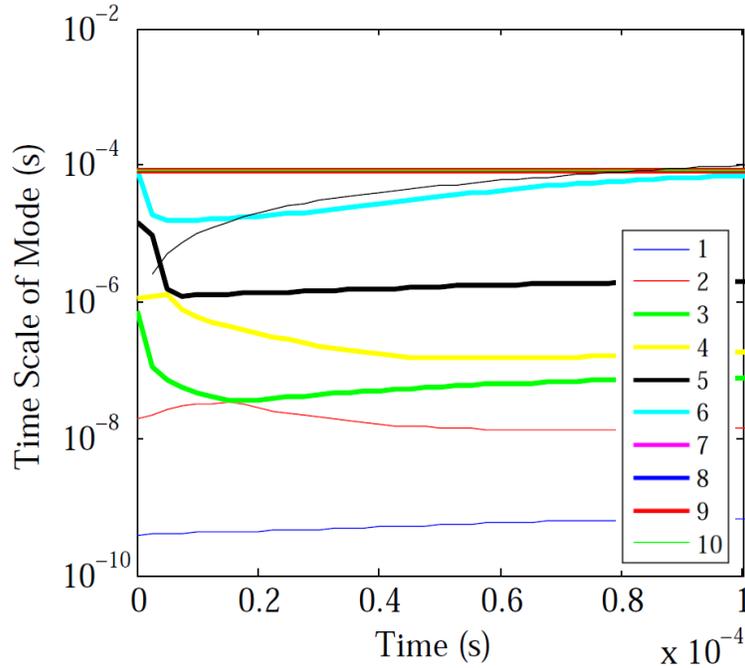


Figure 4.3 Time scale of each mode for the ammonia decomposition catalytic system combined with a PFR model. The thin black line indicates the current time sample. Curves lying above this line are the active modes for the corresponding time sample

Another important quantity given by CSP analysis is the *radical* pointer. There are some species whose concentrations are considerably low but are believed to be important intermediates in the whole reaction scheme. These are referred to as *radicals*. According to the CSP theory, these radical pointers are numerical quantities that suggest which species can be considered as quasi-steady, and hence these ODEs can be approximated as algebraic equations. Figure 4.4 is a plot which shows the value of radical pointer for each species in a particular mode (mode 9 in this case). Once again, we cannot qualitatively say as to which species is a *radical* at which time instant, and these conclusions can be taken only by quantitative analysis at each instant. However, we can definitely say that species 6 and 8 (θ_{NH} and θ_{NH_3}) should not be considered as *radicals*.

