**Master Thesis in Statistics and Data Mining**

# Optimal Experimental Design and Parameter Estimation of a Stacked Bed Hydrotreating Process

Patrik Pavlov


Division of Statistics and Machine Learning
Department of Computer and Information Science
Linköping University

**Supervisor**
Anders Nordgaard

**Examiner**
Mattias Villani

# Contents

# Abstract

This master thesis deals with a project conducted at IFP Énergies nouvelles during the period from April to August 2014. It describes a methodology for estimating the kinetic parameters of stacked catalysts for the hydrotreating process. This methodology is built on extending the kinetic model for one catalyst and knowing that the feed has to pass all the catalysts in the order they are put in the reactor. The parameters of this new model are estimated by an adaptive version of the differential evolution algorithm. Another part of the work investigates how the hydrotreating experiments have to be conducted for obtaining the best precision of the estimates. This is done using a Bayesian optimal design methodology and the uncertainty for some of the best designs is further quantified in a simulation study. The results show that the estimation procedure is successful in estimating the kinetic parameters in a precise and effective way. Further it shows that a necessary requirement for obtaining reasonable precision is to use a scheme of sampling more temperatures than catalyst ratios and to obtain experimental points with as large differences in the temperatures and as large differences in the catalyst ratios as possible.

Keywords: Hydrotreating, Stacking of catalysts, Nonlinear modeling, Adaptive differential evolution, Bayesian optimal design

# Acknowledgments

I wish to thank my supervisor Anders Nordgaard at Linköping University for his help and support during this work. I am also grateful to IFP Énergies nouvelles and my supervisors Pascal Chatron-Michaud and Alberto Servia Silva for making this work possible. Last but not least I want to thank Isak Hietala for reviewing this thesis.

# 1 Introduction

## 1.1 IFP Énergies nouvelles

This thesis is based on work conducted at IFP Énergies nouvelles (IFPEN) from April to August 2014 where Pascal Chatron-Michaud and Alberto Servia Silva, both chemical engineering researchers were supervisors. IFPEN is a French research institute that specializes in energy, transport and the environment. Historically, the institute were mainly focused on petroleum research but have in recent times reformed the organization to also deal with research in new and future energies. IFPEN is with its nearly 1800 employees one of the leading petroleum and energy research institutes in the world (IFPEN, 2015).

## 1.2 Hydrotreating

The hydrotreating or hydrotreatment process can be described as a process where a petroleum feed, most often petroleum residue, is treated under hydrogen with one or more catalysts. The purpose of the hydrotreating is to improve the feed so that it meets a desired level of emission and quality, or to prepare it for other refinery processes. With improving the feed it is meant to lower the concentration of impurities, commonly sulfur and nitrogen (Topsøe et al., 1996).

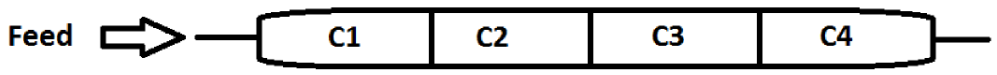


**Figure 1.1:** Example with 4 stacked-catalysts

The standard for hydrotreating is to use more than one catalyst (stacking) in the same reactor. This helps to improve the efficiency of the process and also removing more varying types of impurities. The hydrotreating process using 4 catalysts is illustrated in Figure 1.1. The feed moves from left to right, in through the reactor inlet, mixed with hydrogen it gets in contact with the first catalyst where a chemical reaction occurs. After meeting with the first catalyst, the feed is a bit cleaner and the concentration of impurities is lower. The feed with the new concentration of

impurities gets now in contact with the second catalyst and so on until it reaches the reactor outlet and the process is finished.

A catalyst's kinetic abilities can be characterized by its kinetic parameters. These kinetic parameters describe how the catalyst influences the hydrotreating process and can give information such as how much energy is necessary to activate the catalyst or at what rate the catalyst allows the reaction to occur. Learning the kinetic characteristics of catalysts is important when e.g. researching new catalysts.

## 1.3 Background

Modeling chemical processes is very common in the field of chemical engineering. Given a good model for a process the researchers are able to understand and analyze the process in an exact and effective way. This paper deals with a stacked catalyst hydrotreating process where the focus is on modeling the stacking of catalysts, creating a methodology for estimating the kinetic parameters of the stacked model and also optimal process designs.

The procedure of estimating kinetic parameters of catalysts has previously been done using a one-catalyst model at IFPEN. By modeling stacking of catalysts it is hoped to more cost efficiently be able to estimate kinetic parameters of several catalysts. Cost is measured in the number of changes in operating variables necessary for precise estimation. Every change to the process costs a lot of money and it is hoped that by stacking catalysts in the same reactor one could lower the number of changes necessary to estimate the parameters of several catalysts. This would be achieved by learning the kinetic parameters of several catalyst at the same time in the same reactor instead of using one reactor to estimate parameters of only one catalyst.

To streamline the estimation procedure it is necessary to find some optimal designs to use for the operating variables. By finding optimal process designs one would be able to estimate the kinetic parameters of several catalysts with as few data points as possible. The hydrotreating process is limited when considering what are possible designs. Therefore it is only of interest to find out how some common designs compete with each other and which of these design is to be preferred when constructing future hydrotreating experiments.

## 1.4 Previous work

Estimating the kinetic parameters of the stacked model can be seen as a typical nonlinear regression problem where it is desired to obtain parameter estimates that minimize some objective function. For easily differentiable objective functions one has traditionally used derivative based methods like gradient descent or the Newton-Raphson method, both described in Luenberger and Ye (2008). These methods allow

the search of the global minimum to be done with the help of derivatives that at each iteration says at what direction of the parameter space to move the search (Finlayson, 2003). Derivative based methods are not always plausible and can fail for example when there are parts of the parameter space with zero derivatives or when the parameter space contains multiple minima. For these more complicated problems it is common to apply direct search approaches. Direct search approaches do not rely on any derivative information to guide the search, instead they are generating variations of the parameter vectors at each iteration based on values of the objective function. Commonly used methods are genetic algorithms: (Goldberg, 1989), simulated annealing: (Kirkpatrick et al., 1983) and the differential evolution algorithm: (Storn and Price, 1995). Simulated annealing has in Eftaxias et al. (2002) been successfully used for estimating the kinetic parameters in a model for the catalytic wet air oxidation of phenol chemical process while a modified differential evolution algorithm has been shown to be effective for the optimization of several non-linear chemical processes in Babu and Angira (2006).

Choosing the best optimization algorithm for a given problem is not a straightforward task since the best algorithm for a given problem highly depends on the problem itself. In this paper we adopt a method called JADE which is described in Zhang and Sanderson (2009) and is based on the differential evolution algorithm where the control parameters of the algorithm change adaptively during the run of the algorithm.

Traditionally, optimal designs are derived so as to optimize functions of the Fisher information matrix. Common optimal design criteria include A-optimality and D-optimality. A-optimality aims to find an optimal design that minimizes the average variance of the parameters by minimizing the trace of the inverse of the Fisher information matrix while D-optimality aims to minimize the generalized variance of the parameters by maximizing the determinant of the information matrix (Pukelsheim, 1993). These methods assume that the data depend linearly on the model parameters which might not always be reasonable. In the case of the stacked hydrotreating model we have a model that is highly nonlinear and it is thus not feasible to make this assumption. For nonlinear optimal design problems it is common to adopt Bayesian optimal designs, as described in Cook et al. (2008); Huan and Marzouk (2013); Müller (2005); Ryan et al. (2014).

## 1.5 Research objectives

The first objective of this thesis is to learn how the kinetic parameters of the stacked-catalyst hydrotreating model can be estimated in an efficient and precise way. The second objective is to learn how to conduct hydrotreating experiments in order to obtain good parameter precision. Further it is desired to quantify the precisions given by different good designs.

# 2 Methods

This chapter covers the scientific methods used in the thesis. The first section is dedicated to the hydrotreating model and the work done extending the original one-catalyst model to a model including more than one catalyst. The following section (sec. 2.2) describes the methodology of how and using what settings the data in the thesis were simulated. Another part of the thesis (sec. 2.3) presents a method which is an answer to the question of how to estimate the kinetic parameters in an exact and effective way. Methods answering the question of how to conduct hydrotreating experiments yielding highest precision and quantifying this precision are described in sec. 2.4.

## 2.1 Modeling

### 2.1.1 One-catalyst model

A standard one-catalyst hydrotreating model is associating the output concentration of impurities $x_1$ to the inlet concentration of impurities $x_0$, the operating variables residence time $rT_1$ and temperature $T$, and the three kinetic parameters reaction order $n_1$, rate constant $k_1$ and activation energy $Ea_1$. $R$ is the universal gas constant. The one-catalyst model has the form

$$x_1 = \left\{ (n_1 - 1)(k_1 e^{\frac{-Ea_1}{R(T+273)}} rT_1 - \frac{x_0^{1-n_1}}{1-n_1}) \right\}^{\frac{1}{1-n_1}} . \tag{2.1}$$

Note that the $x_0$ and $x_1$ both measure the total percentage of impurities, meaning that they include all different kinds of impurities existing in the feed.

### 2.1.2 Stacked-catalyst model

The extension of the one-catalyst model to the stacked-catalyst model is based on the knowledge that the catalysts are stacked in some order in the reactor and that there is no mixing between the catalysts. Then it is easy to see that the outlet concentration of impurities, after the feed has passed one catalyst-region, is the

input concentration to the following catalyst-region. One could thus nest several one-catalyst models together (one for every catalyst) by this relation. The expression for the stacked-catalyst model with $K$ number of catalysts is

$$x_l = \left\{ (n_l - 1)(k_l e^{\frac{-Ea_l}{R(T+273)}} rT_l - \frac{x_{l-1}^{1-n_l}}{1-n_l}) \right\}^{\frac{1}{1-n_l}} \quad l = 1, ..., K \tag{2.2}$$

where $x_l$ denotes the output concentration after the feed has passed catalyst-region $l$. The catalysts are stacked so that the feed first gets in contact with the first catalyst $C_1$, then the second catalyst $C_2$ and so on until it reaches the last catalyst $C_K$ which yields the final output concentration $x_K$. The $x_K$ is thus a function of the the input concentration $x_0$, the kinetic parameters of every catalyst $n_l$, $k_l$ and $Ea_l$ and the operating variables $rT_l$ and $T$. Note that the temperature is assumed to be constant in the whole reactor while every other term is catalyst specific.

By this scheme the two-catalyst model ($K = 2$) has the form

$$\begin{aligned} x_1 &= \left\{ (n_1 - 1)(k_1 e^{\frac{-Ea_1}{R(T+273)}} rT_1 - \frac{x_0^{1-n_1}}{1-n_1}) \right\}^{\frac{1}{1-n_1}} \\ x_2 &= \left\{ (n_2 - 1)(k_2 e^{\frac{-Ea_2}{R(T+273)}} rT_2 - \frac{x_1^{1-n_2}}{1-n_2}) \right\}^{\frac{1}{1-n_2}} \end{aligned} \tag{2.3}$$

where the output concentration of the first catalyst-region $x_1$ is used as the inlet concentration to the second catalyst-region.

The process performance will in order to comply with IFPEN standards be expressed as $HDX$, the percentage of impurities cleaned from the feed. For the stacked-catalyst model with $K$ catalysts the $HDX$ is

$$HDX = 100 \times \frac{x_0 - x_K}{x_0}. \tag{2.4}$$

### 2.1.3 Kinetic parameters and operating settings

The operating settings or variables used by the model are the average reactor temperature $T$ and the residence time $rT_l$ for each catalyst-region $l$. The residence time $rT_l$ is a measure of how long time the feed is exposed to catalyst $l$. The longer the feed is exposed to catalyst $l$, the larger will be the effect of catalyst $l$ on the feed. The temperature $T$ is measured on the Celsius scale.