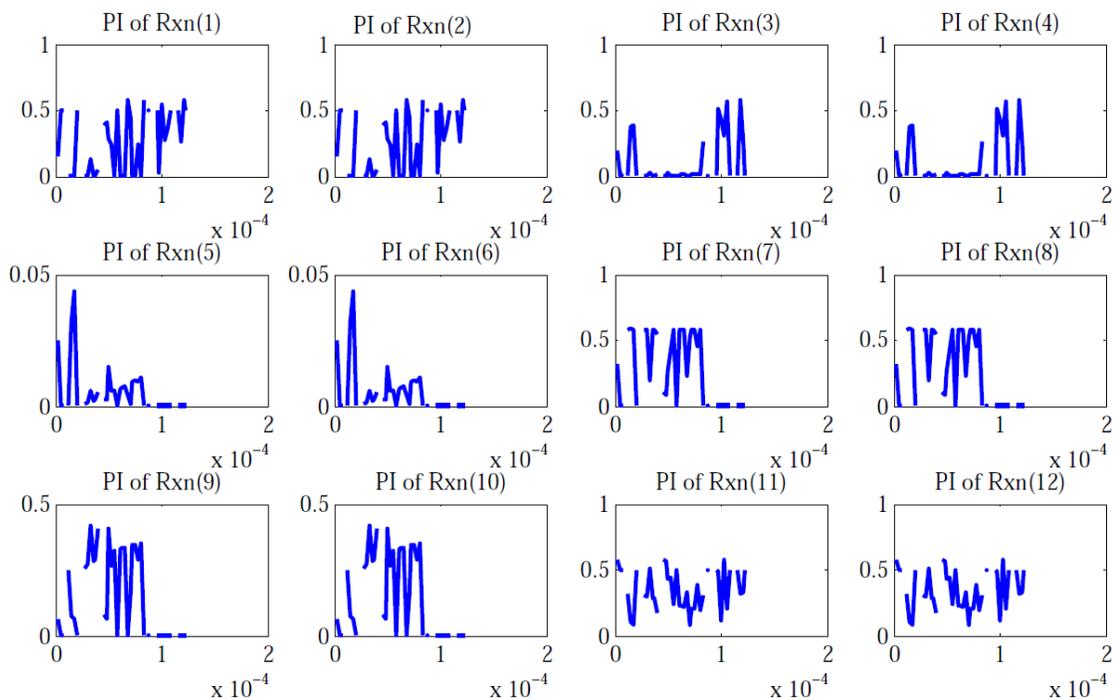


Figure 4.4 Participation Index of each reaction in Mode 9, which is an active mode for most of the times. Broken curves suggest that CSP data fails to give information at some time samples. Reactions 5 and 6 have negligible participation in this mode, however they might be participating in some other active mode

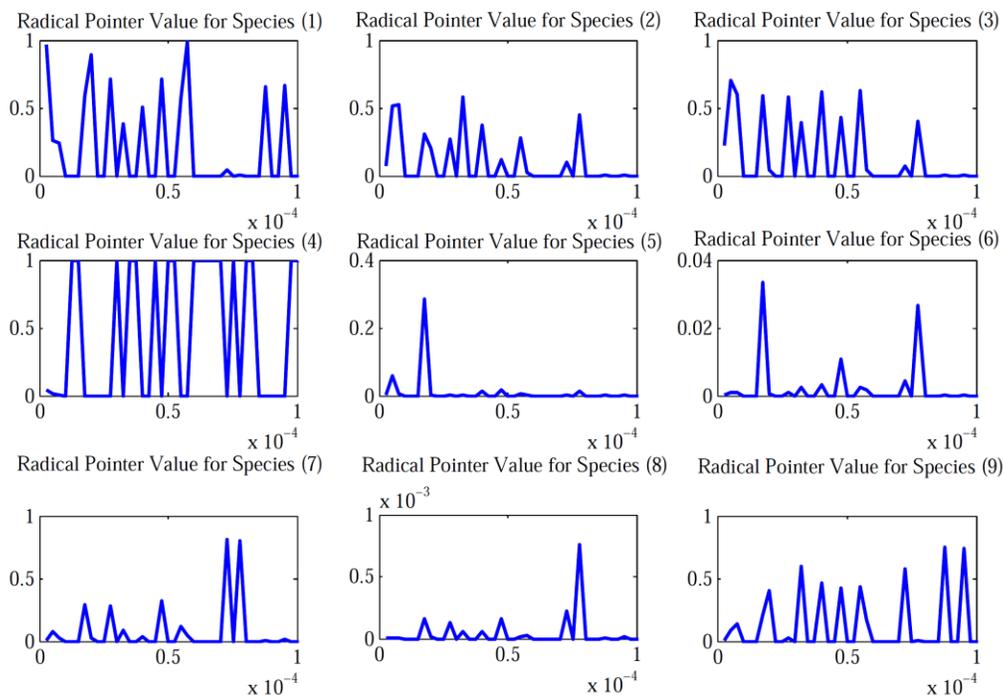


Figure 4.5 Radical Pointer for each species of mode 9, which is an active mode for most of the times. We can see that species 5, 6, 7 and 8 have negligible probability of being the radical in this mode, however they can be radicals in some other active mode

4.4 Design of Experiments using CSP Analysis

This section describes the method proposed for designing experiments for parameter estimation (pre-exponentials) of a desired reaction from the given kinetic system. The search for the optimal inlet conditions for an experiment is driven by an optimization technique. The optimizer solves for a unique combination of inlet conditions in order to maximize the participation of the desired reaction in the dynamics of the reaction system. The objective function used by the optimizer is a proposed reaction metric based on the participation indices and amplitudes of active modes calculated by the CSP.