The global residence time $grT$ is the sum of all the catalyst's residence times. It can be seen from the model in equation (2.2) that the higher the $grT$, the higher is

the response $HDX$ expected to be. Common values of the $grT$ lie in the interval $[0.5, 1.5]$.

In this thesis it is beneficial to speak about the catalyst ratio $CR$ of catalyst $l$. $CR_l$ is defined as the proportion of reactor space occupied by catalyst $l$. More specifically, the catalyst ratio for catalyst $l$ is $CR_l = rT_l/grT$.

The $T$ highly affects the rate of reaction by controlling how much energy can be delivered to a system. From the model it is seen that the higher the $T$, the lower is the output concentration $x_K$ expected to be. Common values of $T$ are ranging from 350 to 390 °C.

Every catalyst's kinetic abilities can be described by three kinetic parameters, the reaction order $n_l$, the activation energy $Ea_l$ and the rate constant $k_l$. The reaction order of a catalyst $l$ is defined as the exponent to which the rate of reaction is raised by when the feed gets in contact with catalyst $l$. A catalyst with higher reaction order will have larger impact on the process than a catalyst with a smaller reaction order, given other values being equal (McNaught and McNaught, 1997).

The activation energy is defined as the minimum energy required to start a chemical reaction. The lower the activation energy, the earlier in the process can the reaction occur and the lower temperature is required to activate the process (Atkins and De Paula, 2010).

While the two parameters $n_l$ and $Ea_l$ measure specific kinetic abilities of the catalyst, the $k_l$ is a product of other parameters unaccounted-for. The implications of this definition of $k_l$ is that common values of $k_l$ are very high, lying in the interval $[10^{10}, 10^{18}]$. Common values of $n_l$ lie in the interval $(1, 2.5]$ and common values of $Ea_l$ lie in the interval $[20000, 100000]$.

The large values for $k_l$ has shown to be a problem when estimating the kinetic parameters because of the resulting large parameter space. A solution to this problem is to scale down the term $k_l$ so that the parameter space becomes narrower and the scale of the parameter is more similar to the other parameters $Ea_l$ and $n_l$. The scaling of $k_l$ was in this thesis done by replacing $k_l$ with $k_l^{1/10}$ in the model. Using $k_l^{1/10}$ instead of $k_l$ results in a much narrower parameter space and simplifies the estimation procedure. Note that $k_l^{1/10}$ will from here on out be referred to as $k_l$.

## 2.2 Data simulation

Because of limited supply of data at IFPEN, it is necessary to simulate data to get some material to work on. The data used in this thesis, with an exception of the perfect data used in sec. 2.3.4, is simulated from the model itself where some random noise is added to every response $HDX$ in order for this simulated data to better resemble real data. The added noise is randomly drawn from a normal distribution

with mean 0 and a standard deviation of 0.25. The error distribution is chosen to resemble historical data collected at IFPEN.

The two catalyst pairs from which data in this thesis will be simulated are shown in Table 2.1. These parameter values belong to some common catalysts used by IFPEN. For the first catalyst pair we have two catalysts that are very different from each other. The features of the first catalyst result in a catalyst that activates early, meaning that it will have a relatively good effect even on low temperatures as opposed to the second catalyst. For the second catalyst pair we see quite equally performing catalysts, with the second catalyst being slightly more effective, because of a high value of $k$.

**Table 2.1:** Catalysts used in the thesis

| Par | Pair 1 | | Pair 2 | |
|---|---|---|---|---|
| $n$ | 1.3 | 2.1 | 1.4 | 1.2 |
| $Ea$ | 32000 | 52000 | 45000 | 50000 |
| $k$ | 11.96 | 47.62 | 30.05 | 47.62 |

Another setting that will be universally used throughout the thesis is a input concentration of impurities $x_0$ of 4%. This is a common value of the impurity concentration for petroleum residue.

## 2.3 Parameter estimation

### 2.3.1 Nonlinear regression

Estimating the kinetic parameters of the stacked model can be seen as a typical nonlinear regression problem where one wants to obtain parameter estimates that minimize some objective function. A nonlinear regression model can be written as $\boldsymbol{Y} = f(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{Z}$ where $f$ is a nonlinear expectation function that depends on $\boldsymbol{X}$ which is a matrix of independent variables or design matrix containing $n$ rows and $p$ columns. The parameter vector $\boldsymbol{\theta}$ is a vector containing $p$ model parameters. The error vector is denoted by $\mathbf{Z}$ and has length $n$.

The choice of objective function is based on assumptions of the underlying process that creates the error vector $\mathbf{Z}$. Assuming that the errors are independently and symmetrically distributed around a common mean, i.e. $E[\mathbf{Z}] = 0$ and $Var(\mathbf{Z}) = \sigma^2\mathbf{I}$, makes the sum of squared error (SSE) an appropriate objective function. Using the SSE as objective function one wants to find an estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ that minimizes

$$SSE(\boldsymbol{\theta}) = (\boldsymbol{Y} - f(\boldsymbol{X}, \boldsymbol{\theta}))^T(\boldsymbol{Y} - f(\boldsymbol{X}, \boldsymbol{\theta})). \tag{2.5}$$

### 2.3.2 Differential evolution

The differential evolution (DE) algorithm is a population based stochastic direct search method used for global optimization. DE is a method that optimizes a problem by iteratively trying to improve its population members with regard to an objective function.

The following description of the differential evolution algorithm follow Price et al. (2006).

The DE algorithm utilizes a population containing $NP$ number of target vectors or individuals each of length $p$, $p$ being the number of parameters. The initial population is chosen uniformly random from the parameter space containing some user specified upper and lower bounds. Once initialized, the DE algorithm mutates and crosses the target vectors in a particular way to create $NP$ number of trial vectors, one for every target vector. A trial vector is at the end of every generation (iteration) compared to its target vector, and the vector having the lowest SSE will in the next generation be the new, evolved target vector. This process is repeated until a user specified optimum or a user specified number of iterations are reached.

Every target vector will at each iteration first be mutated and then recombined (crossed) with another vector. The mutation can be done in many different ways but in this paper we have chosen to evaluate two different mutation strategies called rand and local-to-best. The rand strategy creates the $i^{th}$ mutant vector at generation $g$ as

$$\mathbf{v}_{i,g} = \mathbf{x}_{r1,g} + F(\mathbf{x}_{r2,g} - \mathbf{x}_{r3,g}) \tag{2.6}$$

where $\mathbf{x}_{r1,g}$,$\mathbf{x}_{r2,g}$ and $\mathbf{x}_{r3,g}$ denotes three randomly chosen target vectors at generation $g$. The $F \in (0, 2)$ is a mutation scale factor which controls the rate at which the population evolves. The other considered evolution strategy, local-to-best, creates the $i^{th}$ mutant vector at generation $g$ as

$$\mathbf{v}_{i,g} = \mathbf{x}_{i,g} + (\mathbf{best}_g - \mathbf{x}_{i,g}) + \mathbf{x}_{r1,g} + F(\mathbf{x}_{r2,g} - \mathbf{x}_{r3,g}) \tag{2.7}$$

where $\mathbf{x}_{i,g}$ denotes the $i^{th}$ target vector at generation $g$, $\mathbf{best}_g$ denotes the absolute best target vector (lowest SSE) at generation $g$ and $\mathbf{x}_{r1,g}$,$\mathbf{x}_{r2,g}$, $\mathbf{x}_{r3,g}$ and $F$ are the same as before. By including the best vector for creating every mutant vector in the local-to-best strategy one hopes to faster steer the algorithm towards the global minimum. The three random vectors are randomly generated for each vector $i$, and are included to the evolution in order to increase the diversity in the population and minimize the risk that the algorithm gets stuck at a local minimum.

To add additional diversity to the population, the DE will after the mutation perform discrete recombination between each pair of mutant and target vector, creating $NP$ trial vectors. The recombination is such that the trial vector inherits some parameter values from the target vector and the other parameter values from the mutant vector. The trial vectors are the vectors that will be compared to the target vectors in order to create the coming generation. Trial vector $i$ at generation $g$ will inherit parameter value $j$, $j = 1, ..., p$, as

$$\begin{aligned} \mathbf{u}_{j,i,g} &= \mathbf{v}_{j,i,g}, \ if \ (rand_j[0,1] < CP \ or \ j = j_{rand}) \\ \mathbf{u}_{j,i,g} &= \mathbf{x}_{j,i,g}, \ Otherwise. \end{aligned} \tag{2.8}$$

The crossover probability $CP \in [0, 1]$ controls the level of crossover between the mutant and target vector. A higher $CP$ means that the trial vector more probably will inherit values from the mutant vector than from the target vector. In addition to the crossover probability, there is one randomly chosen parameter $j_{rand}$ that inherits its value from the mutant vector. This is done so that it is impossible for the trial vector to duplicate the target vector. The pseudo code of the DE algorithm with the local-to-best evolution strategy is shown in Algorithm 2.1.

It is important to note that the tuning of the control parameters of the algorithm, the population size $NP$, the number of iterations, the mutation scale factor $F$ and the crossover probability all play an important role in the performance of the algorithm and need to be chosen to fit the specific problem. The choice of the population size $NP$ is often based on finding a good balance between the convergence speed and the risk of getting stuck at a local minimum. The reasoning for using a larger rather than a smaller $NP$ is based on the fact that a larger amount of the parameter space will be covered at every iteration and the risk of getting stuck at a local minimum decreases. Another benefit with a larger population size is that the probability of finding the optimum at an earlier stage increases. Although the DE algorithm generally have a low computational complexity compared to many other similar algorithms, choosing too many population members will result in an unnecessarily slow algorithm because the computational complexity increases more than linearly with $NP$. Choosing the mutation scale factor $F$ and the crossover probability $CP$ is another problem but there are some successful ways of dealing with that, see sec. 2.3.3.

### 2.3.3 Adaptive differential evolution

The classical DE algorithm has since its introduction to the market been a very successful optimization algorithm used by researchers in many real-world problems. The performance of the algorithm is however highly dependent on the setting of its control parameters. One approach to automatically deal with the calibration

**Algorithm 2.1** DE algorithm

**Require:** $NP, CP, F, p$

Initiate random population $\{ \mathbf{x}_{i,0},\ i = 1, ..., NP\}$
g=0
**while** not converged **do**
    $g = g + 1$
    **for** $i = 1$ to $NP$ **do**
        Set $\mathbf{best}_g$ as the best vector from the current population $g$
        Randomly pick $\mathbf{x}_{r1,g}$, $\mathbf{x}_{r2,g}$, $\mathbf{x}_{r3,g}$ s.t. $\mathbf{x}_{r1,g} \neq \mathbf{x}_{r2,g} \neq \mathbf{x}_{r3,g} \neq \mathbf{x}_{i,g}$
        $\mathbf{v}_{i,g} = \mathbf{x}_{i,g} + (\mathbf{best}_g - \mathbf{x}_{i,g}) + \mathbf{x}_{r1,g} + F(\mathbf{x}_{r2,g} - \mathbf{x}_{r3,g})$
        Generate $rand_j = randint(1, p)$
        **for** $j = 1$ to $p$ **do**
            **if** $j = rand_j$ or $U(0, 1) < CP$ **then**
                $u_{i,g,j} = v_{i,g,j}$
            **else**
                $u_{i,g,j} = x_{i,g,j}$
            **end if**
        **end for**
        **if** $SSE(\mathbf{x}_{i,g}) \leq SSE(\mathbf{u}_{i,g})$ **then**
            $\mathbf{x}_{i,g+1} = \mathbf{x}_{i,g}$
        **else**
            $\mathbf{x}_{i,g+1} = \mathbf{u}_{i,g}$
        **end if**
    **end for**
**end while**

of control parameters is to utilize self-adaptive parameter control extensions to the differential evolution algorithm. One of the more successful self-adaptive versions of the algorithm is used in JADE and described in Zhang and Sanderson (2009).

The parameter adaptation in JADE automatically sets the two control parameters mutation scale factor $F$ and the crossover probability $CP$ during the run of the algorithm. It does so by the reasoning that better parameter values tend to generate more successful (more frequently evolving) target vectors and that the parameter settings that generated these successful target vectors can beneficially be used by the coming generations.

This adaptive version of the DE algorithm will dynamically adapt the control parameters without the need for human input and will by this adaptivity most often lead to an improved convergence rate. The convergence rate will be improved by the reasoning that different values of the control parameters will be successful at different stages of the algorithm. The settings that generate the most successful individuals at an early stage of the algorithm, when the individuals are far away from the global minimum, might not be the same optimal settings for a later stage, when

the individuals are closer to the global minimum. From this, one clearly sees that generating successful individuals at all stages than just at some stages, like with the standard DE, will increase the efficiency and convergence rate of the algorithm.

The parameter adaptation works as follows. The crossover probability $CP_i$ for each individual $i$ at generation $g$ is independently generated from a truncated normal distribution with mean $\mu_{CP,g}$ and standard deviation 0.1, truncated to [0,1]

$$CP_i \sim TN(\mu_{CP,g}, 0.1, 0, 1). \tag{2.9}$$

The mean of the truncated normal distribution is at the end of every generation updated as

$$\mu_{CP,g+1} = (1 - c) * \mu_{CP,g} + c * mean(S_{CP,g}) \tag{2.10}$$

where $c \in [0, 1]$ and $S_{CP,g}$ is denoted as the set of all successful $CP_i$ at generation $g$. A $CP_i$ is said to be successful if it leads to the evolution of individual $i$. The starting value of $\mu_{CP,g}$ is initialized to be 0.5. One can thus see that the $CP_i$ for each individual at each $g$ is updated as a weighted average between the previous mean crossover probability $\mu_{CP,g}$ and the arithmetic mean of all current successful crossover probabilities, with some random noise added. The parameter $c$ controls the rate of parameter adaptation. A higher value of $c$ will mean that $\mu_{CP,g+1}$ to a higher degree will be composed of recent successful crossover probabilities $S_{CP,g}$ than $\mu_{CP,g}$. The noise from the normal distribution is added to $CP_i$ in order to explore potential successful crossover probabilities. A standard deviation of 0.1 gives a sampling distribution rather concentrated around the $\mu_{CP,g}$ so as to direct the search of future crossover probabilities to a near neighborhood of $\mu_{CP,g}$.

The second control parameter, the mutation factor is for each individual $i$ at generation $g$ independently generated from a Cauchy distribution with location parameter $\mu_{F,g}$ and scale parameter 0.1 as

$$F_i \sim TC_i(\mu_{F,g}, 0.1) \tag{2.11}$$

truncated to 1 if $F_i \geq 1$ and $F_{i-1}$ if $F_i \leq 0$. The location parameter of the truncated Cauchy distribution is at the end of every generation updated as

$$\mu_{F,g+1} = (1 - c) * \mu_{F,g} + c * mean_L(S_{F,g}) \tag{2.12}$$

where $c$ is the same control parameter as before and $S_{F,g}$ denotes the set of all successful mutation factors $F_i$ at generation $g$. $mean_L(S_{F,g})$ is the Lehmer mean described in Bullen (2003) and defined as

$$mean_L(S_{F,g}) = \frac{\sum_{F \in S_{F,g}} F^2}{\sum_{F \in S_{F,g}} F}. \tag{2.13}$$

By using the Cauchy distribution instead of the normal distribution when generating $F_i$ one diversifies the mutation factor, because of the heavier tails of the Cauchy distribution, which in turn will help avoiding premature convergence. By using the Lehmer mean instead of the arithmetic mean in the creation of $\mu_{F,g}$ one gives more weight to larger successful mutation factors, which has shown to increase the progress rate of the algorithm.

### 2.3.4 Algorithm evaluation

The convergence properties of the standard DE algorithm and the adaptive version JADE will be evaluated. It is of interest to learn how fast and how frequently the algorithms converge on the two-catalyst model.

The performance evaluation will be carried out using eight different model settings created by using different combinations of kinetic parameters and data sets. The datasets will be considered perfect in the sense that there are no errors added to the response variable, $HDX$, so that the objective value, SSE, at the global minimum will be 0, making it easy to identify algorithm convergence.

The kinetic parameters of the two catalyst pairs used in this performance evaluation are shown in Table 2.1. Catalyst pair 1 is assigned to settings 1, 3, 5 and 7 while catalyst pair 2 is assigned to settings 2, 4, 6 and 8.

The four datasets used for the performance evaluation are shown in Table 2.2, Table 2.3, Table 2.4 and Table 2.5. Dataset 1 is used by setting 1 and 2, dataset 2 is used by setting 3 and 4, dataset 3 is used by setting 5 and 6 and dataset 4 is used by setting 7 and 8. The difference between the datasets in Table 2.2 and Table 2.3 and the datasets in Table 2.4 and Table 2.5 is that the two first contain designs with six data points while the last two contain the same designs but with one additional data point added to each design. These 4 specific datasets are chosen so that we can learn how different number of data points, six and seven, and how different types of designs affect the convergence abilities of the algorithms. The structures of these datasets will further be discussed in sec. 2.4

For each setting we will run the different algorithms 50 times and record the proportion of runs that converge to the global minimum and the number of iterations required to reach this minimum. The two evolution strategies described in sec. 2.3.2

**Table 2.2:** Dataset 1

| $T$ | $rT_1$ | $rT_2$ | $x_0$ |
|---|---|---|---|
| 350 | 0.2 | 0.8 | 4 |
| 370 | 0.2 | 0.8 | 4 |
| 390 | 0.2 | 0.8 | 4 |
| 350 | 0.8 | 0.2 | 4 |
| 370 | 0.8 | 0.2 | 4 |
| 390 | 0.8 | 0.2 | 4 |

**Table 2.3:** Dataset 2

| $T$ | $rT_1$ | $rT_2$ | $x_0$ |
|---|---|---|---|
| 350 | 0.2 | 0.8 | 4 |
| 390 | 0.2 | 0.8 | 4 |
| 350 | 0.5 | 0.5 | 4 |
| 390 | 0.5 | 0.5 | 4 |
| 350 | 0.8 | 0.2 | 4 |
| 390 | 0.8 | 0.2 | 4 |

**Table 2.4:** Dataset 3

| $T$ | $rT_1$ | $rT_2$ | $x_0$ |
|---|---|---|---|
| 350 | 0.2 | 0.8 | 4 |
| 370 | 0.2 | 0.8 | 4 |
| 390 | 0.2 | 0.8 | 4 |
| 350 | 0.8 | 0.2 | 4 |
| 370 | 0.8 | 0.2 | 4 |
| 390 | 0.8 | 0.2 | 4 |
| 360 | 0.8 | 0.2 | 4 |

**Table 2.5:** Dataset 4

| $T$ | $rT_1$ | $rT_2$ | $x_0$ |
|---|---|---|---|
| 350 | 0.2 | 0.8 | 4 |
| 390 | 0.2 | 0.8 | 4 |
| 350 | 0.5 | 0.5 | 4 |
| 390 | 0.5 | 0.5 | 4 |
| 350 | 0.8 | 0.2 | 4 |
| 390 | 0.8 | 0.2 | 4 |
| 370 | 0.8 | 0.2 | 4 |

will be evaluated using both the standard DE, without parameter adaptation, and the JADE extension. The control parameters of the standard DE are set to the default, proposed by its authors, of $CP = 0.5$ and $F = 0.8$. The adaptivity of the JADE extension will be evaluated using three different values of $c$, $c = (0.01, 0.1, 0.5)$. The number of population members $NP$ are set to 120 and a run of the algorithm is seen to be converged when the SSE reaches a value of $10^{-8}$.

## 2.4 Optimal design

The optimal design methodology presented in this section aims at finding good experimental designs in order to estimate the kinetic parameters of the stacked-catalysts model with reasonable precision and with the lowest cost possible. The experimental cost is measured in terms of the number of experimental runs and also the type of changes done between the experimental runs. Every run of the hydrotreating reactor must be unique to be informative, which means that it is often necessary to change the temperature, the catalyst ratio and the global residence time in the reactor. The cost of changing each of these three factors vary greatly since different long resets and changes has to be done to the process. A change in $T$ is known to be less costly than a change in $grT$ and a change in $grT$ is known to be

less costly than a change in $CR$. By keeping this in mind one can prioritize the consideration of designs with as few changes in $CR$ as possible.

In this thesis we will focus on optimal designs for the two-catalyst model since that is a very common stacked catalysts model and it is reasonably assumed that the general results obtained for this model easily can be generalized to other stacked-catalyst models containing more than two catalysts.

The foundation of all experimental designs considered will build on a saturated model where the number of data points are equal to the number of parameters, $n = p$. This means that the minimum number of experimental data points considered for the two-catalyst model will be $n = p = 6$.

When stacking two catalysts in the model one can construct the experiments in many different ways. In this thesis the focus is on two general designs that has been chosen as more important to analyze than others. These are the designs of sampling three different temperatures for two different catalyst ratios and to sample two different temperatures for three different catalyst ratios. These two designs will from here on be referred to as the $2rT$ and the $3rT$ designs. Both of these designs lead to a saturated model with six data points. An example of a $2rT$ design is shown in Table 2.2 and an example of a $3rT$ design is shown in Table 2.3 .

Within the two general designs, some sub-designs will be considered which are created by sampling the $T$, $CR$ and $grT$ in some specific ways. For the $T$ and $CR$, the focus is on the sizes of changes between the sampled experimental points. Different sizes of changes in $CR$ and $T$ are defined by the variables $CRd$ and $Td$, respectively. They define the distance between the sampled $CR$ and the distance between the sampled $T$ in a data set. The higher the values of $CRd$ and $Td$, the more the $CR$ and $T$ differ in a specific data set. As an example, when sampling three temperatures and two catalyst ratios ($2rT$), $Td = 10$ might give the sample $T = (370, 380, 390)$, the sample $T = (360, 370, 380)$ or any other set with three temperatures 10 units from each other. On the other hand, a $CRd$ of 0.3 will for the same design always result in $CR = (0.35/0.65, 0.65/0.35)$.

In addition, we will compare designs where only changes in $Td$ and $CRd$ are considered with designs where also changes in $grT$ are considered. Two different types of changes in $grT$ will be considered. The first design $grT(T)$ consists of utilizing different $grT$ for every different $T$, i.e. when sampling $k$ temperatures we will have $k$ unique $grT$, each associated with one temperature. The other considered design $grT(CR)$ consists of utilizing different $grT$ for every different $CR$, i.e. when sampling $k$ catalyst ratios we will have $k$ global residence times, each associated with one catalyst ratio. By this analysis one can thus learn whether it is more beneficial in terms of information gain to conduct experiments with changes to $grT$ for every change in $CR$, changes in $grT$ for every change in $T$ or simply with no change in $grT$ at all. Examples of a $grT(T)$ and a $grT(CR)$ design using the general design $2rT$ are shown in Table 2.6 and Table 2.7, respectively.