4.4.1 The Reaction Metric DOE

In order to accomplish the task of DOE to maximize the impact of particular reaction chemistry, we have formulated a suitable metric. If we want maximize the impact of the r^{th} reaction in the dynamics of the system, then the optimizer should maximize the following reaction metric:

$$RM = \sum_{t=t_i}^{t_f} \sum_{i=1}^{N-M} \frac{f_i(t) \times PI_i^r(t)}{\sqrt{\sum f_i(t)^2}} \quad (4.25)$$

where $N - M$ = the number of active modes, $f_i(t)$ = the amplitude of i^{th} mode at time t , $PI_i^r(t)$ = participation index of the r^{th} reaction in the i^{th} mode at time t . We are trying to calculate a quantity which equals the participation of the r^{th} reaction in each active mode multiplied by the amplitude of the active mode and normalized over the amplitudes of all the active modes. Hence, this is a dimensionless quantity. This quantity, when summed up over all the time samples, gives the reaction metric. Conventional optimal DOE methods calculate the sensitivities of outputs with respect to the desired parameters. In these methods, the choice of desired parameters is driven by prior knowledge of the system and remains constant throughout the process of designing the experiment. In contrast, this reaction metric takes the advantage of the numerical analysis of CSP, thus identifying important reactions (parameters) at every time sample.

4.4.2 The Search Method

The ideal search method is the employment of a global optimization technique maximizing the reaction metric. However, the computational cost of a single CSP analysis is itself very high, and if we use a global optimization technique over it, the computational cost will increase many folds. Hence, in order not to increase the computational cost any further, we select a grid search over the input variables. The procedure is simple and is described below:

1. Select a grid of any two inputs
2. Evaluate the following reaction metric at each point in the grid
3. Select the point in the grid which has the maximum value of the reaction metric