**Table 2.6:** Example data $grT(T)$

| $T$ | $rT_1$ | $rT_2$ | $x_0$ |
|---|---|---|---|
| 350 | 0.1 | 0.4 | 4 |
| 370 | 0.2 | 0.8 | 4 |
| 390 | 0.3 | 1.2 | 4 |
| 350 | 0.4 | 0.1 | 4 |
| 370 | 0.8 | 0.2 | 4 |
| 390 | 1.2 | 0.3 | 4 |

**Table 2.7:** Example data $grT(CR)$

| $T$ | $rT_1$ | $rT_2$ | $x_0$ |
|---|---|---|---|
| 350 | 0.2 | 0.8 | 4 |
| 370 | 0.2 | 0.8 | 4 |
| 390 | 0.2 | 0.8 | 4 |
| 350 | 1.2 | 0.3 | 4 |
| 370 | 1.2 | 0.3 | 4 |
| 390 | 1.2 | 0.3 | 4 |

### 2.4.1 Bayesian optimal design

A Bayesian optimal design (BOD) approach that determines an optimal design based on maximizing the expectation of some utility function was used to asses which experimental designs are more suitable than others in this specific problem. Looking at the problem from a decision theoretic point of view, one can construct the objective function of the experimental design as

$$\begin{aligned} U(\mathbf{d}) &= \int_{\mathbf{Y}} \int_{\mathbf{\Theta}} u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \\ &= \int_{\mathbf{Y}} \int_{\mathbf{\Theta}} u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y}) p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \end{aligned} \tag{2.14}$$

where $u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta})$ is a utility function, $\mathbf{Y}$ is the support of $\mathbf{y}$ and $\mathbf{\Theta}$ is the support of $\boldsymbol{\theta}$. With this general objective function one chooses the best experimental design $\mathbf{d}^*$ that maximizes $U(\mathbf{d})$ over the considered design space with respect to unknown future observations $\mathbf{y}$ and model parameters $\boldsymbol{\theta}$ (Lindley et al., 1972).

The utility function considered in this thesis is the Kullback-Leibler (KL) divergence, described in Kullback and Leibler (1951), from posterior to prior which is expressed as

$$u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta}) = \int_{\mathbf{\Theta}} p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y}) ln\Big[\frac{p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y})}{p(\boldsymbol{\theta})}\Big] d\boldsymbol{\theta} = u(\mathbf{d}, \mathbf{y}) \tag{2.15}$$

where $p(\boldsymbol{\theta})$ is the prior and $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})$ is the posterior. In the expression it is easy to see that the KL divergence from posterior to prior is independent of the parameters $\boldsymbol{\theta}$ which allows us to rewrite the expected utility in equation 2.14 as

$$
\begin{aligned}
U(\mathbf{d}) &= \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} u(\mathbf{d}, \mathbf{y}) p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y}) p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \\
&= \int_{\mathbf{Y}} u(\mathbf{d}, \mathbf{y}) p(\mathbf{y}|\mathbf{d}) d\mathbf{y} \\
&= \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y}) ln\Big[\frac{p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y})}{p(\boldsymbol{\theta})}\Big] p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y}. \qquad (2.16)
\end{aligned}
$$

By using the KL divergence as utility criterion one will by this methodology maximize the expected information gain on $\boldsymbol{\theta}$ due to obtaining $\mathbf{y}$ by performing an experiment at a design point $\mathbf{d}$. The implications of a higher KL divergence from posterior to prior is that more entropy will be decreased in $\boldsymbol{\theta}$ thanks to $\mathbf{y}$ and that this $\mathbf{y}$ therefore is more informative (Huan and Marzouk, 2013).

The expected utility in equation 2.16 is analytically intractable and needs to be approximated numerically. By the use of Bayes theorem to the quantities in the expression of $U(\mathbf{d})$ one can rewrite $U(\mathbf{d})$ as

$$
\begin{aligned}
U(\mathbf{d}) &= \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y}) ln\Big[\frac{p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y})}{p(\boldsymbol{\theta})}\Big] p(\mathbf{y}|\mathbf{d}) d\mathbf{y} d\boldsymbol{\theta} \qquad (2.17) \\
&= \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} \{ln[p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{d})] - ln[p(\mathbf{y}|\mathbf{d})]\} p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{d}) p(\boldsymbol{\theta}) d\mathbf{y} d\boldsymbol{\theta} \qquad (2.18)
\end{aligned}
$$

and then use importance sampling to estimate the integral in equation 2.18. The importance sampling estimate is

$$
\hat{U(\mathbf{d})} = \sum_{i=1}^{n} \Big\{ \frac{p(\mathbf{y}_i|\boldsymbol{\theta}_i, \mathbf{d})}{\sum_{i=1}^{n} p(\mathbf{y}_i|\boldsymbol{\theta}_i, \mathbf{d})} ln[p(\mathbf{y}_i|\boldsymbol{\theta}_i, \mathbf{d})] - ln[\frac{1}{n} \sum_{i=1}^{n} p(\mathbf{y}_i|\boldsymbol{\theta}_i, \mathbf{d})] \Big\}. \qquad (2.19)
$$

where $n$ denotes the number of Monte Carlo samples the approximation is based on and the importance weight of sample $i$ is $p(\mathbf{y}_i|\boldsymbol{\theta}_i, \mathbf{d}) / \sum_{i=1}^{n} p(\mathbf{y}_i|\boldsymbol{\theta}_i, \mathbf{d})$. A sample $i$ is created by first sampling $\boldsymbol{\theta}_i$ from the prior $p(\boldsymbol{\theta})$, then sampling $\mathbf{y}_i$ from the likelihood $p(\mathbf{y}_i|\boldsymbol{\theta} = \boldsymbol{\theta}_i, \mathbf{d})$ (Ryan et al., 2014).

In the application of finding the best optimal designs for the stacked hydrotreating model we have used uniform priors $p(\boldsymbol{\theta})$ covering the whole possible parameter space, defined the distribution of responses $\mathbf{y}$ that might be observed as normally distributed and used 20000 samples $n$ to approximate the $U(\mathbf{d})$. By using 20000 samples one gets a reasonable exact estimate of $U(\mathbf{d})$, enough precise to be able to assert the differences in expected utility for different data sets correctly without a too great computational cost.

### 2.4.2 Utility evaluation of designs

The expected utility for datasets generated using the two general designs $2rT$ and $3rT$ with different combinations of $CRd$, $Td$ and $grT$ will be estimated. By this we will learn how the distance between the measurements $CR$ and $T$, additional changes in $grT$ and choice of general design affect the expected information gain from posterior to prior.

Since a specific $Td$ randomly generates temperatures with $Td$ units from each other one would need to repeat the process of estimating the utility for a specific combination of $Td$, $grT$ and $CRd$ in order to better be able to generalize the results of a specific $Td$. For every considered combination of $Td$, $grT$ and $CRd$ we therefore simulate 100 datasets on which we average the utility values to obtain a final average expected utility.

### 2.4.3 Inference of optimal designs

From the Bayesian optimal design one learns how to construct hydrotreating experiments yielding the highest KL divergence from posterior to prior. With these results it will not be possible to say anything more than how different designs on average compare with each other. In order to learn more about the designs which are shown to be optimal by the BOD we need to conduct some statistical inference. The statistical inference will allow us to learn how precise the kinetic parameters can be estimated given a specific experimental design.

The statistical inference will also be used as a comparative method to the BOD since it can be reasoned that the two methodologies will give similar conclusions in what designs are the best. This observation results from that the KL divergence from posterior to prior is closely related to the D-optimality condition which tries to minimize the generalized variance of the parameters. For linear Gaussian design problems it can be proven that the BOD using the KL divergence from posterior to prior in fact will equal the D-optimality (Huan and Marzouk, 2013).

The way one can conduct statistical inference without being able to sample real data is by simulation. With a given experimental design and some predefined catalysts (fixed parameter values) one simulates the data from the model by additionally adding some reasonable error to the response. The added noise is randomly drawn from a normal distribution with mean 0 and standard deviation of 0.25 as described in sec. 2.2. Every data set constructed in this way can be seen as a sample from the population of all possible data sets from that common experimental design. Having a design with $n$ data point we thus create a sample $b$, $b = 1, ..., B$, by sampling the response of data point $m$, $m = 1, .., n$, as

$$BHDX_{b,m} \sim N(HDX_m, 0.25) \tag{2.20}$$

where $HDX_m$ is the response value given from the model when using the experimental design of point $m$ and a certain catalyst-pair.

The process of generating data samples is repeated until $B$ number of data samples are obtained. For each of these $B$ samples we apply the algorithm proposed in sec. 2.3.3 to obtain $B$ parameter estimates $\hat{\boldsymbol{\theta}}^*(b)$. The $B$ parameter estimates will jointly result in sampling distributions for every kinetic parameter, conditioned on the specific experimental design and it will from these distributions be possible to learn the variability in the estimates obtained by using the specific design.

The standard error (SE) described in McDonald (2009) is chosen as a measure of variability with which the precision given by different designs can be compared. This measure is for kinetic parameter $j$, $j = 1, ..., p$, defined as

$$SE_j = \left( \frac{\sum_{b=1}^{B} \left\{ \hat{\theta}_j^*(b) - \hat{\theta}_j^*(\cdot) \right\}^2}{B-1} \right)^{1/2} \tag{2.21}$$

where

$$\hat{\theta}_j^*(\cdot) = \frac{\sum_{b=1}^{B} \hat{\theta}_j^*(b)}{B} \tag{2.22}$$

is the estimated mean.

The inference of the experimental designs will be done using the two different catalyst pairs from Table 2.1. The reason for investigating two different catalyst pairs is that it is believed that the precision of parameter estimates depends on the catalysts themselves and it is desired to learn more about this by using two very different catalyst pairs.

In addition to analyzing the $2rT$ and $3rT$ designs with $n = p = 6$ we will also analyze experimental designs constructed by adding one additional data point to the experimental data. This means that the $p$ model parameters will be estimated using $n = p + 1 = 7$ experimental data points. The new data point will be constructed by adding one extra $T$ to one of the already existing $CR$. This is considered to be the least costly way of sampling one additional data point since one only has to change the $T$ and not the $CR$ or the $grT$. By adding this extra data point one can learn how the addition of one extra data point affect the precision of the parameter estimates. A $2rT$ design with one extra point is seen in Table 2.4 while a $3rT$ design with one extra point is seen in Table 2.5.

# 3 Results

## 3.1 Algorithm evaluation on the two-catalyst model

The speed and success of estimating the kinetic parameters, using the local-to-best evolution strategy, are for the eight different two-catalyst settings (presented in sec. 2.3.4) shown in Table 3.1. The performance measures are for every setting the proportion of individual algorithm runs that reach the global minimum and the mean number of iterations it takes to reach this minimum. The results show that both algorithms, using the local-to-best strategy, succeeded in finding the global minimum for all the 50 runs and all the settings. Further it is seen that the convergence occurred much faster for the adaptive version JADE than for the standard DE algorithm. On average, JADE succeed in estimating the kinetic parameters about 80 times faster than the standard DE. It can also be seen that both algorithms require more time to find the global minimum when using settings with six data points (setting 1 to 4) compared to when using corresponding settings with seven data points (setting 5 to 8). Further we see that the convergence times for the $3rT$ settings with $n = 6$ are 1.5 to 14 times larger than for the corresponding $2rT$ setting.

**Table 3.1:** Evaluation of speed and success of DE and JADE, using local-to-best strategy

| Setting | **DE** | | **JADE** | |
|---|---|---|---|---|
| | Mean | Conv. rate(%) | Mean | Conv. rate(%) |
| 1: $2rT(n = 6), C_1$ | 13922 | 100 | 332 | 100 |
| 2: $2rT(n = 6), C_2$ | 14744 | 100 | 268 | 100 |
| 3: $3rT(n = 6), C_1$ | 58185 | 100 | 756 | 100 |
| 4: $3rT(n = 6), C_2$ | 186129 | 100 | 1125 | 100 |
| 5: $2rT(n = 7), C_1$ | 14309 | 100 | 216 | 100 |
| 6: $2rT(n = 7), C_2$ | 17598 | 100 | 215 | 100 |
| 7: $3rT(n = 7), C_1$ | 26378 | 100 | 269 | 100 |
| 8: $3rT(n = 7), C_2$ | 20133 | 100 | 227 | 100 |

The speed and success of estimating the kinetic parameters using the rand evolution strategy are shown in Table 3.2. The results show that both algorithms, using the rand strategy, succeeded in finding the global minimum for all the 50 runs and all the settings. Just as for the local-to-best strategy, it is seen that the convergence

occurred much faster for the adaptive version JADE than for the standard DE algorithm. On average, JADE succeed in estimating the kinetic parameters about 20 times faster than the standard DE. Again it can be seen that both algorithms require more time to find the global minimum when using settings with six data points (setting 1 to 4) compared to when using corresponding settings with seven data points (setting 5 to 8). Additionally we see that the convergence times for the $3rT$ settings with $n = 6$ are 3 to 9 times larger than for the corresponding $2rT$ setting.

It should be noted that the three values of $c$, $c = 0.01$, $c = 0.1$ and $c = 0.5$ resulted in almost identical convergence speeds and the same convergence rates, which is the reason only one general result is shown for the JADE extensions.

**Table 3.2:** Evaluation of speed and success of DE and JADE, using rand strategy

| Setting | **DE** | | **JADE** | |
|---|---|---|---|---|
| | Mean | Conv. rate(%) | Mean | Conv. rate(%) |
| 1: $2rT(n = 6), C_1$ | 15700 | 100 | 941 | 100 |
| 2: $2rT(n = 6), C_2$ | 12054 | 100 | 607 | 100 |
| 3: $3rT(n = 6), C_1$ | 51042 | 100 | 2249 | 100 |
| 4: $3rT(n = 6), C_2$ | 98177 | 100 | 3970 | 100 |
| 5: $2rT(n = 7), C_1$ | 13425 | 100 | 608 | 100 |
| 6: $2rT(n = 7), C_2$ | 12166 | 100 | 613 | 100 |
| 7: $3rT(n = 7), C_1$ | 13862 | 100 | 809 | 100 |
| 8: $3rT(n = 7), C_2$ | 17894 | 100 | 762 | 100 |

## 3.2 Bayesian optimal design for the the two-catalyst model

For each general design we will create 25 different subdesigns by utilizing different values of $CRd$ and $Td$. These 25 designs will be presented in image plots where colors are used to represent the expected utilities given by the designs. The expected utility is coded on a color scale from red to blue where blue represents higher utility than red. The analysis will be repeated for different ways of changing $grT$.

The considered $CRd$ and $Td$ differ for the two general designs, $2rT$ and $3rT$, since they use different number of $T$ and $CR$ and since it is important to use the same variable range for the two designs, to compare them fairly. The $2rT$ design has three temperatures that together with a $Td$ of 20 creates the same range of $T$, from 350 to 390, as the $3rT$ design with $Td = 40$.

### 3.2.1 Bayesian optimal design for the 2rT design

The $2rT$ design is the design constructed by three changes in temperature and two changes in catalyst ratio. This design results in a sample of six data points where the sampling procedure consists of first setting a catalyst ratio, sampling three different temperatures, then change the catalyst ratio to once again sample the same three temperatures.

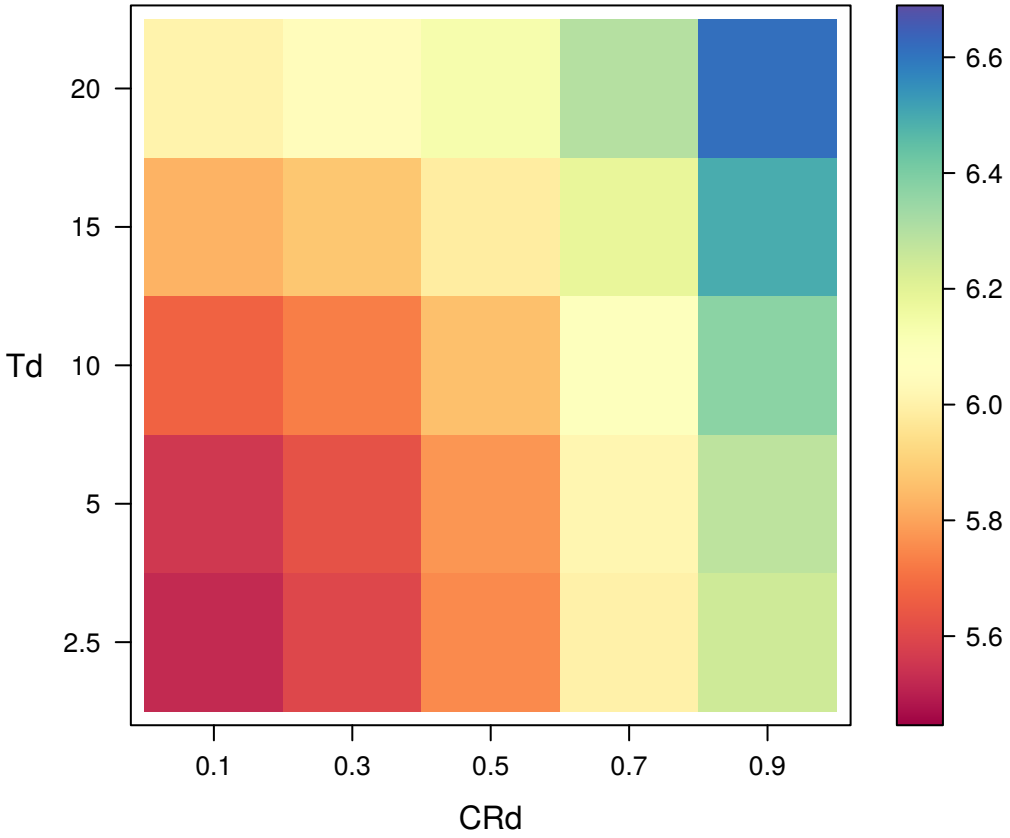


**Figure 3.1:** $U(d)$ for the $2rT$ design with changes in $Td$ and $CRd$ when $grT = 1$

The expected utilities for 25 different designs or combinations of changes in $T$ and $CR$ when $grT = 1$ are shown in Figure 3.1. The results show that the highest expected utility is obtained when both $Td$ and $CRd$ are set as high as possible. These values are $Td = 20$ and $CRd = 0.9$ and give a $U(d)$, measured by the KL divergence from posterior to prior, of 6.61. The results further show that different combinations of $CRd$ and $Td$ around the best combination decay asymmetrically in $U(d)$. This can be seen by that the $U(d)$ decreases slower with decreasing $Td$ compared to decreasing $CRd$. Concluding this observation one can say that the $CR$ plays a higher role when searching for high utility than does the $T$ for the $2rT$ design.

Investigating how different changes in the $grT$ affect the expected utility were further considered in the same manner as in the previous analysis when the $grT$ were fixed to 1. Three different changes in $grT$ were analyzed. The first change simply consists of using a different $grT$, $grT = 1.3$, while the two other consist of changing the $grT$ either for every different $T$ or for every different $CR$, as described in sec. 2.4. When changing the $grT$ for the three temperatures we use $grT = (0.7, 1, 1.3)$ and when changing the $grT$ for the two catalyst ratios we use $grT = (1, 1.3)$. The results in all the three cases are very similar to the case when $grT = 1$. Meaning that the resulting image plots have the same structure and that the same combination of $Td = 20$ and $CRd = 0.9$ give the highest value of the expected utility. The best expected utility values seen for the three ways of changing $grT$ are shown in Table 3.3. It is seen that the best expected utilities are all very similar for all the $2rT$ designs and one can thus conclude that different ways of changing the $grT$ play a minor role when designing informative experiments.

**Table 3.3:** Best values for $U(d)$ for $2rT$ designs created by changing $grT$

| Design | $U(d)$ |
|---|---|
| $grT = 1.3$ | 6.62 |
| $grT(T)$ | 6.61 |
| $grT(CR)$ | 6.60 |

The results from the Bayesian optimal design show that one optimal design found by changing all the three factors of $T$, $CR$ and $grT$ for the $2rT$ is the design in Table 3.4.

**Table 3.4:** Optimal $2rT$ design with $Td = 20$ and $CRd = 0.9$

| $T$ | $rT_1$ | $rT_2$ |
|---|---|---|
| 350 | 0.05 | 0.95 |
| 370 | 0.05 | 0.95 |
| 390 | 0.05 | 0.95 |
| 350 | 0.95 | 0.05 |
| 370 | 0.95 | 0.05 |
| 390 | 0.95 | 0.05 |

### 3.2.2 Bayesian optimal design for the 3rT design

The $3rT$ design is the design constructed by two changes in temperature and three changes in catalyst ratio. This design will just like the $2rT$ design result in a sample of six data points, but with the difference that the sampling procedure now consists of sampling two temperatures for three different catalyst ratios.

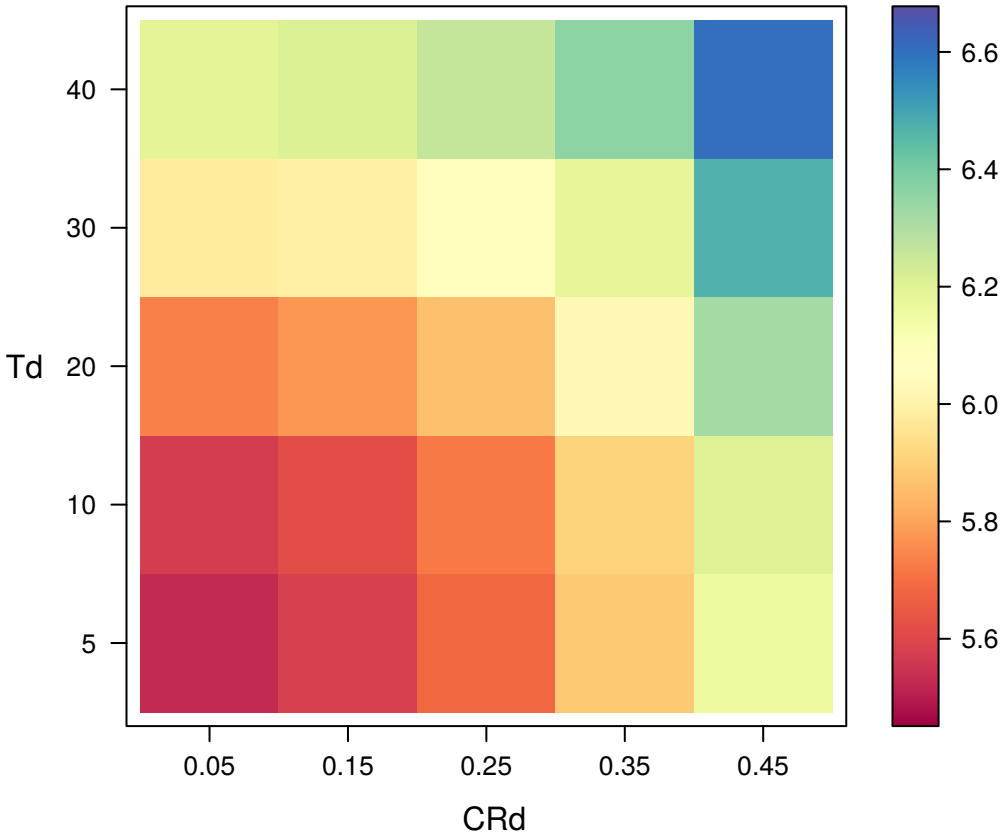


**Figure 3.2:** $U(d)$ for the $3rT$ design with changes in $Td$ and $CRd$ when $grT = 1$

The expected utility for 25 different designs or combinations of changes in $T$ and $CR$ when $grT = 1$ are shown in Figure 3.2. In the plot one sees the same trend as in the analysis of the $2rT$ design, which is that the higher the $Td$ and $CRd$, the higher $U(d)$ is obtained. The highest values of $Td$ and $CRd$ are for this design set to $Td = 40$ and $CRd = 0.45$ which give a $U(d)$, measured by the KL divergence from posterior to prior, of 6.59. This value of the expected utility is similar to the best value seen for the $2rT$ design. The results further show that different combinations of $CRd$ and $Td$ around the best combination decay more symmetrically in $U(d)$ than what we saw for the $2rT$ design, indicating that the $CR$ plays a smaller role when searching for high utility in the $3rT$ design than in the $2rT$ design.