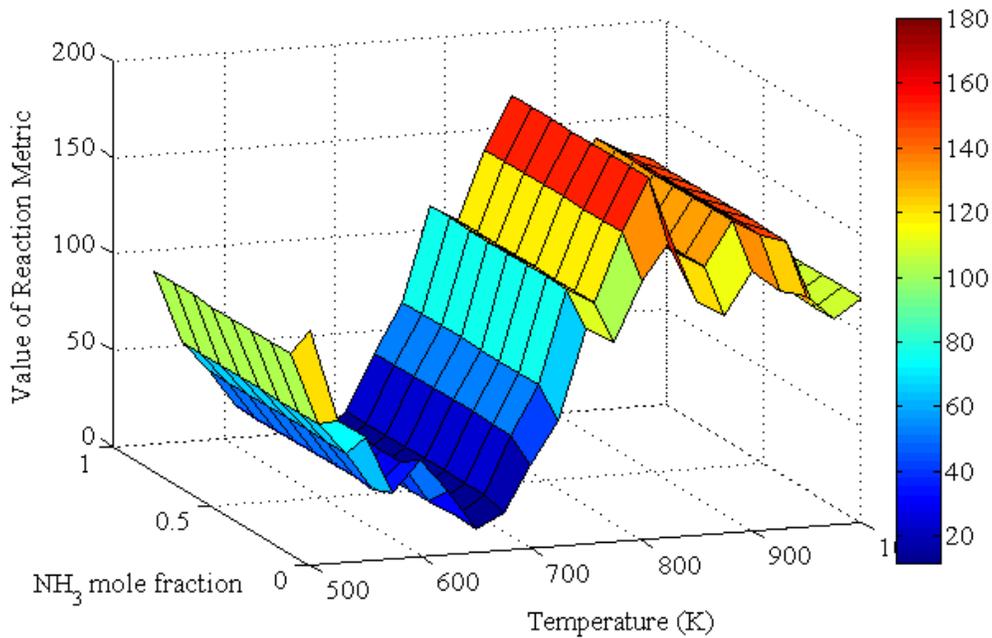


Figure 4.6 Variation in the value of reaction metric with respect to temperature (K) and NH₃ mole fraction. The plot reaches a peak at 830K, but is not affected much by the NH₃ mole fraction

4.4.3 Application to Ammonia decomposition model

The method proposed in the previous section was applied to our system of ammonia decomposition. The input variables chosen for gridding the input space and their bounds are shown in Table 4.2. Figures 4.5 and 4.6 are plots that have

been generated by evaluating the reaction metric at each point in the selected grid. These results have been derived for maximizing the impact of the reactions 9 and 10 of the Table 1.2. This is the reaction step which involves the formation of NH_3^* .

Table 4-2 Bounds of Decision Variables

Decision Variable	Lower Bound (<i>LB</i>)	Upper Bound(<i>UB</i>)
$T(K)$	500	1000
$x_1(\text{NH}_3)$	0	1
$x_2(\text{H}_2)$	0	1

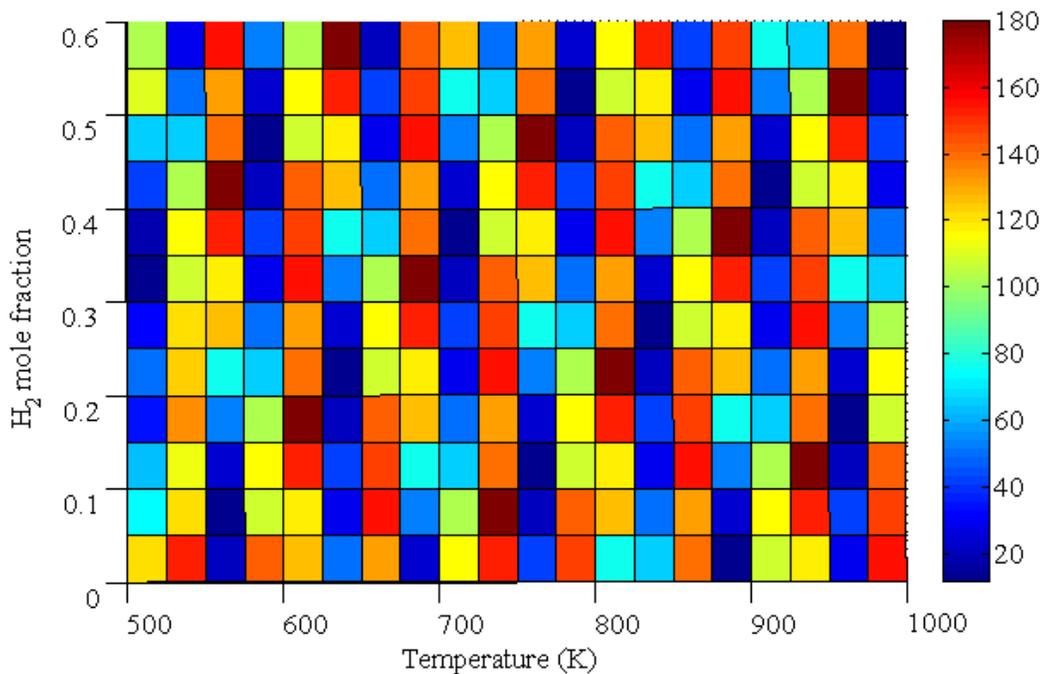


Figure 4.7 Variation in the value of reaction metric shown with respect to NH_3 and H_2 mole fraction. The plot shows no specific shape and no direct relationship can be deduced.

Note that in Figure 4.5, NH_3 inlet mole fraction has no considerable impact on the value of reaction metric. However, temperature has very smooth curve with a distinct minima and maxima. In Figure 4.6, neither the NH_3 inlet mole fraction, nor the H_2 inlet mole fraction shows any visible trend. Although there are points

in the grid where the value of reaction metric reaches high values, no clear relationship can be deduced.

4.5 Conclusions

The design of experiments using the computational singular perturbation is advantageous in many ways. The conclusions of this chapter are:

- 1) A two-variable-at-a-time (TVAT) approach for DOE, has been tried which although does not speak much about the interactions between other potential inputs, but the method is outlined and can be formulated with all the variables included.
- 2) DOE for maximizing the impact of the formation of NH_3^* has been performed. A new reaction metric formulated is considered to maximize the impact of a desired reaction on the dynamics of the catalytic system. Experiments when performed at such conditions will provide data better for the parameter estimation of that desired reaction.
- 3) This method for experimental design is not statistical, as are most of the conventional methods for optimal design. This method is driven by the calculation of important species, reactions and their contributions to the current dynamics of the system.
- 4) CSP alone, as a mathematical tool, gives us considerable amount of information on the kinetically important reactants and species at any instant of time.

5. Concluding Remarks and Future Work

5.1 Concluding Remarks

While this work focuses on different variations required for experimental design for large scale catalytic systems, the results and the methodologies are applicable to all fields of science and engineering that may need good parameter estimation. The D-optimal design metric has been used in this study owing to its easy and straightforward procedure. Three different problems have been addressed in this study. Large scale catalytic systems may have the problem of non-identifiability, i.e. not all parameters can be identified. In such cases, this study presents a comparison between principal component analysis, singular value decomposition and hierarchical clustering. Hierarchical clustering is the preferred methodology because PCA fails to give information about uniquely coupled chemistries at times. The system under investigation for this identifiability study was the preferential oxidation system for hydrogen production for fuel cells. The study was performed for models of two different catalytic systems and the method yielded different solutions for both of them.

The next important problem discussed is the problem of experimental design when there exists uncertainty in the nominal values of parameters used for D-optimal design. This problem is also important because many times, the equipment involved in performing experiments may not be able to dial in the exact values of inlet conditions as suggested by the experimental design. In this part of the study, stochastic optimization has been proposed as a potential solution and the stochastic results have been compared with the deterministic results. Constrained particle swarm optimization has been used for optimization. Contrary to the stochastic results, the deterministic results show that the value of the objective function is highly sensitive to very small changes in the values of decision variables (inlet conditions), and hence stochastic results are accepted. Also, the stochastic case yields a much larger value (VSS of 28.97 on logarithmic scale) as compared to the deterministic solution and hence, the solution provided by stochastic optimization is considered to be better for estimating parameters. The system under investigation for this part of the study was the ammonia decomposition catalytic system on Ru.

The third and the last problem discussed is the formulation of an experimental design strategy when parameter estimation of only some user-specific reactions is to be performed. The solution proposed is based on computational singular perturbation, a popular mathematical tool used mainly for identifying active modes and important reactions and species at any desired time scale. This tool has been used only for investigating combustion chemistries. In this study, this tool has been used for experimental design, where a reaction metric is maximized such that a desired reaction has maximum impact on the dynamics of the system. The special reaction metric is evaluated using the results of the CSP analysis for the catalytic system. A mere CSP analysis of the system alone gives us insights about the important reaction chemistries at any time scale. A grid based search mechanism was used to identify the optimal value of operating conditions for the experiment. Once again the system under investigation was the ammonia decomposition catalytic system on Ru.