Investigating how different changes in the global residence time affect the expected utility were further considered in the same manner as in the previous analysis when the $grT$ were fixed to 1. Three different changes in $grT$ were analyzed. The first change simply consists of using a different $grT$, $grT = 1.3$, while the two other consist of changing the $grT$ either for every different $T$ or for every different $CR$, as described in sec. 2.4. When changing the $grT$ for the two temperatures we use $grT = (1, 1.3)$ and when changing the $grT$ for the three catalyst ratios we use $grT = (0.7, 1, 1.3)$. The results in all the three cases are very similar to the case when $grT = 1$. Meaning that the resulting image plots have the same structure

and that the same combination of $Td = 40$ and $CRd = 0.45$ gives the highest value of the expected utility. The best expected utility values seen for the three ways of changing $grT$ are shown in Table 3.5. It is seen that the best expected utilities are all very similar for all the $3rT$ designs and one can thus conclude that different ways of changing the $grT$ play a minor role when designing informative experiments.

**Table 3.5:** Best values for $U(d)$ for $3rT$ designs created by changing $grT$

| Design | $U(d)$ |
|---|---|
| $grT = 1.3$ | 6.60 |
| $grT(T)$ | 6.61 |
| $grT(CR)$ | 6.59 |

The results from the Bayesian optimal design show that one optimal design found by changing all the three factors of $T$, $CR$ and $grT$ for the $3rT$ is the design in Table 3.6.

**Table 3.6:** Optimal $3rT$ design with $Td = 40$ and $CRd = 0.45$

| $T$ | $rT_1$ | $rT_2$ |
|---|---|---|
| 350 | 0.05 | 0.95 |
| 390 | 0.05 | 0.95 |
| 350 | 0.50 | 0.50 |
| 390 | 0.50 | 0.50 |
| 350 | 0.95 | 0.05 |
| 390 | 0.95 | 0.05 |

## 3.3 Inference of optimal designs for the two-catalyst model

By using simulations we are going to learn more about the precision given by some good designs, indicated by the BOD in sec. 3.2. The number of samples used for estimating the parameter means and standard errors are $B = 500$. For the density plots, a larger number of samples are used, $B = 5000$, in order to get relatively smooth looking distributions. All the parameter estimates were obtained using JADE, as described in sec. 2.3.3 with the local-to-best evolution strategy described in sec. 2.3.2, $NP = 250$ and $c = 0.1$. Based on the results in sec. 3.1 we use 1000 iterations for the $2rT$ designs and 3000 iterations for the $3rT$ designs. When investigating the effect of the addition of a seventh data point for the $2rT$ designs we randomly reuse one of the $CR$ already in the dataset and set $T = 360$. The procedure is similar for the $3rT$ design, with the only difference in setting $T = 370$.

### 3.3.1 Inference of the first catalyst pair

The first catalyst pair is shown to the left in Table 2.1.

#### 3.3.1.1 Inference of 2rT optimal designs

The mean and standard errors of the kinetic parameters estimated using $2rT$ designs, the first catalyst pair and the saturated model $n = p = 6$ are shown in Table 3.7. The results show that the standard errors of parameter estimates are highly correlated with the expected utility as seen from the BOD in sec. 3.2. The experimental design yielding the highest parameter precision, i.e. the lowest SE is the design with largest values of $CRd$ and $Td$, while the design yielding the lowest precision is the design with the smallest values of $CRd$ and $Td$. The difference in precision given by the best performing design and the worst performing design is noteworthy. By using $(CRd, Td) = (0.7, 15)$ instead of $(CRd, Td = 0.9, 20)$ one can on average expect to obtain a twice as large SE. The results further show that the standard errors are generally higher for the second catalyst than for the first catalyst. These differences are larger when utilizing $CRd = 0.7$ compared to $CRd = 0.9$ and also larger for the rate constant $k_l$ than for other parameters.

**Table 3.7:** Mean and SE for $2rT$ designs, $n = 6$

| | $CRd, Td$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Par | $0.9, 20$ | | $0.7, 20$ | | $0.9, 15$ | | $0.7, 15$ | |
| | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** |
| $n_1$ | 1.30 | 0.09 | 1.32 | 0.10 | 1.30 | 0.15 | 1.30 | 0.16 |
| $n_2$ | 2.10 | 0.15 | 2.11 | 0.19 | 2.11 | 0.25 | 2.11 | 0.29 |
| $Ea_1$ | 32056 | 1408 | 31956 | 1417 | 31996 | 2341 | 31986 | 2212 |
| $Ea_2$ | 51951 | 2067 | 52054 | 2888 | 52014 | 3147 | 52024 | 4245 |
| $k_1$ | 11.95 | 1.27 | 11.97 | 1.26 | 11.97 | 2.34 | 11.96 | 2.10 |
| $k_2$ | 47.61 | 7.04 | 47.60 | 10.07 | 47.62 | 10.52 | 47.62 | 13.94 |

The impact of adding one extra data point on the parameter precision is shown in Table 3.8. The results show the same structure as when $n = 6$ and it is seen that the effect on the precision of the added point is very small.

**Table 3.8:** Mean and SE for $2rT$ designs, $n = 7$

| $CRd, Td$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Par | $0.9, 20$ | | $0.7, 20$ | | $0.9, 15$ | | $0.7, 15$ | |
| | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** |
| $n_1$ | 1.31 | 0.09 | 1.29 | 0.10 | 1.30 | 0.15 | 1.30 | 0.16 |
| $n_2$ | 2.11 | 0.15 | 2.10 | 0.18 | 2.09 | 0.23 | 2.10 | 0.28 |
| $Ea_1$ | 32052 | 1371 | 32002 | 1430 | 31992 | 2339 | 31995 | 2248 |
| $Ea_2$ | 51952 | 2079 | 51982 | 2820 | 51962 | 2956 | 51967 | 4005 |
| $k_1$ | 11.98 | 1.23 | 11.95 | 1.26 | 11.97 | 2.31 | 11.99 | 2.14 |
| $k_2$ | 47.64 | 7.07 | 47.62 | 9.98 | 47.59 | 10.00 | 47.63 | 13.44 |

#### 3.3.1.2 Inference of 3rT optimal designs

The means and standard errors of the kinetic parameters estimated using $3rT$ designs, the first catalyst pair and the saturated model $n = p = 6$ are shown in Table 3.9. The results show that the estimates are biased and generally tend to be slightly higher than the true parameter values. Further differences to the previously seen $2rT$ designs are that the SE generally are smaller for the first catalyst and larger for the second catalyst.

**Table 3.9:** Mean and SE for $3rT$ designs,$n = 6$

| $CRd, Td$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Par | $0.45, 40$ | | $0.35, 40$ | | $0.45, 30$ | | $0.35, 30$ | |
| | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** |
| $n_1$ | 1.38 | 0.09 | 1.39 | 0.11 | 1.39 | 0.10 | 1.41 | 0.11 |
| $n_2$ | 2.32 | 0.19 | 2.33 | 0.19 | 2.31 | 0.19 | 2.34 | 0.18 |
| $Ea_1$ | 33350 | 1379 | 33348 | 1539 | 33388 | 1409 | 33622 | 1544 |
| $Ea_2$ | 54690 | 2461 | 54911 | 2661 | 54473 | 2406 | 55031 | 2525 |
| $k_1$ | 13.28 | 1.33 | 13.29 | 1.48 | 13.31 | 1.36 | 13.55 | 1.5 |
| $k_2$ | 58.35 | 9.77 | 59.49 | 10.63 | 57.33 | 9.44 | 59.77 | 10.13 |

The impact of adding one extra data point on the parameter precision is shown in Table 3.10. The results show similar structure as when $n = 6$ but with significantly less biased estimates. The best $3rT$ design with $n = 7$, $(CRd, Td) = (0.45, 40)$, will yield similar precision in parameter estimates as the best $2rT$ designs with $n = 6$ and $n = 7$ data points.

**Table 3.10:** Mean and SE for $3rT$ designs, $n = 7$

| $CRd, Td$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Par | 0.45, 40 | | 0.35, 40 | | 0.45, 30 | | 0.35, 30 | |
| | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** |
| $n_1$ | 1.30 | 0.09 | 1.31 | 0.09 | 1.29 | 0.12 | 1.31 | 0.13 |
| $n_2$ | 2.10 | 0.20 | 2.12 | 0.21 | 2.09 | 0.26 | 2.12 | 0.28 |
| $Ea_1$ | 31983 | 1327 | 32143 | 1359 | 31893 | 1886 | 32093 | 1888 |
| $Ea_2$ | 52058 | 2359 | 52324 | 2671 | 51998 | 3121 | 52400 | 3502 |
| $k_1$ | 11.99 | 1.18 | 12.15 | 1.22 | 11.99 | 1.68 | 12.16 | 1.70 |
| $k_2$ | 48.51 | 8.23 | 49.63 | 9.45 | 48.82 | 10.89 | 50.50 | 12.23 |

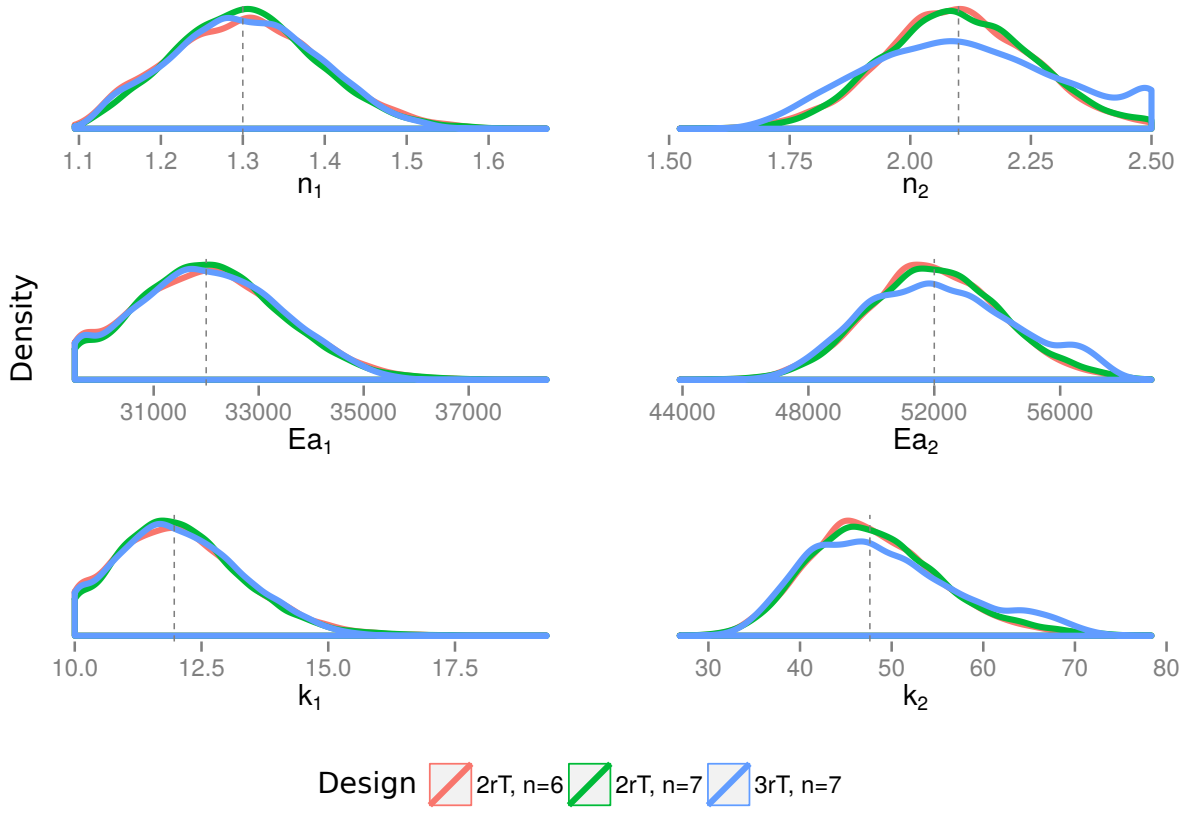


**Figure 3.3:** Sampling distributions for different designs $1^{st}$catalyst pair

Density plots showing the sampling distributions for the two most successful $2rT$ designs and the most successful $3rT$ design are further shown in Figure 3.3. Each color corresponds to the density obtained using one design. The reference lines in the plots show the true parameter values, using which the data was generated. In the plot it is seen that the sampling distributions for the kinetic parameters using

the $3rT$ design with $n = 7$ almost are as concentrated around the true value as the distributions obtained using the two $2rT$ designs. The results that the precision is higher for the first catalyst and that these contrasts are larger for the rate constant $k$ than for other parameters are quite evident. For $k_1$ one sees most of the values between 10 and 16 while $k_2$ is spread from around 30 to 70. For the reaction order one sees that almost all the mass of $n_1$ is concentrated between 1.1 and 1.5 while the $n_2$ is spread between 1.75 and 2.5.

## 3.3.2 Inference of the second catalyst pair

The second catalyst pair is shown to the right in Table 2.1.

### 3.3.2.1 Inference of 2rT optimal designs

The means and standard errors of the kinetic parameters estimated using $2rT$ designs, the second catalyst pair and the saturated model $n = p = 6$ are shown in Table 3.11. The results show that the standard errors of parameter estimates are highly correlated with the expected utility as seen from the BOD in sec. 3.2. The experimental design yielding the highest parameter precision, i.e. the lowest SE is the design with the largest values of $CRd$ and $Td$ while the design yielding the lowest precision is the design with the smallest values of $CRd$ and $Td$. The difference in precision given by the best performing design and the worst performing design is noteworthy. By using $(CRd, Td = 0.7, 15)$ instead of $(CRd, Td = 0.9, 20)$ one will on average obtain a almost twice as large SE. Compared to the corresponding inference on the first catalyst pair one notes that the precision is more evenly distributed among the two catalysts. Actually, we see higher SE for $n_1$ and $Ea_1$ than we do for $n_2$ and $Ea_2$. Further difference to the corresponding inference on the first catalyst pair is that the precision overall is a bit better.

**Table 3.11:** Mean and SE for $2rT$ designs, $n = 6$

| $CRd, Td$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Par | $0.9, 20$ | | $0.7, 20$ | | $0.9, 15$ | | $0.7, 15$ | |
| | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** |
| $n_1$ | 1.40 | 0.12 | 1.40 | 0.15 | 1.41 | 0.18 | 1.41 | 0.24 |
| $n_2$ | 1.20 | 0.05 | 1.20 | 0.05 | 1.19 | 0.08 | 1.20 | 0.08 |
| $Ea_1$ | 45001 | 1556 | 45003 | 1774 | 45020 | 2189 | 44983 | 2670 |
| $Ea_2$ | 49987 | 1173 | 50012 | 1315 | 50011 | 1657 | 49987 | 1853 |
| $k_1$ | 30.04 | 3.36 | 30.05 | 3.84 | 30.07 | 4.73 | 30.05 | 5.89 |
| $k_2$ | 47.66 | 4.2 | 47.62 | 4.72 | 47.62 | 5.97 | 47.64 | 6.71 |

The impact of adding one extra data point on the parameter precision is shown in Table 3.12. The results show no or insignificant differences to the case when $n = 6$.

**Table 3.12:** Mean and SE for $2rT$ designs, $n = 7$

| $CRd, Td$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Par | $0.9, 20$ | | $0.7, 20$ | | $0.9, 15$ | | $0.7, 15$ | |
| | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** |
| $n_1$ | 1.41 | 0.11 | 1.40 | 0.15 | 1.41 | 0.21 | 1.40 | 0.23 |
| $n_2$ | 1.19 | 0.05 | 1.19 | 0.05 | 1.19 | 0.08 | 1.20 | 0.09 |
| $Ea_1$ | 45021 | 1465 | 45004 | 1851 | 44993 | 2445 | 45003 | 2419 |
| $Ea_2$ | 500015 | 1147 | 49987 | 1329 | 49999 | 1671 | 50012 | 1887 |
| $k_1$ | 30.04 | 3.17 | 30.02 | 3.92 | 30.05 | 5.29 | 30.02 | 5.15 |
| $k_2$ | 47.61 | 4.11 | 47.66 | 4.87 | 47.61 | 6.16 | 47.59 | 6.82 |

#### 3.3.2.2 Inference of 3rT optimal designs

The mean and standard errors of the kinetic parameters estimated using $3rT$ designs, the second catalyst pair and the saturated model $n = 6$ are shown in Table 3.13. Overall one can observe much higher standard errors compared to when using the $2rT$ designs, especially for the second catalyst where we see a six times larger SE for $k_2$ when estimated using this $3rT$ design. We also see bias in the estimates, just as we did in the previous analysis with the first catalyst pair.

**Table 3.13:** Mean and SE for $3rT$ designs, $n = 6$

| $CRd, Td$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Par | $0.45, 40$ | | $0.35, 40$ | | $0.45, 30$ | | $0.35, 30$ | |
| | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** |
| $n_1$ | 1.39 | 0.24 | 1.45 | 0.36 | 1.43 | 0.24 | 1.44 | 0.33 |
| $n_2$ | 1.32 | 0.30 | 1.34 | 0.29 | 1.34 | 0.32 | 1.33 | 0.31 |
| $Ea_1$ | 44815 | 2728 | 45564 | 4003 | 45495 | 2582 | 45405 | 3405 |
| $Ea_2$ | 52937 | 6497 | 53248 | 6323 | 53153 | 6450 | 52821 | 6298 |
| $k_1$ | 30.37 | 6.45 | 32.62 | 10.95 | 31.3 | 5.83 | 31.81 | 8.33 |
| $k_2$ | 66.74 | 30.50 | 67.69 | 29.76 | 66.78 | 30.19 | 65.44 | 29.61 |

The impact of adding one extra data point on the parameter precision is shown in Table 3.14. The results show a slight improvement to the $n = 6$ case, in terms of SE and bias. Despite the improvement to the case with six data points, the precision is far from being as good as with the $2rT$ design. The only difference to the $2rT$ designs is seen in the precision for the second catalyst pair where we see about six times larger SE when using the $3rT$ design.

**Table 3.14:** Mean and SE for $3rT$ designs, $n = 7$

| $CRd, Td$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Par | 0.45, 40 | | 0.35, 40 | | 0.45, 30 | | 0.35, 30 | |
| | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** | **Mean** | **SE** |
| $n_1$ | 1.39 | 0.10 | 1.36 | 0.16 | 1.39 | 0.13 | 1.40 | 0.20 |
| $n_2$ | 1.30 | 0.25 | 1.26 | 0.22 | 1.28 | 0.25 | 1.24 | 0.21 |
| $Ea_1$ | 44496 | 1348 | 44594 | 1815 | 44389 | 1589 | 45105 | 2165 |
| $Ea_2$ | 52994 | 5260 | 51398 | 4903 | 51758 | 5273 | 50952 | 4585 |
| $k_1$ | 29.14 | 2.84 | 29.47 | 3.65 | 30.06 | 3.39 | 30.32 | 4.68 |
| $k_2$ | 61.86 | 24.10 | 55.90 | 22.79 | 58.69 | 24.44 | 54.26 | 20.34 |

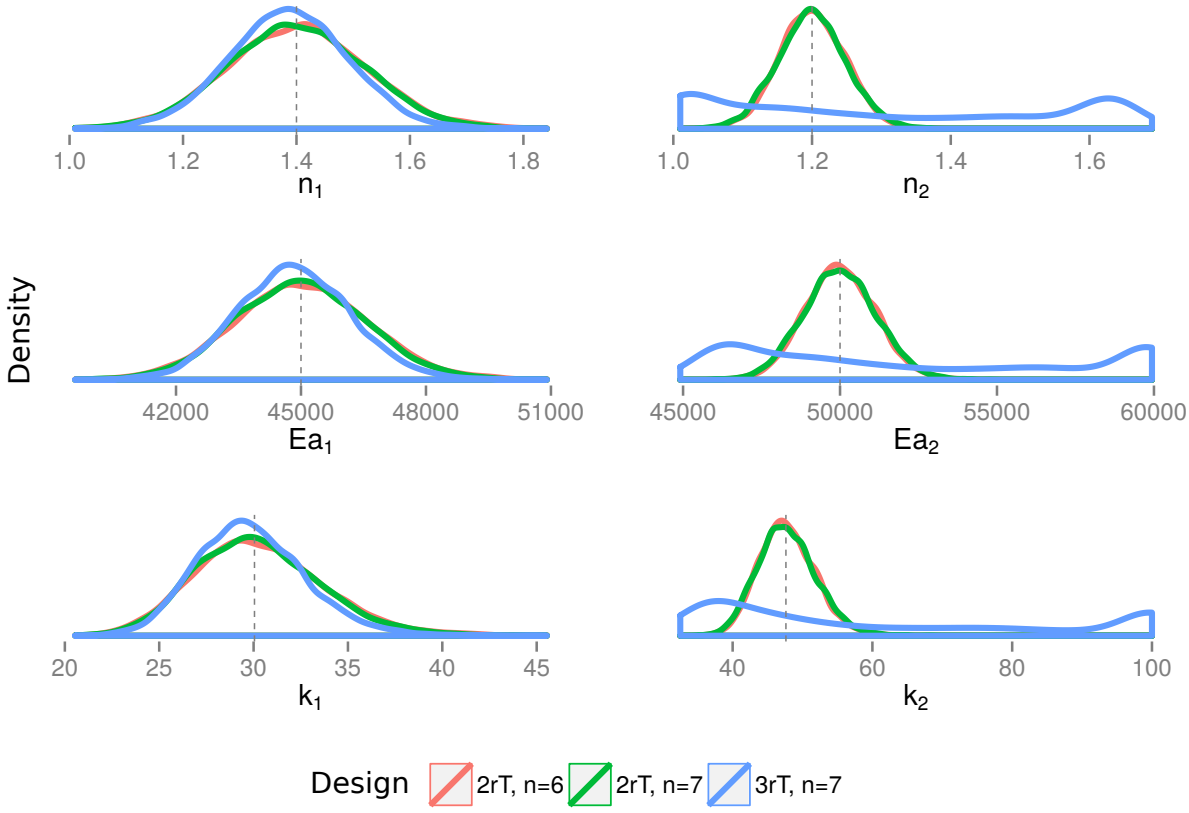


**Figure 3.4:** Sampling distributions for different designs, $2^{nd}$catalyst pair

Density plots showing the sampling distribution for the two most successful $2rT$ designs and the most successful $3rT$ design are shown in Figure 3.4. Each color corresponds to the density obtained using one design. The reference lines in the plots show the true parameter values, using which the data was generated. In the plot it is seen that the sampling distributions for the kinetic parameters using the

$3rT$ design with $n = 7$ greatly differ from the sampling distributions given by the two $2rT$ designs. The spread is much larger for the second catalyst using the $3rT$ design than for the same catalyst using the two $2rT$ designs, indicating that the $3rT$ design is a worse choice when creating hydrotreating experiments yielding high precision. It is further seen that the $n_1$, for the $2rT$ designs, has most of its mass between 1.1 and 1.7 and that the $n_2$ has most of its mass between 1.1 and 1.3. For the activation energy it is seen that the $Ea_1$, using the $2rT$ designs, has most of its mass between 42000 and 48000 and that $n_2$ has most of its mass between 47000 and 53000.