5.2 Future Work

5.2.1 Stochastic Sampling

As explained in chapter 3, the use of stochastic techniques for experimental design is very necessary. However, rigorous stochastic sampling was not a part of this study because of huge computational time requirements. The number of samples was reduced from 16384 to 16. In order to get more accurate results, the same procedure can be repeated taking all the 16384 samples into consideration without any reductions. The computational cost will increase tremendously but the predictions will be much more accurate.

5.2.2 Input Sequence design for Design of Experiments

Input sequence design for the purpose of design of experiments is beneficial for systems where the observed system changes are very quick. If the input sequence is designed dynamically for the experiment, rather than performing the experiment at just one inlet condition, the fast dynamics of the system can also be given due consideration. One of the novel methods proposed is similar to the model predictive control techniques where the control sequence is designed for a *control horizon* and a *prediction horizon*. The idea for an input sequence design is based on obtaining the current value of inlet conditions by solving, at each sampling instant, a finite horizon D-optimal experiment design based on sensitivity matrix evaluations using important reactions and species given by CSP analysis at that sampling instant. The optimization then yields a time-variant sequence for the inlet conditions, which when used for experiments, will give data tailored for optimal precision at every sampling instant.

6. Bibliography

- Balsa-Canto, E., Rodriguez-Fernandez, M. and Banga, J. R. (2007). Optimal Design of Dynamic Experiments for Improved Estimation of Kinetic Parameters of Thermal Degradation. *Journal of Food Engineering*, 82(2), 178-188.
- Beltrán-Oviedo, T. A., Batyrshin, I. and Domínguez, J. M. (2009). The Optimal Design of Experiments (ODOE) as an Alternative Method for Catalysts Libraries Optimization. *Catalysis Today*, 148(1-2), 28-35.
- Bernaerts, K., Versyck, K. J. and Van Impe, J. F. (2000). On the Design of Optimal Dynamic Experiments for Parameter Estimation of a Ratkowsky-type Growth Kinetics at Suboptimal Temperatures. *International Journal of Food Microbiology*, 54, 27-38.
- Brad, R. B., Tomlin, A. S., Fairweather, M. and Griffiths J. F., (2007). The Application of Chemical Reduction Methods to a Combustion System Exhibiting Complex Dynamics. *Proceeding of the Combustion Institute*, 31, 455-463.
- Bykov, V., Gol'dshtein, V., and Maas, U. (2008). Simple Global Reduction Technique based on Decomposition Approach. *Combustion Theory and Modelling*, 12(2), 389-405.
- Chu, Y. and Hahn, J. (2009). Parameter Set Selection via Clustering of Parameters into Pairwise Indistinguishable Groups of Parameters. *Industrial & Engineering Chemistry Research*, 48(13), 6000-6009.
- Esbensen, K. H., Wold, S. and Geladi P. (1987). Principal Component Analysis, *Chemometrics and Intelligent Laboratory Systems*, 3(2) 37-52.
- Gou, X., Sun, W., Chen, Z. and Ju, Y. (2010). A Dynamic Multi-timescale Method for Combustion Modeling with Detailed and Reduced Chemical Kinetic Mechanisms. *Combustion and Flame*. 157, 1111-1121.
- Goussis, D. and Valorani, M. (2006). An Efficient Iterative Algorithm for the Approximation of the Fast and Slow Dynamics of Stiff Systems. *Journal of Computational Physics*, 214(1), 316-346.
- Hunter, W. G. and Box, G. E. P. (1965). Sequential Design of Experiments for Nonlinear Models, in *Proceedings IBM Scientific Computing Symposium: Statistics*, J.J. Korth, ed., IBM, White Plains, 113-137.

- Jia, D., Zheng, G., Qu, B. and Khan, M. K. (2011). A Hybrid Particle Swarm Optimization Algorithm for High-dimensional Problems. *Computers & Industrial Engineering*, In press.
- Kall, P. and Wallace, S. W. (1994). Stochastic Programming. *John Wiley & Sons Inc.*, Chichester, England.
- Lam, S. H. (1993). Using CSP to Understand Complex Chemical Kinetics. *Combustion Science and Technology*, 89(5), 375-404.
- Lam, S. H., and Goussis, D. A. (1994). The CSP method for Simplifying Kinetics. *International Journal of Chemical Kinetics*, 26(4), 461-486.
- Lee, J. C., Najm, H. N., Lefantzi, S., Ray, J., Frenklach, M., Valorani, M. and Goussis, D. A. (2007). A CSP and Tabulation-based Adaptive Chemistry Model. *Combustion Theory and Modelling*. 11(1), 73-102.
- Lee, C. J., Prasad, V. and Lee, J. M. (2010). Stochastic Nonlinear Optimization for Robust Design of Catalysts. *Industrial & Engineering Chemistry Research*, 50(7), 3938-3946.
- Ljung, L. (1999). System Identification: Theory for the User, 2nd ed., *Prentice Hall*, Upper Saddle River, NY.
- Lu, T., Ju, Y. and Law, C. K. (2001). Complex CSP for Chemistry Reduction and Analysis. *Combustion and Flame*, 126:1445-1455.
- Lu, T. and Law, C. K. 2008. A Criterion based on Computational Singular Perturbation for the Identification of Quasi Steady State Species: A Reduced Mechanism for Methane Oxidation with NO Chemistry. *Combustion and Flame*, 154, 761-774.
- Lucas H.L. and Box, G. E. P. (1959). Design of Experiments in Non-Linear Situations. *Biometrika* 46, 77-90.
- Maestri, M., Vlachos, D., Beretta, A., Groppi, G. and Tronconi, E. (2008). Steam and Dry Reforming of Methane on Rh: Microkinetic Analysis and Hierarchy of Kinetic Models. *Journal of Catalysis*, 259(2), 211-222.
- Marzouk, Y., Debusschere, B., Najm, H. N., Goussis, D. A., Valorani, M. and Frenklach, M. (2009). Time Integration of Reacting Flows with CSP Tabulation. 4 pages, 2nd *International Conference in Model Reduction in Reacting Flows*. South Bend, Indiana, USA, April 2009.

- Massias, A., Diamantis, D., Mastorakos, E. and Goussis, D. A. (1999). An Algorithm for the Construction of Global Reduced Mechanisms with CSP data. *Combustion and Flame*, 117(4), 685-708.
- Mhadeshwar, A. B., Wang, H. and Vlachos, D. G. (2003). Thermodynamic Consistency in Microkinetic Development of Surface Reaction Mechanisms. *The Journal of Physical Chemistry B*, 107(46), 12721-12733.
- Mhadeshwar, A. B. and Vlachos, D. G. (2004). Microkinetic Modeling for Water-Promoted CO Oxidation, Water-Gas Shift, and Preferential Oxidation of CO on Pt. *The Journal of Physical Chemistry B*, 108(39), 15246-15258.
- Mhadeshwar, A. B. and Vlachos, D. G. (2005). Hierarchical Multiscale Mechanism Development for Methane Partial Oxidation and Reforming and for Thermal Decomposition of Oxygenates on Rh. *The Journal of Physical Chemistry. B*, 109(35), 16819-35.
- Polifke, W., Geng, W. and Dobbeling, K. (1998). Optimization of Rate Coefficients for Simplified Reaction Mechanisms with Genetic Algorithms, *Combustion and Flame*, 113:119-135.
- Prasad, V. and Vlachos, D. G. (2008). Multiscale Model and Informatics-Based Optimal Design of Experiments: Application to the Catalytic Decomposition of Ammonia on Ruthenium. *Industrial & Engineering Chemistry Research*, 47(17), 6555-6567.
- Prasad, V., Karim, A. M., Arya, A. and Vlachos, D. G. (2009). Assessment of Overall Rate Expressions and Multiscale, Microkinetic Model Uniqueness via Experimental Data Injection : Ammonia Decomposition on Ru / γ -Al₂O₃ for Hydrogen Production. *Industrial & Engineering Chemistry Research*, 48, 5255-5265.
- Prasad, V., Karim, A. M., Ulissi, Z., Zagrobelny, M., and Vlachos, D. G. (2010). High Throughput Multiscale Modeling for Design of Experiments, Catalysts, and Reactors: Application to Hydrogen Production from Ammonia. *Chemical Engineering Science*, 65(1), 240-246.
- Pukelsheim F. (1993). Optimal Design of Experiments, 2nd Edition, *John Wiley and Sons Inc.* New York.
- Ryan, Thomas P. (2007). Modern Experimental Design. 1st Edition, *John Wiley and Sons Inc.* Hoboken, New Jersey.
- Sarabia, L. A. and Ortiz, M. C. (2009). Response Surface Methodology, Comprehensive Chemometrics: Chemical and Biochemical Data Analysis. *Elsevier Science*, 345-390.