# 4 Discussion

## 4.1 Discussion of optimal designs

The general design with two changes in catalyst ratio ($2rT$) were shown to be a better way of designing experiments than the general design with three changes in catalyst ratio ($3rT$). This is a good result since it means that the design with the lower experimental cost also is the one giving the better precision. With these results one can also conclude that the information given by the temperatures is more important than the information given by the catalyst ratios. The choice of temperatures to use in the hydrotreating process are very limited with common values from 350 to 390 °C. It is thus not possible to vary the temperature very much which might be the reason that the addition of a third temperatures, as in the $2rT$ design, is more important for the parameter precision than the addition of a third catalyst ratio. Another idea for the $2rT$ being superior is that the $3rT$ design includes 50/50 catalyst ratios and that this makes it difficult to identify the effect of each catalyst.

From the Bayesian optimal design we saw similar performance between the two general designs, while the simulation study later on showed that the parameters estimated using the $3rT$ design were badly estimated, with large bias and bad precision. One reason the results differ between the two methods may come from that the BOD do not incorporate any information about the optimization problem and the algorithm. Another reason might be the choice of catalyst parameters and that the used parameters work better for the $2rT$ design than for the $3rT$ design. This is supported by that the BOD integrates out the parameters, meaning that the utility values seen are averaged over many possible catalyst parameters, and that some of these parameters work better with a given design than others.

The conclusion from the result that larger differences in catalyst ratio give higher precision in parameter estimates, compared to small ones points to that estimates obtained using a one-catalyst model will be more precise than estimates using a stacked-catalyst model. This observation was expected since a lot of uncertainty is placed in the stacked-catalyst model when internally estimating the impurity concentrations between different catalyst regions. The arising question is whether this decrease in precision using the stacked-catalyst model is large enough to consider staying with the old method, estimating one catalyst at a time. The answer to this question is not straightforward and estimating the parameters using the stacked-catalyst model should rather be seen as a possible alternative way of estimating the

parameters which might or might not be more appropriate than the old method in a given situation. For situations where other research experiments simultaneously are conducted using the stacked-catalyst model, it will most often be more appropriate to use the same data for the parameter estimation than to additionally start new and costly one-catalyst experiments. This will of course only be a good idea if the experiments are conducted using a design yielding a reasonable precision.

Changing the global residence time between the experimental runs yielded no or very small effect on the expected utility. This shows that it only is a matter of how much the residence times differ in percentage (catalyst ratio) and not absolute magnitude of the changes. The small differences in expected utility seen between the best designs, shown in Table 3.4 and Table 3.6, are mainly due to the errors introduced by approximating the integrals in the expression of the expected utility with a Monte Carlo estimate.

Another noteworthy result shown by the Bayesian optimal design is that the information given by the temperature and catalyst ratio both play a large importance in the precision and that it is crucial to focus on both these factors when conducting experiments. This was seen in the results when designs with more information in temperature, $2rT$ designs, benefited more from sampling the two catalyst ratios with large differences between each other than sampling the three temperatures with large differences between each other. The opposite is also seen, that designs with more information given by the catalyst ratios benefit more from well designed temperature changes than well designed catalyst ratio changes. A similar observation is also very evident in the simulation study where one extra data point were added to datasets in order to observe how the addition of this extra data point affected the parameter precision. The added point consisted of sampling one of the already used catalyst ratios with a completely new temperature. The only new information added to the data sets was thus another temperature. When adding this data point to the dataset with two previous temperatures ($3rT$) we saw increases in parameter precision but no real difference in precision were seen when adding it to the dataset already containing three different temperatures ($2rT$).

Further observations in the simulation study and for the $2rT$ designs is that the parameter precision is highly catalyst dependent, and a higher parameter precision is seen for stronger and more effective catalysts. For the first catalyst pair we could note better precision for the first and more effective catalyst than for the second catalyst. The same observation is seen for the second catalyst pair where the second catalyst is slightly more effective and also gives higher parameter precision. The idea that stronger catalysts will obtain better precision is supported by that the effect of the stronger catalyst will be dominating the process and resulting in more information identifying this catalyst than information identifying the weaker one.

## 4.2 Discussion of methods

The adaptivity of JADE has shown to play a great role in the estimation of the kinetic parameters of the model. With standard settings we have seen that the adaptive version can be about 80 times faster to convergence than the standard method, without parameter adaptation. These results are important and results from that the standard settings used in the standard DE ($CP = 0.5, F = 0.8$) not are very optimal for the hydrotreating model, failing in defining the relationship between the kinetic parameters, but also because there is a need for different control parameters at different stages of the run of the algorithm.

Further results show that the choice of control parameter $c$ in the JADE extension carries no importance in the success of the algorithm. The three different values $c = (0.01, 0.1, 0.5)$ performed equally well. This parameter was though to play a larger role, performing worse for low values like $c = 0.01$. It was believed that a too low value would slow down the development of the population, by not being able to adapt fast enough to developments in the population.

The differences in results between the two evolution strategies, rand and local-to-best are small. The differences lie in that the local-to-best strategy performs better of the two when applying parameter adaptation while the rand strategy performs better when there is no parameter adaptation. The reason local-to-best is the worse strategy without parameter adaptation is because its greedy strategy always moves the population towards the best vector, which is not always a good idea. The algorithm gets more often delayed in bad areas, and the fixed control parameter makes it hard to move the best vector in a different and better direction of the parameter space. What happens when using parameter adaptation is that the control parameters will adapt, allowing faster moves from bad to good areas in the parameter space. The parameter adaptation benefits the local-to-best strategy more than the rand strategy since its target vectors are more similar from one generation to another, making successful control parameters in previous generation more appropriate for the coming generation.

The reason for considering the rand strategy is that the local-to-best strategy is known to lead to premature and false convergence since it greedily uses the best target vector in the creation of all the offsprings. For this specific problem there were no issues with premature and false convergence when using the local-to-best strategy and the need for the rand strategy is not necessary and would just lead to a slower convergence.

The performance evaluation of the optimization algorithms further shows that faster convergence occur when adding one extra data point ($n = 7$) compared to when using the saturated model ($n = 6$). The difference in the results are due to the logic that more information exists in seven data points than in six data points and that the more information in the data, the easier it is to find the global minimum. We also observed slower convergence for the $3rT$ design than for the $2rT$ design, which

is another observation why the $2rT$ design is to be preferred.

The utility function used in the Bayesian optimal design was the Kullback-Leibler divergence (Kullback and Leibler, 1951) from posterior to prior . This utility function was chosen based on that it is highly related to the generalized variance of parameters, which is what is of interest in this study. This resemblance to the parameter variance was evident when we also, using simulations, quantified the variability of parameter estimated by different designs.

For general optimal design problems it is common practice to consider all possible designs and apply some optimization algorithm to find close to optimal designs, as in Cook et al. (2008); Huan and Marzouk (2013); Ryan et al. (2014). The problem with this methodology is that one only obtains good designs and do not learn anything about bad designs. In this thesis it was of more interest to learn how some prespecified designs compete to each other in order to observe how different types of changes to the reactor and general designs affect the precision.

# 5 Conclusions

A successfully applied method for estimating the kinetic parameters of the stacked-catalyst hydrotreating model is an adaptive differential evolution algorithm called JADE, described in Zhang and Sanderson (2009). The adaptivity in JADE has shown to be very useful in this problem, resulting in up to 80 times faster convergence compared to when using the standard DE algorithm.

The general design yielding best precision in parameter estimates when conducting two-catalyst experiments is a design consisting of sampling three temperatures for two different catalyst ratios. The design consisting of sampling two temperatures for three different catalyst ratios gives bad parameter precision.

Experiments should be conducted by sampling the temperatures and catalyst ratios to be as different from each other as possible in order to yield good parameter precision. Changing the global catalyst ratio has no effect on the parameter precision.

It is sufficient with six experimental runs or data points to estimate the six parameters in the two-catalyst model with reasonable precision, given a good design. Adding one additional data point to the experimental data set might not always increase the precision in parameter estimates.

The focus in this thesis has been on the two-catalyst model but the same results seen here are reasonable to assume for models larger than the two-catalyst models. The expected differences for larger models are that the parameter uncertainty will grow and larger standard errors will be seen. One possible continuation of this work would be to quantify this uncertainty. Another interesting continuation of this thesis would be to perform a similar analysis but utilize a more complex hydrotreating model containing more parameters and operating variables.

# Bibliography


Atkins, P. and De Paula, J. (2010). *Physical chemistry for the life sciences*. Oxford University Press.

Babu, B. and Angira, R. (2006). Modified differential evolution (mde) for optimization of non-linear chemical processes. *Computers & chemical engineering*, 30(6):989–1002.

Bullen, P. S. (2003). *Handbook of means and their inequalities*. Springer Science & Business Media.

Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Cook, A. R., Gibson, G. J., and Gilligan, C. A. (2008). Optimal observation times in experimental epidemic processes. *Biometrics*, 64(3):860–868.

Dolan, W., Cummings, P., and LeVan, M. (1989). Process optimization via simulated annealing: application to network design. *AIChE Journal*, 35(5):725–736.

Eftaxias, A., Font, J., Fortuny, A., Fabregat, A., and Stüber, F. (2002). Nonlinear kinetic parameter estimation using simulated annealing. *Computers & chemical engineering*, 26(12):1725–1733.

Finlayson, B. A. (2003). *Nonlinear analysis in chemical engineering*. Bruce Alan Finlayson.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Artificial Intelligence. Addison-Wesley.

Huan, X. and Marzouk, Y. M. (2013). Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317.

IFPEN (2015). IFP Energies nouvelles, En bref. Internet: http://www.ifpenergiesnouvelles.fr. Retrieved April 16, 2015.

Kirkpatrick, S., Vecchi, M., et al. (1983). Optimization by simmulated annealing. *science*, 220(4598):671–680.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, pages 79–86.

Lindley, D. V., Lindley, D. V., and Lindley, D. V. (1972). *Bayesian statistics: A review*. SIAM.

Luenberger, D. G. and Ye, Y. (2008). *Linear and nonlinear programming*, volume 116. Springer Science & Business Media.

McDonald, J. H. (2009). *Handbook of biological statistics*, volume 2.

McNaught, A. D. and McNaught, A. D. (1997). *Compendium of chemical terminology*, volume 1669. Blackwell Science Oxford.

Müller, P. (2005). Simulation based optimal design. *Handbook of Statistics*, 25:509–518.

Price, K., Storn, R. M., and Lampinen, J. A. (2006). *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media.

Pukelsheim, F. (1993). *Optimal design of experiments*, volume 50. siam.

Ryan, E. G., Drovandi, C. C., Thompson, M. H., and Pettitt, A. N. (2014). Towards bayesian experimental design for nonlinear models that require a large number of sampling times. *Computational Statistics & Data Analysis*, 70:45–60.

Storn, R. and Price, K. (1995). *Differential evolution-a simple and efficient adaptive scheme for global optimization over continuous spaces*, volume 3. ICSI Berkeley.

Topsøe, H., Clausen, B. S., and Massoth, F. E. (1996). *Hydrotreating catalysis*. Springer.

Zhang, J. and Sanderson, A. C. (2009). Jade: adaptive differential evolution with optional external archive. *Evolutionary Computation, IEEE Transactions on*, 13(5):945–958.

LIU-IDA/STAT-A–15/006—SE