- Schwaab, M., Biscaia, Jr., E. C., Monteiro, J. L. and Pinto, J. C. (2008). Nonlinear Parameter Estimation through Particle Swarm Optimization. *Chemical Engineering Science*, 63(6), 1542-1552.
- Subramanian, K., Kumar, S., Patwardhan, S. C. and Prasad, V. (2011), Extensions to Experiment Design and Identification Algorithms for Large-Scale and Stochastic Processes. *International Journal of Advanced Mechatronic Systems*, 2011 – Vol. 3, No.1 pp.3 - 13.
- Valorani, M., Najm, H. N. and Goussis, D. A. (2003). CSP Analysis of a Transient Flame Vortex Interaction: Time Scales and Manifolds. *Combustion and Flame*, 134:35-53.
- Valorani, M., Creta, F., Goussis, D., Lee, J. and Najm, H. (2006). An Automatic Procedure for the Simplification of Chemical Kinetic Mechanisms based on CSP. *Combustion and Flame*, 146(1-2), 29-51.
- Walter, E. and Pronzato, L. (1990). Qualitative and Quantitative Experiment Design for Phenomenological Models—A Survey. *Automatica*, 26(2), 195-213.
- Walter, E. and Pronzato, L., (1997). Identification of Parametric Models From Experimental Data, *Springer, Masson*, 413 pp.
- Zagaris, A., Kaper, H. G. and Kaper, T. J. (2004). Analysis of the Computational Singular Perturbation Reduction Method for Chemical Kinetics. *Journal of Nonlinear Science*, 14(1), 59-91.
- Zielinski, K., and Laur, R. (2007). Stopping Criteria for a Constrained Single-Objective Particle Swarm Optimization Algorithm. *Informatica*, 31, 51-59.

7. Appendix

MATLAB Functions

1. CLUSTERDATA

This function constructs clusters from data. The syntax is:

$$T = \text{CLUSTERDATA}(X, \text{'PARAM'}, \text{CUTOFF})$$

X is the data matrix of size M by N , treated as M observation of N variables. CUTOFF is a threshold for cutting the hierarchical tree. 'PARAM' is another option for inserting parameter values. For this study, the value of 'PARAM' used was 'maxclust' , which ensures that the maximum number of clusters formed is the CUTOFF .

$$T = \text{CLUSTERDATA}(S, \text{'maxclust'}, 8)$$

where S is the sensitivity matrix.

2. COMBVEC

This function creates all possible combination of vectors.

$$C = \text{COMBVEC}(A1, A2 \dots)$$

The function takes any number of inputs. If $A1$ is a matrix of $N1$ (column) vectors, and $A2$ is a matrix of $N2$ (column) vectors and so on, the function returns a matrix of $(N1*N2*\dots)$ column vectors, where the columns consist of all possibilities of $A2$ vectors, appended to $A1$ vectors, etc.